At the Intersection of Medicinal Chemistry and Chemical Biology: Development of Chemical Probes for Target Inhibition, Degradation, and Labeling

DISSERTATION

zur Erlangung des Grades
der Doktorin der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von

Apothekerin Theresa Manz, Dipl.-Pharm.

Saarbrücken
2021

Tag des Kolloquiums:	27. Juni 2022
Dekan:	Prof. Dr. Jörn Eric Walter
Berichterstatter:	Prof. Dr. Christian Ducho Prof. Dr. Nathanael S. Gray Prof. Dr. Rolf W. Hartmann Prof. Dr. Stefan A. Laufer
Akad. Mitglied:	Dr. Josef Zapp
Vorsitz:	Prof. Dr. Alexander Titz







At the Intersection of Medicinal Chemistry and Chemical Biology: Development of Chemical Probes for Target Inhibition, Degradation, and Labeling

A Dissertation Presented by

Theresa Manz
Dipl.-Pharm./licenced pharmacist

in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in the Subject of Pharmaceutical and Medicinal Chemistry
(Department of Pharmacy)

2021

Die vorliegende Dissertation wurde im Zeitraum von November 2015 bis August 2017 im Arbeitskreis für Pharmazeutische und Medizinische Chemie (Fachrichtung Pharmazie) der Universität des Saarlandes (Saarbrücken, Deutschland) unter der Betreuung von Herrn Prof. Dr. Christian Ducho, und von September 2017 bis Januar 2020 am Department für Cancer Biology des Dana-Farber Cancer Institutes, sowie am Department für Biological Chemistry and Molecular Pharmacology der Harvard Medical School (Boston, MA, USA) im Arbeitskreis von Herrn Prof. Dr. Nathanael S. Gray (unter begleitender Betreuung von Herrn Prof. Dr. Christian Ducho) angefertigt.

This dissertation is the result of the work performed in Prof. Christian Ducho's lab in the subject of Pharmaceutical and Medicinal Chemistry (department of Pharmacy) at the Saarland University (Saarbrücken, Germany) from November 2015 to August 2017, as well as in Prof. Nathanael S. Gray's lab (under supervision of Prof. Christian Ducho) as part of the department of Cancer Biology at the Dana-Farber Cancer Insitute and the department of Biological Chemistry and Molecular Pharmacology at the Harvard Medical School (Boston, MA, USA) from September 2017 to January 2020.

Erstgutachter/First Assessor:	Prof. Dr. Christian Ducho
Zusätzlicher Erstgutachter/ Additional First Assessor:	Prof. Dr. Nathanael S. Gray
Zweitgutachter/Second Assessor:	Prof. Dr. Rolf W. Hartmann

To my parents, Ulrike Modrey and Thomas Manz,

my brothers, Johannes Manz, Dominik and Maximilian Modrey,

and the love of my life, Shalom Halbert,

whithout whom none of this would have been possible.

ACKNOWLEDGEMENTS

While this thesis, and the years of work it summarizes, is one of the most proud accomplishment of my life, I could not have made it this far without the incredible help, support, advice, guidance, motivation, and inspiration of so many talented scientists I had the honor to work with or under.

First and foremost, I would like to thank Prof. Dr. Christian Ducho for taking me on as a Ph.D. student and for giving me the opportunity to grow as a scientist and chemist within his lab. I am especially thankful for his amazing support and care, his passion for detail and excellence, and for being my reliable guide during the time I was able to spend in Prof. Dr. Nathanael S. Gray's lab in Boston. I feel very blessed that I was given the chance to learn and work in Germany and the US, and I am certain that it has helped me immensely in my development as a person as well as a scientist. Thank you, Prof. Ducho, for making this possible and for making my journey as a Ph.D. student so fulfilling and memorable!

On that note, I would also like to thank Prof. Dr. Nathanael S. Gray for his unparalleled support and generosity during not only my Ph.D. but also my "Dipl.-Pharm." degree. I am unbelievably grateful for his welcoming personality, for providing such an amazing place to work and learn, and for sharing his wisdom and guidance to help me find my path. I cannot emphasize enough how my time in Boston has shaped my life and words cannot express how thankful I am for that. I wish you, Prof. Gray, all the best for your new and exciting chapter in California, and I cannot wait to come and visit your new lab space!

I would also like to thank Prof. Dr. Rolf W. Hartmann for giving me my first job in his lab as a student, for supporting me in performing my "Dipl.-Pharm." thesis in Prof. Gray's lab, and for offering to be my scientific tutor and committee member during my time as a Ph.D. student. Prof. Hartmann – I am unbelievably grateful for the opportunities I had because of your support and encouragement. Thank you!

In Germany, I had the pleasure of working among many wonderful students, whom I would like to thank for creating such a warm and caring atmosphere, and for making it so hard to leave for Boston in 2017. In particular, I would like to thank Christian Rohrbacher, Daniel Wiegmann, Giuliana Niro, Katrin Voos, Kristin Leyerer, Malte-Ole Schneemann, Marius Wirth, Melissa Meng, Patrick Fischer, Ruth Linder, Stefan Koppermann, and Verena Böttner. Thank you all for the many fun days we had in lab, for much laughter, for after-lab beers and karaoke, for many helpful discussions, and support in need! Ruth – thank you for being a wonderful hood-neighbor and column-companion. Kristin – thank you for all your help tackling the chemistry of muraymycins. Stefan – thank you so much for testing my final compound and giving me the happy news. Chris – thank you for your bravery in handling our chemicals together. Giuliana – you are my most favorite karaoke-partner and you were always a great motivation for me to become a better chemist. Thank you for your friendship and support throughout those long years. I would also like to thank Dr. Stefan Böttcher for performing HRMS of my compounds, and for always being such a helpful and fun chemist to work with and learn from, as well as Dr. Josef Zapp for his excellent help in performing NMR analysis of my compounds.

In Boston, I was lucky enough to work alongside an amazing team of incredibly talented and smart scientists. I am gladly thinking back to all the wonderful memories I have from this time. Thank you, Gray lab, for being such a special place to learn and grow! I would

like to specifically thank Dr. Tinghu Zhang, who has been my mentor and guide for many years. Thank you, Tinghu, for believing in me and for teaching me all your insights into becoming a great medicinal chemist. Your passion for science is contagious and I could not have been luckier than having you as my supervisor. I would also like to point out Dr. Fleur Ferguson, who is one of the most productive and amazing scientists I have ever had the chance to meet. Thank you, Fleur, for guiding me through my scientific adventures the last couple of years. You have been an amazing support through hard times, and you are such an inspiration in all you do.

As part of former or current members of the Gray lab, I would like to thank Alan L. Leggett, Dr. med. Alexander Parent, Dr. Alyssa Verano, Dr. Baishan Jiang, Dr. Behnam Nabet, Dr. Benika Pinch, Dr. Brian Groendyke, Dr. Calla Olson, Dr. Carmen Sivakumaren, Dr. Chelsea Powell, Dr. Christopher M. Browne, Dr. David Scott, Dr. Dries De Clercq, Dr. Eric Wang, Eriko Koide, Dr. Guangyang Du, Inchul You, Dr. Jaebong Jang, Dr. Jianwei Che, Dr. Jie Jiang, Dr. Jinhua Wang, Dr. Mei Zeng, Dr. Mengyang Fan, Mikaela Mohardt, Dr. Mingxing Teng, Nathanael J. Henning, Dr. Ryosuke Ishida, Dr. Suman Rao, Dr. Wenchao Lu, Dr. Wenzhi Ji, Dr. Yao Liu, Dr. Zainab Doctor, Dr. Zhengnian Li, and Zhixiang He. Thank you all for all the helpful discussions throughout these years of my Ph.D. and for creating a wonderful community to work and thrive in. Yao and Mingxing – thank you for being wonderful hood-neighbors and for all the insightful and educational conversations we have had. David – thank you for being a great desk-neighbor. I highly cherish your compliment regarding my "highly organized" lab book. Jianwei (John) – thank you for teaching me and continuing to teach me so much about computational chemistry and data science. Jie, Benika, Eriko, Behnam, Eric, Carmen, and Inchul – thank you for all your help with experiments and chemistry. I could not have done this without you. Chelsea and Benika – thank you for taking care of my cells when I could not be there to do it myself and for the many lunches and coffees we have shared. Your friendship has been an immense blessing and helped me so, so much to make it through this Ph.D. in one piece. Carmen – thank you for being my PIP4K-partner and for accompanying me at "Dana's got Talent". Inchul – thank you for all the helpful and motivating conversations we had. I wish you all the best for the rest of your studies. You got this! Alyssa, Benika, Brian, Chelsea, Eric, Inchul, and Fleur – thank you for all your helpful comments and edits during the process of writing this thesis.

Since my time in Boston included many collaborations, I would also like to thank the labs of Prof. Dr. Ping Lu and Prof. Dr. Jarrod Marto, especially Dr. Scott Ficarro and Joe Card, as well as the group among Dr. Sirano Dhe-Paganon at the Harvard Medical School, the group among Dr. Matthew D. Hall, especially Dr. Adam Yasgar, at the National Center for Advancing Translational Sciences, as well as the labs of Prof. Dr. Nir London at the Weizman Institute of Science and Prof. Dr. Lewis C. Cantley at the Weill Cornell Medicial College. It has been exceptionally educational to work with these amazing teams.

Finally, I would like to thank my family and friends outside of academia for their unconditional support. I love you all from the bottom of my heart! Tobias Fritz – you have been my dear friend for many years and have been my consistent dose of "home" no matter how far apart we may live. Dr. Ricardo de Matos Simoes – thank you for all the nice words of encouragement over the years and for the amazing gift of bringing music back into my life. Yerri Portillo – thank you for always having time to listen to my endless complaints and lifting my spirit when I need it the most. Hannah Britz – thank you for your uplifting

spirit, our long phone calls, and for your incredible help already during our years as undergraduates and now as Ph.D. students. Leah Axelrod – you and your adorable young family have been amazingly supportive and a much-needed source of encouragement and kind words when I was doubting myself. I am so happy to officially being able to call you "family". Thank you so much! Johannes Manz – you as my older brother have always been a huge influence on me and have inspired me to always reach for higher goals all my life. Maximilian and Dominik Modrey – you are my baby brothers and are always so sweet and caring. Thank you for always cheering me on!

Thomas Manz, my "Paps" – thank you for teaching me from a young age to be curious and to appreciate science and nature. I would not be the women scientist I am today without your hard work, love, and care, as well as your constant support in all my goals in life. You will always be my hero. Ulrike Modrey – you are an amazing mom and give so much love. Thank you for encouraging me through all these years and helping me through the roughest times. I love you both so much!

Finally, I would like to thank my husband, Shalom Halbert, for his incredible patience and love. Shalom – you have seen me go through some of the toughest periods of my life, and through all that you were my pillar without whom I could not stand. You fill each day of my life with so much joy and happiness and I feel so lucky to have met you. You have supported me through so many long hours in lab. Whether on weekends or late at night, you would always be there for me. Thank you for giving me the courage, strength, and love I needed to make it this far. I love you!

KURZZUSAMMENFASSUNG

Chemische Sonden stellen essentielle Werkzeuge für Zielproteinindentifikations- und Valdierungsstudien in der Entdeckung und Entwicklung von Arzneistoffen dar. In dieser Doktorarbeit wird die Entwicklung von chemischen Sonden für eine Reihe therapeutisch relevanter Zielproteine beschrieben. Hierfür wurden vielfältige medizinisch-chemische Ansätze, einschließlich kovalenter und reversibler Proteininhibitoren, sowie Strategien zum gezielten Proteinabbau, angewandt.

Im Hauptteil (Part II) werden drei Sondenentwicklungsprojekte für die Inhibierung der Kinasen PI5P4K und CDK12/13, sowie der Isomerase Pin1 präsentiert. Für alle Zielproteine wurden Biotin/Desthiobiotin-markierte Sonden auf der Grundlage der entsprechenden Proteininhibitoren kreiert, welche zur Identifizierung der chemischen Verbindungen als potenzielle Leitstrukturen oder zur Untersuchung der Zielproteinbiologie dienten. Ausgehend von Screeninghits oder Off-targeteffekten von Kinaseninhibitoren konnten darüber hinaus zwei neuartige chemische Strukturserien als PI5P4K-Inhibitoren identifiziert werden. Medizinisch-chemische Optimierungen führten hier zu verbesserter Potenz und Selektivität der initialen Hits *in vitro* und *in cellulo*.

Im Appendix (Part III) werden unpublizierte Sondenentwicklungsprojekte für eine Fragment-basierte Pin1-inhibierende Sonde, eine PAL-abgeleitete Sonde zur Untersuchung des bakteriellen Proteins MraY, sowie die Ansätze zur Entwicklung von PI5P4K- und Pin1-abbauenden Verbindungen diskutiert. Obwohl die MraY-inhibierende Sonde und die proteinabbauenden Verbindungen weitere Optimierung benötigen, gibt die hier dargelegte Arbeit wertvolle Einblicke in die Entwicklung dieser Sonden.

ABSTRACT

Chemical probes are important tools for target identification and validation studies that guide drug discovery and development. In this thesis, the development of chemical probes for therapeutially relevant protein targets is described. To achieve this, a wide variety of medicinal chemistry approaches were applied and include covalent and reversible protein inhibitors as well as targeted protein degradation strategies.

In the main part (Part II), three chemical probe discovery projects, targeting the kinases PI5P4K and CDK12/13, and the isomerase Pin1, are presented. For all three protein targets, biotin/desthiobiotin-labeled probes of the covalent protein binders were developed and enabled the investigation of the target inhibitors' scaffolds as potential leads for further drug development purposes or served to investigate target biology. In case of the PI5P4K inhibitors, two novel scaffolds were identified, starting from screening hits or kinase inhibitor off-targets. Medicinal chemistry optimization efforts led to improved potency and selectivity properties of initial hit compounds *in vitro* and *in cellulo*.

In the Appendix (Part III), the development of unpublished probes is discussed, including a fragment-based probe targeting Pin1, a PAL-inspired probe to investigate the bacterial target MraY, and approaches to the development of PI5P4K- and Pin1-targeting protein degraders. While the MraY-targeting probe and the degrader molecules require further optimization, the presented work gives important insights into the development of these kinds of probes.

PUBLICATIONS

Complete list of publications included in this thesis:

- Publication A:
  Sivakumaren, S. C.; Shim, H.; Zhang, T.; Ferguson, F. M.; Lundquist, M. R.; Browne, C. M.; Seo, H.-S.; Paddock, M. N.; **Manz, T. D.**; Jiang, B.; Hao, M.-F.; Krishnan, P.; Wang, D. G.; Yang, J.; Kwiatkowski, N. P.; Ficarro, S. B.; Cunningham, J. M.; Marto, J. A.; Dhe-Paganon, S.; Cantley, L.C.[#]; Gray, N. S.[#] Targeting the PI5P4K lipid kinase family in cancer using novel covalent inhibitors. *Cell Chem. Biol.* **2020**, *27* (5), 525-537.

- Publication B:
  Manz, T. D.; Sivakumaren, S. C.; Yasgar, A.; Hall, M. D.; Davis, M. I.; Seo, H.-S.; Card, J. D.; Ficarro, S. B.; Shim, H.; Marto, J. A.; Dhe-Paganon, S.; Sasaki, A. T.; Boxer, M. B.; Simeonov, A.; Cantley, L. C.; Shen, M.; Zhang, T.[#]; Ferguson, F. M.[#]; Gray, N. S.[#] Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors. *ACS Med. Chem. Lett.* **2020**, *11* (3), 346-352.

- Publication C:
  Manz, T. D.[*]; Sivakumaren, S. C.[*]; Ferguson, F. M.; Zhang, T.; Yasgar, A.; Seo, H.-S.; Ficarro, S. B.; Card, J. D.; Shim, H.; Miduturu, C. V.; Simeonov, A.; Shen, M.; Marto, J. A.; Dhe-Paganon, S.; Hall, M. D.; Cantley, L. C.; Gray, N. S. Discovery and Structure-Activity Relationship Study of (*Z*)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors. *J. Med. Chem.* **2020**, *63* (9), 4880-4895.

- Publication D:
  Zhang, T.[*]; Kwiatkowski, N.[*]; Olson, C. M.; Dixon-Clarke, S. E.; Abraham, B. J.; Greifenberg, A. K.; Ficarro, S. B.; Elkins, J. M.; Liang, Y.; Hannett, N. M.; **Manz, T.**; Hao, M.; Bartkowiak, B.; Greenleaf, A. L.; Marto, J. A.; Geyer, M.; Bullock, A. N.; Young, R. A.; Gray, N. S. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **2016**, *12* (10), 876-884.

- Publication E:
  Pinch, B. J.; Doctor, Z. M.[*]; Nabet, B.[*]; Browne, C. M.; Seo, H.-S.; Mohardt, M. L.; Kozono, S.; Lian, X.; **Manz, T. D**.; Chun, Y.; Kibe, S.; Zaidman, D.; Daitchman, D.; Yeoh, Z. C.; Vangos, N. E.; Geffken, E. A.; Tan, L.; Ficarro, S. B.; London, N.; Marto, J. A.; Buratowski, S.; Dhe-Paganon, S.; Zhou, X. Z.; Lu, K. P.[#]; Gray, N. S.[#] Identification of a potent and selective covalent Pin1 inhibitor. *Nat. Chem. Biol.* **2020**, *16* (9), 979-987.

[*]these authors contributed equally to this work; [#]co-corresponding authors

CONTRIBUTIONS

Confirmation of Author Contributions

We, authors of one or more of the below mentioned articles, hereby attest the valuable contributions of **Theresa Manz** as described in the respective "Author Contributions" sections of the listed articles (provided at the end of the document).

List of publications:

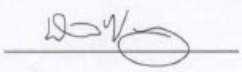
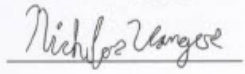
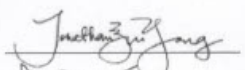
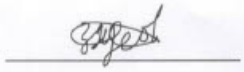
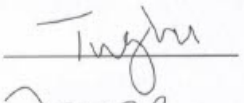
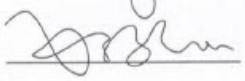
A. Sivakumaren, Shim, Zhang *et al.* Targeting the PI5P4K Lipid Kinase Family in Cancer Using Covalent Inhibitors. *Cell Chem Biol.* **2020**, *27* (5), 525-537.
B. Manz *et al.* Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors. *ACS Med Chem Lett.* **2020**, *11* (3), 346-352.
C. Manz, Sivakumaren *et al.* Discovery and Structure-Activity Relationship Study of (*Z*)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors. *J Med Chem.* **2020**, *63* (9), 4880-4895.
D. Zhang, Kwiatkowski *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol.* **2016**, *12* (10), 876-884.
E. Pinch *et al.* Identification of a potent and selective covalent Pin1 inhibitor. *Nat Chem Biol.* **2020**; DOI: 10.1038/s41589-020-0550-9.

List of Authors:	(Date)	(Signature)
1. Abraham, Brian J.	6/14/2020	
2. Bartkowiak, Bartlomiej	6/22/2020	
3. Boxer, Matthew B.	6/15/2020	
4. Browne, Christopher M.	06/15/2020	Christopher Browne
5. Bullock, Alex N.	22 June 2020	
6. Buratowski, Stephen	June 14, 2020	
7. Cantley, Lewis C.	7/11/2020	

List of Authors (continued):	(Date)	(Signature)
8. Card, Joseph D.	6/15/2020	Joseph D Card
9. Chun, Yujin	06/23/2020	Lim
10. Cunningham, James M.	6/15/20	James M. Cunningham
11. Daitchman, Dina	22/6/20	
12. Davis, Mindy I.	6/15/2020	
13. Doctor, Zainab M.	6/14/20	
14. Dhe-Paganon, Sirano	7/27/20	Sirano Dhe-Paganon
15. Dixon-Clarke, Sarah E.	06/14/2020	Sarah Clarke.
16. Elkins, Jonathan M.	14th June 2020	Jon Elkins
17. Ferguson, Fleur M.	07/13/2020	
18. Ficarro, Scott B.	6/22/2020	
19. Geffken, Ezekiel A.	06/15/2020	
20. Geyer, Matthias	06/29/2020	
21. Gray, Nathanael S.	7/3/2020	

List of Authors (continued):	(Date)	(Signature)
22. Greenleaf, Arno L.	16June2020	Arno L Greenleaf
23. Greifenberg, Ann K.	01.07.2020	A. Greifenberg
24. Hall, Matthew D.	6/15/2020	
25. Hannett, Nancy M.	6/24/20	
26. Hao, Mingfeng	07/25/2020	Mingfeng Hao
27. Jiang, Baishan	06-17-2020	Baishan
28. Kibe, Shin	6/14/2020	Shin Kibe
29. Kozono, Shingo	6/15/2020	
30. Krishnan, Pranav	7/4/2020	Pranav Krishnan
31. Kwiatkowski, Nicholas P.	6/22/2020	
32. Lian, Xiaolan	07/24/2020	Xiaolan Lian
33. Liang, Yanke	06/14/2020	Yanke Liang
34. London, Nir	15/06/2020	
35. Lu, Kun Ping	6/15/2020	

List of Authors (continued):	(Date)	(Signature)
36. Lundquist, Mark R.	06/14/2020	
37. Marto, Jarrod A.	30 July 2020	
38. Miduturu, Chandrasekhar V.	6/22/2020	
39. Mohardt, Mikaela L.	6/15/2020	
40. Nabet, Behnam	06-14-2020	
41. Olson, Calla M.	06/22/2020	
42. Paddock, Marcia N.	7/11/20	
43. Pinch, Benika J.	6/13/2020	
44. Sasaki, Atsuo T.	6/14/2020	
45. Seo, Hyuk-Soo	6/15/2020	
46. Shen, Min	6/14/2020	
47. Shim, Hyeseok	06/14/2020	
48. Simeonov, Anton	6/14/2020	
49. Sivakumaren, S. Carmen	June 15th, 2020	

List of Authors (continued):	(Date)	(Signature)
50. Tan, Li	06/13/2020	
51. Wang, Diana G.	6/15/20	
52. Vangos, Nicholas E.	06/22/2020	
53. Yang, T. Jonathan	6-14-2020	
54. Yasgar, Adam	06-15-2020	
55. Yeoh, Zoe C.	06/22/2020	
56. Young, Richard A.	6/29/2020	
57. Zaidman, Daniel	June 14th, 2020	
58. Zhang, Tinghu	June 14, 2020	
59. Zhou, Xiao Zhen	June 29, 2020	

Publication A: L.C.C., H.S., T.Z. and N.S.G. conceived original project. S.C.S. designed and performed the majority of experiments. T.H.Z., B.S.J., M.F.H. and **T.D.M.** designed and synthesized compounds. C.M.B. and S.B.F. conducted MS experiments. M.R.L. performed autophagy experiments. H.S.S. performed protein purification and crystallography experiments. H.S.S. and S.D.P. performed crystallography data collection and analysis. M.N.P., P.K., D.G.W., and T.J.Y. performed experiments. N.P.K. provided expertise and feedback. J.M.C., J.A.M., S.D.P., L.C.C. and N.S.G. supervised the research. L.C.C. and N.S.G. secured funding. S.C.S. wrote the manuscript with guidance from F.M.F., T.Z., and N.P.K.. S.C.S., H.S., F.M.F., **T.D.M.**, and N.S.G. revised the manuscript. All authors read and provided feedback on the manuscript.

Publication B: **T.D.M**., T.Z., F.M.F., and N.S.G. conceived the project and designed the research strategy. **T.D.M**., T.Z., and F.M.F. designed and synthesized the compounds. S.C.S. performed compound testing and helped with planning experimental strategies. M.B.B. synthesized additional compounds for the biochemical assay validation. A.Y., M.I.D., A.S., and M.D.H. developed and validated the biochemical assays. A.Y. performed the biochemical assays under supervision of M.D.H. and M.S. H.-S.S. and S.D.-P. generated the co-crystal structure. A.T.S. provided reagents and methods for the biochemical assays. S.C.S., J.D.C., S.B.F., and J.A.M. performed the intact mass spectrometry analysis. **T.D.M.** planned and performed the cellular assays. H.S. and L.C.C. were involved in the planning and provided expertise and feedback for the project. The manuscript was written by **T.D.M**. and contributions of all authors. All authors have given approval to the final version of the manuscript. T.Z., F.M.F., and N.S.G. supervised the research.

Publication C: **T.D.M.**, S.C.S., T.Z., F.M.F., and N.S.G. conceived the project and designed the research strategy. S.C.S. performed the library screening and heled with planning experimental strategies. **T.D.M.**, F.M.F., and T.Z. designed the compounds. C.V.M. synthesized the initial hit compound. **T.D.M.** synthesized the compounds presented. H.-S.S. and S.D.-P. generated the co-crystal structure. A.Y. and A.S. performed the biological evaluation of the compounds under supervision of M.D.H. and M.S.. J.D.C. and S.B.F. performed mass spectrometry analysis under supervision of J.A.M.. **T.D.M.** planned and performed the cellular assays. H.S. and L.C.C. were involved in the planning and provided expertise and feedback for the project. The manuscript was written by **T.D.M.** and the contributions of all authors. All authors have given approval of the final version of the manuscript. N.S.G. supervised the research.

Publication D: N.S.G., R.A.Y., N.K., and T.Z. conceived the project. N.S.G. and T.Z. conceived and directed chemistry effort with input from Y.L. Chemical synthesis and small molecule structure determination was performed by T.Z., M.H., and **T.M**. R.A.Y., N.S.G., B.J.A. and N.K. conceived genomics effort. N.K. and C.M.O. designed and executed cellular biological experimental research with input from N.S.G. and R.A.Y. S.E.D.-C. solved co-crystal structure of CDK12-cyclin K with THZ531 with guidance from J.M.E. and A.N.B. A.K.G. designed and executed CDK *in vitro* kinase assays with input from M.G. S.B.F. designed and performed protein mass spectrometry on THZ531-CDK12 adducts with input from J.A.M. N.M.H. provided cloning expertise. B.B. and A.L.G. developed and provided CDK12 and CDK13 antibodies. B.J.A. designed and performed genomics data analyses. N.K., T.Z., N.S.G. and R.A.Y. co-wrote the paper. All authors edited the manuscript.

Publication E: B.J.P. performed the chemical synthesis and biological characterization and wrote the manuscript. Z.M.D. generated cell lines and performed the GFP dropout experiment and RNA-sequencing analysis. B.N. generated cell lines and designed and supervised biological studies. C.M.B., S.B.F. and J.A.M. performed CITe-Id and mass spectrometry studies. M.L.M. assisted in Pin1 mutant cellular studies. S. Kozono, S. Kibe and X.L. performed the FP assay. D.Z., D.D. and N.L. provided computational docking support. L.T. advised on compound design and **T.D.M.** provided support for the chemical synthesis and biological characterization. Z.C.Y., N.E.V., E.A.G. and H. S.S. performed the protein expression and crystallography; H.-S.S. and S.D.-P. analyzed the crystallography data. Y.C. and S.B. performed the yeast complementation and experiments and analysis. K.P.L. and X.Z.Z. helped supervise the project and provided experimental support. N.S.G. conceived of and led this study.

”It’s a dangerous business (…) going out your door. You step onto the road, and if you don’t keep your feet, there’s no knowing where you might be swept off to.”

- J.R.R. Tolkien, The Lord of the Rings

CONTENTS

ACKNOWLEDGEMENTS ..... **i**

KURZZUSAMMENFASSUNG ..... **v**

ABSTRACT ..... **vii**

PUBLICATIONS ..... **ix**

CONTRIBUTIONS ..... **xi**

LIST OF ABBREVIATIONS AND SYMBOLS ..... **xxv**

ABBREVIATIONS OF AMINO ACIDS ..... **xxxiii**

Part I ..... **xxxv**

GENERAL INTRODUCTION ..... **xxxv**

1 Drug and Probe Discovery and Development ..... **- 1 -**

1.1 General Approaches in Drug and Probe Discovery and Development ..... - 1 -

1.1.1 How Medicinal Research Influences Human Health ..... - 1 -

1.1.2 From Bench to Clinic – A Short Overview ..... - 3 -

1.1.3 Approaches in the Drug and Probe Development and Discovery Process ..... - 4 -

1.2 Methods of Target Deconvolution and Validation ..... - 7 -

1.2.1 Techniques of Target Deconvolution ..... - 7 -

1.2.2 Techniques of Target Validation ..... - 10 -

2 The Development and Application of Chemical Probes ..... **- 12 -**

2.1 Medicinal Chemistry and Chemical Biology Orchestrate Drug Development ..... - 12 -

2.2 Chemical Probes – Useful Tools in the Drug Discovery and Development Process ..... - 13 -

2.2.1 Development of Chemical Probes ..... - 13 -

2.2.2 Applications of Chemical Probes in Drug Discovery and Development ..... - 14 -

3 References ..... **- 16 -**

Part II ..... **- 19 -**

PUBLICATIONS ..... **- 19 -**

4 Aims and Scope ..... **- 20 -**

5 Discovery and Development of Chemical Probes for the Study of Lipid Kinases ..... **- 22 -**

5.1 Introduction ..... - 22 -

5.1.1 Label-based and Label-free Methods for Target Identification Studies in Drug Development ..... - 22 -

5.1.2 Targeting Strategies in the Treatment of Cancer – A Brief Overview ..... - 27 -

5.1.3 Targeting Kinases with Small Molecules ..... - 42 -
5.1.4 The Lipid Kinase Phosphatidylinositol 5-Phosphate 4-Kinase (PI5P4K)... - 46 -
5.2 Publication A: *Targeting the PI5P4K Lipid Kinase Family in Cancer Using Novel Covalent Inhibitors* ..... - 61 -
5.2.1 Specific Aims ..... - 61 -
5.2.2 Main Outcomes and Context ..... - 61 -
5.2.3 Published Article ..... - 62 -
5.3 Publication B: *Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors* ..... - 98 -
5.3.1 Specific Aims ..... - 98 -
5.3.2 Main Outcomes and Context ..... - 98 -
5.3.3 Published Article ..... - 99 -
5.4 Publication C: *Discovery and Structure-Activity Relationship Study of (Z)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors* ..... - 160 -
5.4.1 Specific Aims ..... - 160 -
5.4.2 Main Outcomes and Context ..... - 160 -
5.4.3 Published Article ..... - 161 -
5.5 References ..... - 205 -
6 Desthiobiotin Labeling of a Covalent Kinase Inhibitor for the Study of CDK12 and CDK13 ..... - 216 -
6.1 Introduction ..... - 216 -
6.1.1 The Cyclin-Dependent Kinase Family – An Overview ..... - 216 -
6.1.2 CDK12 and CDK13 as Therapeutic Targets ..... - 221 -
6.2 Publication D: *Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors* ..... - 223 -
6.2.1 Specific Aims ..... - 223 -
6.2.2 Main Outcomes and Context ..... - 223 -
6.2.3 Published Article ..... - 224 -
6.3 References ..... - 281 -
7 Novel Chemical Probes for the Study of the Peptidyl-prolyl Isomerase Pin1 . - 284 -
7.1 Introduction ..... - 284 -
7.1.1 Determining Proteome-wide Selectivity of a Covalent Probe *via* a Novel Desthiobiotin-mediated Pulldown Technique ..... - 284 -
7.1.2 The Peptidyl-prolyl Isomerase Pin1 – A Key Regulator of Cellular Processes ..... - 285 -

7.2 Publication E: *Identification of a potent and selective covalent Pin1 inhibitor* - 294 -
7.2.1 Specific Aims ... - 294 -
7.2.2 Main Outcomes and Context ... - 294 -
7.2.3 Published Article ... - 296 -
7.3 References ... - 377 -
8 Conclusion and Perspectives ... - 381 -
Part III ... - 384 -
APPENDIX ... - 384 -
9 Novel Chemical Probes for the Study of the Peptidyl-prolyl Isomerase Pin1 (continued) ... - 385 -
9.1 Introduction ... - 385 -
9.2 Pre-Print: *Sulfopin, a selective covalent inhibitor of Pin1, blocks Myc-driven tumor initiation and growth in vivo* ... - 386 -
9.2.1 Specific Aims ... - 386 -
9.2.2 Main Outcomes and Context ... - 386 -
9.2.3 Pre-print Article ... - 388 -
10 Diazirine-tagged Muraymycin Analogs as Photoaffinity Labeling Probes to Study a Novel Class of Antibiotics ... - 455 -
10.1 Introduction ... - 455 -
10.1.1 Activity-Based Protein Profiling and Photoaffinity Labeling as Tools in Drug Development ... - 455 -
10.1.2 Antibiotics – An Overview ... - 466 -
10.1.3 The Bacterial Cell Wall as a Promising Target for Antibacterial Treatments ... - 474 -
10.1.4 Muraymycin-based Nucleoside Antibiotics ... - 489 -
10.2 Aim of this Project ... - 500 -
10.2.1 Design of Muraymycin-based A*f*BPs ... - 500 -
10.2.2 Retrosynthetic Considerations for the Synthesis of Muraymycin-based Chemical Probes ... - 502 -
10.3 Results and Discussion ... - 505 -
10.3.1 Synthesis of the Main Building Blocks ... - 505 -
10.3.2 Synthesis of the Fatty Acid Building Blocks ... - 509 -
10.3.3 Synthetic Endgame ... - 519 -
10.3.4 Biological Evaluation ... - 527 -
10.4 Conclusion ... - 528 -
10.5 Experimental ... - 533 -

10.5.1 General Methods ..... - 533 -
10.5.2 Analytics ..... - 535 -
10.5.3 General Procedures ..... - 536 -
10.5.4 Synthesis ..... - 537 -
10.6 References ..... - 564 -
11 Development of Protein Degraders from Small Molecule Protein Inhibitors - 578 -
11.1 Introduction ..... - 578 -
11.1.1 Targeted Protein Degradation – A Promising Tool of Chemical Biology ..... - 578 -
11.2 Development of Potential PI5P4K Degraders ..... - 587 -
11.2.1 Aim of This Project ..... - 587 -
11.2.2 Results and Discussion ..... - 589 -
11.2.3 Conclusion ..... - 592 -
11.3 Development of Potential Pin1 Degraders ..... - 593 -
11.3.1 Aim of This Project ..... - 593 -
11.3.2 Results and Discussion ..... - 594 -
11.3.3 Conclusion ..... - 597 -
11.3.4 Experimental ..... - 598 -
11.4 References ..... - 599 -

LIST OF ABBREVIATIONS AND SYMBOLS

^{1}H-NMR	proton nuclear magnetic resonance
Å	Angstrom
aaRS	aminoacyl tRNA synthetase
ABC	ATP-binding cassette transporter
ABP	activity-based probe
ABPP	activity-based protein profiling
acetyl CoA	acetyl coenzyme A
ACN	acetonitrile
ACP	acyl carrier protein
AD	Alzheimer's disease
ADAM	a disintegrin and metalloproteinase
ADME(T)	absorption, distribution, metabolism, excretion, (toxicity)
ADP	adenosine diphosphate
A*f*BP	affinity-based probe
AgOTf	silver trifluoromethanesulfonate
AI	artificial intelligence
ALK	anaplastic lymphoma kinase
ALL	acute lymphoid leukemia
alr	alanine racemase
AML	acute myeloid leukemia
AMP	adenosine monophosphate
AMPK	AMP-activated protein kinase
ANTH	AP180 N-terminal homology
aPBP	class A penicillin-binding protein
APC/C	anaphase-promoting complex/cyclosome
aPK	atypical protein kinase
APP	amyloid precursor protein
APPA	2-amino-5-phosphono-3-pentenoic acid
aq.	aqueous
Ar	argon (gas)
AR	androgen receptor
ARG	antibiotic resistance gene
ARNT	aryl hydrocarbon receptor nuclear translocator
ATF6	activating transcription factor 6
ATM	ataxia telangiectasia mutated
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3-related protein
ATRA	all-*trans* retinoic acid
ATTEC	autophagosome-tethering compound
AUTAC	autophagy-targeting chimeras
BAIB	bis(acetoxy)iodobenzene
BAX	BCL-2 associated X protein
BAK1	BCL-2-antagonist/killer1
BBC3	BCL-2 binding component 3
B.C.	before Christ
BCL-2	B-cell CLL/lymphoma 2
BCR	B-cell receptor

BER	base excision repair
BH3	BCL-2 homolog 3
BID	BH3 interacting domain death agonist
BIM	BCL-2 interacting mediator of cell death
Bn	benzyl
Boc	*tert*-butyloxycarbonyl
bPBP	class B penicillin-binding protein
BRCA1/2	breast cancer type 1/2 susceptibility proteins
BRD4	bromodomain-containing 4
BTK	Bruton's tyrosine kinase
C_{55}-P	undecaprenyl phosphate
Cbz	benzyloxycarbonyl
CD8	cluster of differentiation 8 antigen
CDC25C	cell division cycle 25C
CDK	cyclin-dependent kinase
cDNA	complementary DNA
CETSA	cellular thermal shift assay
CHK	cell cycle checkpoint kinase
cIAP1	cellular inhibitor of apoptosis protein 1
CITe-Id	covalent inhibitor target-site identification
CK2	casein kinase 2
conc.	concentrated
COPD	chronic obstructive pulmonary disease
COT	cancer Osaka thyroid kinase
CRABP	cellular retinol- and retinoic acid-binding protein
CRBN	cereblon
CRISPR	clustered regularly interspaced short palindromic repeats
CRL	cullin RING ligase
CTD	C-terminal domain
CYP	cytochrome P450
d	doublet (NMR)
d	day/s (unit)
Dab1	disabled homolog 1
DAG	diacylglycerol
DARTS	drug affinity responsive target stability
DCAF16	DDB1 and CUL4-associated factor 16
DCM	dichloromethane
DDB1	DNA damage-binding protein 1
ddl	D-Ala-D-Ala ligase
DDR	DNA damage response
DEA	diethylamine
der.	derivative
δ	chemical shift (NMR)
DHFR	dihydrofolate reductase
DHPS	dihydropteroate synthase
DIEA	*N*,*N*-diisopropylethylamine
D-*i*Gln	D-*iso*-glutamine
D-*i*Glu	D-*iso*-glutamate
Dipl.-Pharm.	Diplom Pharmazeut(in)
DMEM	Dulbecco's modified Eagle's medium
DMF	dimethyl formamide

DMPU	*N*,*N*'-dimethylpropyleneurea
DMS	dimethyl sulfide
DMSO	dimethyl sulfoxide
DNA-PK	DNA-dependent protein kinase
DNMT	DNA methyltransferase
d.r.	diastereomeric ratio
dt	doublet of triplets (NMR)
dTAG	degradation Tag (methodology)
dTMP	2'-deoxythymidine-5'-monophosphate
DTX1	deltex-1
DUB	de-ubiquitinating enzyme
ECM	extracellular matrix
EDC·HCl	*N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride
EDTA	ethylenediaminetetraacetic acid
EEA1	early endosome antigen 1
e.g.	*exempli gratia* (lat. for “for example”)
EGCG	epigallo-catechin-3-gallate
EGFR	epidermal growth factor receptor
eIF	eukaryotic translation initiation factor
ENTH	epsin N-terminal homology
ePK	eukaryotic protein kinase
eq.	equivalent(s)
ER	endoplasmic reticulum
ERα	estrogen receptor alpha
ERK	extracellular signal-regulated kinase
ESS1	essential protein 1 (in yeast)
Et_2O	diethyl ether
et al.	*et alii* (lat. for “and others”)
EtOAc	ethyl acetate
EZH2	enhancer of zeste 2 homolog 2
Fab1	1-phosphatidylinositol 3-phosphate 5-kinase
FAK	focal adhesion kinase
FAS	fatty acid synthetase
FBS	fetal bovine serum
FDA	U.S. Food and Drug Administration
FDTS	flavin-dependent thymidylate synthase
$Fe(acac)_3$	iron (III) acetylacetonate
FERM	4.1 Protein, erzin, radixin, moiesin
FKBP	FK506-binding protein
FluoPol	fluorescence polarization
Fmoc	fluorenylmethyloxycarbonyl
FOX	forkhead box
FTIR	fourier-transform infrared (spectrometer)
FtsZ	filamenting temperature-sensitive protein Z
FYVE	Fab-1, YGL023, Vps27, EEA1
GlcNAc	*N*-acetylglucosamine
GLI	glioma-associated oncogene
GLUT1	glucose transporter type 1
GLUT4	glucose transporter type 4
GMP	guanosine 5'-monophosphate
GPT	glutamate pyruvate transaminase 1

GSK3	glycogen synthase kinase 3
GTP	guanosine triphosphate
GTPase	GTP hydrolase
HAOSA	hydroxylamine-*O*-sulfonic acid
HBD	helical-bundle domain
HDAC6	histone deacetylase 6
HECT	homologous to the E6AP carboxyl terminus
Het	heteroatom
HIF-1α	hypoxia-inducible factor 1-alpha
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	hydroxybenzotriazole
HPLC	high-performance liquid chromatography
h	hour(s) (unit)
HR	homologous recombination
HRMS	high-resolution mass spectrometry
HSP90	heat shock protein 90
HTS	high-throughput screen
i.a.	*inter alia* (lat. for “among other things”)
IAP	inhibitor of apoptosis protein
IBX	2-iodoxybenzoic acid
IC_{50}	half maximal inhibitory concentration
IED	inverse electron-demand
IGF-1R	insulin-like growth factor 1 receptor
IKZF1	Ikaros
IKZF3	Aiolos
IMiD	immunomodulatory imide drug
IMPDH	inosine 5'-monophosphate dehydrogenase
ING2	inhibitor of growth protein 2
INPP4	inositol polyphosphate-4-phosphatase
INPP5	inositol polyphosphate-5-phosphatase
IP_3	inositol-1,4,5-trisphosphate
i-Pr	isopropyl
IR	infrared (spectroscopy)
IRAK4	interleukin-1 receptor-associated kinase 4
IRE1	inositol-requiring enzyme 1
IRS-1	insulin receptor substrate 1
$ITDRF_{CETSA}$	isothermal dose-response fingerprint CETSA
iTRAQ	isobaric tags for relative and absolute quantification
IUPAC	International Union of Pure and Applied Chemistry
J	coupling constant (NMR)
JAK	Janus kinase
JNK	Jun-N-terminal kinase
k_{agg}	aggregation constant (rate constant)
K_D	dissociation constant (equilibrium constant)
kDa	kilo Dalton
Keap1	Kelch like ECH-associated protein 1
KHMDS	potassium bis(trimethylsilyl)amide
KOH	potassium hydroxide
K_u	unfolding constant (here: of a protein) (equilibrium constant)
λ_{max}	maximal absorption (UV/Vis spectroscopy)
lat.	Latin

LC-MS/MS	liquid chromatography-tandem mass spectrometry
LPS	lipopolysaccharide
LYTAC	lysosome targeting chimeras
m	multiplet (NMR)
M	molar
MAO	monoamine oxidase
MAPK	mitogen-activated protein kinase
MarR	multiple antibiotic resistance repressor (transcription factor)
MATE	multidrug and toxic compound extrusion transporter
MAX	MYC associated factor X
MCL1	myeloid cell leukemia sequence 1
m-DAP	*meso*-diaminopimelic acid
MDM2	mouse double minute 2
MEK	MAPK kinase
MeOH	methanol
MetAp-2	methionine aminopeptidase 2
MFS	major facilitator superfamily transporter
MIC	minimum inhibitory concentration
MIC_{90}	minimum inhibitory concentration required to inhibit growth of 90% of organisms
min	minute(s)
miRNA	microRNA
MMR	mismatch repair
MNK	MAPK-interacting kinase
MoA	mode of action
$MraY_{AA}$	MraY isolated from *Aquifex eaolicus* strains
MreB	murein region “e” B gene
mRNA	messenger RNA
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry
MS	molecular sieve (chemical synthesis)
MTM	myotubularin
mTOR	mechanistic target of rapamycin
mTORC	mTOR complex
MudPIT	multidimensional protein modification technology
MurNAc	*N*-acetylmuramic acid
MW	molecular weight
m/z calc.	calculated mass-to-charge ratio
N_2	molecular nitrogen
Na_2CO_3	sodium carbonate
$Na_2S_2O_3$	sodium thiosulfate
Na_2SO_4	sodium sulfate
NAADP	nicotinic acid adenine dinucleotide phosphate
$NaBH_4$	sodium borohydride
$NaBH(OAc)_3$	sodium triacetoxyborohydride
$NaClO_2$	sodium chlorite
NAD	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
$NaHCO_3$	sodium bicarbonate
NaOCl	sodium hypochlorite
NBS	*N*-bromosuccinimide

n-BuLi	*n*-butyllithium
NCI	National Cancer Institute
NER	nucleotide excision repair
NF-AT	nuclear factor of activated T cells
NF-κB	nuclear factor κB
NH_3	ammonia
NH_4Cl	ammonium chloride
NIMA	never-in-mitosis, gene A
NLP	natural language processing
nm	nanometer
Nrf2	nuclear factor erythroid 2-related factor 2
ns	nanosecond
NTD	N-terminal domain
O_3	ozone
OCRL1	lowe oculocerebrorenal syndrome protein
O-GlcNAc	*O*-linked β-*N*-acetylglucosamine hydrolase
Omp	outer membrane protein
OXONE®	$2KHSO_5 \cdot KHSO_4 \cdot K_2SO_4$
P2HPT	polyprenyl-phosphate *N*-acetylhexosamine 1-phosphate transferase
p53	tumor protein 53
PAGE	polyacrylamide gel electrophoresis
PAINS	pan-assay interference compound
PAL	photoaffinity labeling
PARP	poly-(ADP)-ribose polymerase
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PDAC	pancreatic ductal adenocarcinoma
Pd/C	palladium on charcoal (10 wt.% loading)
PDE	phosphodiesterase
PE	petroleum ether
PEG	polyethylene glycol
PERK	RNA-like ER kinase
PH	Pleckstrin homology
phox	phagocytic oxidase
PI-3,4-P_2	phosphatidylinositol-3,4-bisphosphate
PI-3,5-P_2	phosphatidylinositol-3,5-bisphosphate
PI3K	phosphatidylinositol 3-kinase
PI3P5K	phosphatidylinositol 3-phosphate 5-kinase (also: PIKfyve)
PI-4,5-P_2	phosphatidylinositol-4,5-bisphosphate
PI4K	phosphatidylinositol 4-kinase
PI-4-P	phosphatidylinositol 4-phosphate
PI5P4K	phosphatidylinositol 5-phosphate 4-kinase
PIKK	PI3K-related kinase
Pin1	peptidyl-prolyl *cis*/*trans* isomerase NIMA-interacting 1
PIP	phosphatidylinositol phosphate
PIP_3	phosphatidylinositol-3,4,5-trisphosphate
PIPK	PIP kinase
PKA	protein kinase A
PK/PD	pharmacokinetic/pharmacodynamic
PLK	Polo-like kinase
PMB	*p*-methoxybenzyl
PO_3^{2-}	phosphoryl anion

POI	protein of interest
PP2A	protein phosphatase 2A
PPARγ	peroxisome proliferator activated receptor γ
PPh_3	triphenylphosphine
PPI	protein-protein interaction
PPIase	peptidyl-prolyl isomerase
ppm	parts per million
PRDM16	PR domain 16
prim.	primarily
PROTACS	proteolysis targeting chimera
PRPP	phosphoribosyl pyrophosphate
PTB	phosphotyrosine binding
ptch	patched
PtdIns	phosphatidylinositol
PTEN	phosphatase and tensin homolog
PX	phox homology
PyBOP	benzotriazole-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
q	quartet (NMR)
qNIRF	quenched near-infrared fluorescent
R&D	research and development
RAR	retinoic acid receptor
Rb	retinoblastoma protein
RBR	RING-between-RING
red.	reductive
RIES	rearrangement in the excited state
RING	really interesting new gene
RIPK2	receptor interacting serine/threonine kinase 2
RNAi	RNA interference
RNAPII	RNA polymerase II
RNase P	ribonuclease P
RND	resistance nodulation division transporter
ROS	reactive oxygen species
RPA	replication protein A
rRNA	ribosomal RNA
RT q-PCR	reverse transcription quantitative polymerase chain reaction
s	singlet (NMR)
S6K	S6 kinase
SAR	structure-activity relationship
SAE	sialic acid 9-*O*-acetylesterase
SCF	Skp1-cullin-F box
SEDS	shape, elongation, division, sporulation
SH2	Src homology 2
SHIP1/2	SH2 domain-containing inositol phosphatases 1/2
SHP2	Src homology 2 domain-containing phosphatase 2
shRNA	short hairpin RNA
SILAC	stable isotope labeling with amino acids in cell culture
siRNA	small interfering RNA
Sirt2	sirtuin 2
SKIP	Ski-interacting protein
SLFN11	Schlafen 11
SMARCA	SWI/SNF related, matrix-associated, actin-dependent regulator of chromatin A

SMILES	simplified molecular input line entry specification
smo	smoothened
SMR	small multidrug resistant transporter
SNP	single nucleotide polymorphism
SPROX	stability of proteins from rates of oxidation
sRNA	small regulatory RNA
ssDNA	single-stranded DNA
STAT	signal transducer and activator of transcription
SUMO	small ubiquitin-like modifier
SV	secretory vesicle
SYNJ	synaptojanin
t	triplet (NMR)
target ID	target identification
TB	tuberculosis
TBK1	TANK-binding kinase 1
TβL	tabtoxinine-β-lactam
TBDMS	*tert*-butyldimethylsilyl
TBST	tris buffered saline with tween
t-BuOK	potassium *tert*-butoxide
TCEP	tris(2-carboxyethyl)phosphine
Tces	2,2,2-trichloroethoxysulfonyl
TCR	T-cell receptor
TEA	triethylamine
TEMPO	2,2,6,6-tetramethylpiperidine 1-oxyl
tert-Bu	*tert*-butyl
tert-BuOH	*tert*-butanol
TFA	trifluoro acetic acid
TGF-β	transforming growth factor beta
THC	$\Delta^{8/9}$-tetrahydrocannabinol
THF	tetrahydrofuran
TLC	thin-layer chromatography
TLS	translation synthesis
T_m	melting temperature/point (here: for the unfolding of a protein)
TMEM55	transmembrane protein 55
TMS	trimethyl silane
TNF	tumor necrosis factor
TNFR	TNF receptors
TOP	tandem orthogonal proteolysis
TPIP	transmembrane phosphoinositide 3-phosphatase and tensin homolog 2
TPP	thermal proteome profiling
t_R	retention time
TRAIL	TNF-related apoptosis-inducing ligand
tRNA	transfer RNA
TrkC	tropomyosin receptor kinase C
TS	thymidylate synthase
TSA	thermal shift assay
TSP	Thermo Scientific Spectra
tt	triplet of triplets (NMR)
UDP	uridine diphosphate
UHPLC	ultra-high-performance liquid chromatography
UMP	uridine monophosphate

UppP	undecaprenyl pyrophosphate phosphatase
UppS	undecaprenyl diphosphate synthase
UV	ultraviolet
VHL	Von Hippel-Lindau (E3 ligase)
VPAC1	vasoactive intestinal peptide receptor 1
Vps27	vascuolar protein sorting-associated protein 27
VRE	vancomycin-resistant *Enterococci*
vs.	versus
WHO	World Health Organization
Wnt	wingless-related integration site
ZipA	Z interaction protein A

ABBREVIATIONS OF AMINO ACIDS

Amino Acid	Single Letter Code	Triple Letter Code
alanine	A	Ala
arginine	R	Arg
asparagine	N	Asn
aspartic acid	D	Asp
cysteine	C	Cys
glutamic acid	E	Glu
glutamine	Q	Gln
glycine	G	Gly
histidine	H	His
isoleucine	I	Ile
leucine	L	Leu
lysine	K	Lys
methionine	M	Met
phenylalanine	F	Phe
proline	P	Pro
serine	S	Ser
threonine	T	Thr
tryptophan	W	Try
tyrosine	Y	Tyr
valine	V	Val

Part I

GENERAL INTRODUCTION

1 Drug and Probe Discovery and Development

1.1 General Approaches in Drug and Probe Discovery and Development

1.1.1 How Medicinal Research Influences Human Health

Drug discovery and development processes aim to introduce a new drug for the treatment or cure of a disease. While these processes are naturally long and expensive – currently estimated to cost over $2.5 billion per newly developed drug[1] – there is a consistent need for novel therapies. Constant progress in our understanding of disease biology and the completion of successful research and development (R&D) programs broaden our options to treat or cure pathological conditions, therefore majorly improving our quality of life and life expectancy. Obtaining information on current threats to human health and survival is therefore essential in directing research and drug development efforts.

The World Health Organization (WHO) frequently reports several statistics on global causes of death. For example, in the reports of 2000 and 2016, ischaemic heart disease and stroke were not only the leading causes of death world-wide, but also showed an increase in absolute numbers of deaths over time[2] (Figure 1.1.1). Interestingly, while infectious diseases, such as lower respiratory infections, tuberculosis, and HIV/AIDS, were the second major causes of death in 2000, this was not the case in 2016[2]. Instead, the second most common causes of death posed different lung-based diseases, including chronic obstructive pulmonary disease (COPD), lower respiratory infections, and lung-based cancers[2]. Notably, age-related conditions, such as dementia and diabetes mellitus arose as some of the leading causes of death in 2016[2] (Figure 1.1.1). Although this shift from infectious diseases to metabolic and age-related conditions could be interpreted as a positive effect of improvements to modern health care systems, several areas of research focusing on pathological conditions, such as heart diseases, cancer, COPD, infectious diseases, and dementia, require further progress to deepening our understanding in their disease biology, thus enabling the development of novel treatments.

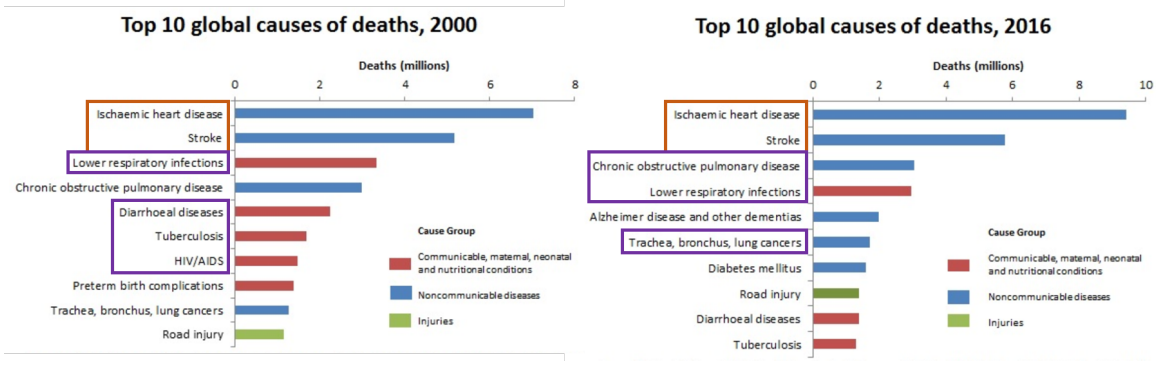


Figure 1.1.1 | Top 10 Causes of Death World-Wide. Global causes of death from 2000 (left) and 2016 (right) are compared. While the top causes of death (orange boxes) in both years were ischeamic heart disease and stroke, the second most common causes of death (purple boxes) were either infectious disease-related deaths in 2000 or lung-based diseases in 2016. Adapted from WHO, "The top 10 causes of death.", **2018**[2].

The WHO further provides statistics that present the main causes of death in relation to the wealth of a country. While Figure 1.1.1 could indicate that infectious diseases do not pose a major problem for global causes of death anymore, Figure 1.1.2 clearly shows that this is only true for high-income countries[2]. In low-income countries, infectious diseases, such as lower respiratory infections and HIV/AIDS, are still lead causes of deaths, and only in high-income countries are these effectively replaced by metabolic- or age-related diseases, such as cardiovascular conditions, dementia, cancer, diabetes mellitus, and kidney diseases[2]. In summary, this data implies that advances in health care and lifestyle improve with the wealth of a country and majorly impact the overall health of its population.

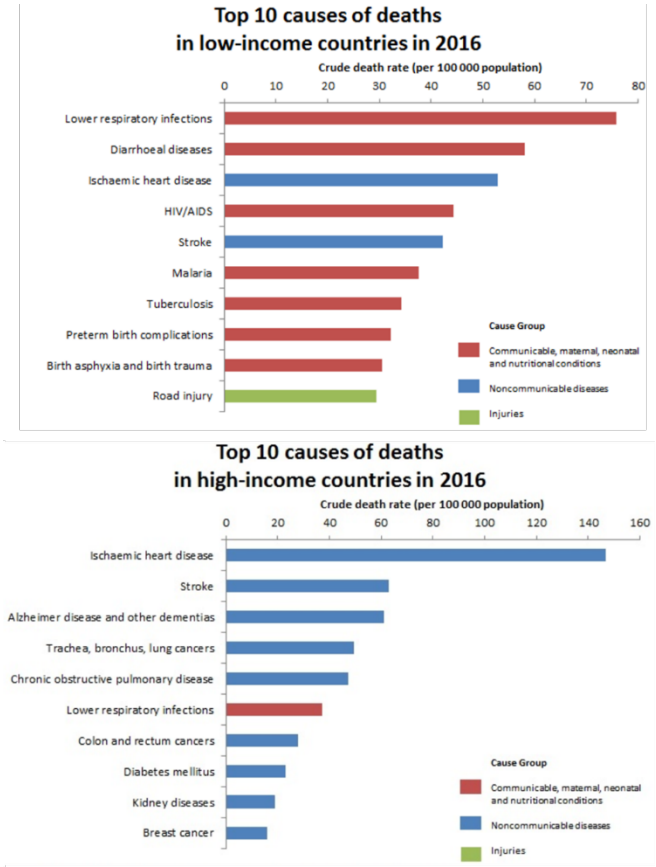


Figure 1.1.2 | Top 10 Causes of Death for Low- and High-Income Countries. Compared are the top 10 causes of death for low- (top) and high-income (bottom) countries in 2016. While low-income countries are mainly suffering from infectious disease-related deaths, high-income countries show a majority of age- and metabolic disease-related deaths. Adapted from WHO, "The top 10 causes of death.", **2018**[2].

While the presented data further implies that it is possible to shift or alter global causes of death within just 16 years – possibly due to changes in treatment availability, improvements in health care, and changes in lifestyle – it also indicates that there is great potential for current and future R&D programs to have a great impact on global health. Especially cardiovascular conditions, lung diseases, cancer, dementia-related, and infectious diseases still require a more thorough understanding of their underlying mechanisms to advance the development of novel or improved treatments.

1.1.2 From Bench to Clinic – A Short Overview

Drug discovery and development starts with the identification of disease-relevant target(s) or phenotype(s), and ends in the clinical evaluation of highly optimized and characterized drugs[1]. Generally, the pre-clinical process of drug discovery can be divided into four major stages: concept development, lead generation, lead optimization, and preclinical evaluation[3] (Figure 1.1.3). While the last stage aims to characterize a drug candidate for clinical trials, the first three stages pose the important phases of generating biologically active compounds[3]. Depending on the starting point, a drug discovery process can be target- or phenotype-based, meaning that compounds are either screened against a target of interest or for a specific phenotype, respectively[3] (Figure 1.1.3).

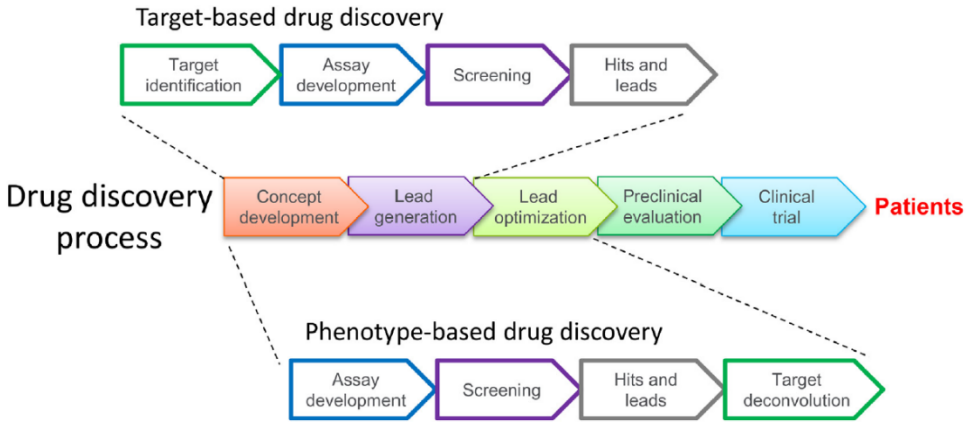


Figure 1.1.3 | Basic Concept of a Drug Discovery Process. Drug discovery and development processes are commonly initiated starting from target- or phenotype-based approaches. Taken from Kubota *et al.*, *Biochim. Biophys. Acta Proteins Proteom.* **2019**, *1867* (1), 22-27[3].

Phenotype-based approaches were the most widely applied drug discovery strategies in the past, even though the underlying pharmacological targets of screening hits are (initially) not known[3]. However, with advancements in genetic techniques and small molecule libraries, target-based approaches arose, which provide the basis for rational drug design. Thus, while phenotype-based approaches potentially offer better translation to *in vivo* models and clinical effectiveness, the majority of drug discovery campaigns currently utilize target-based approaches due to the compounds' comprehensible modes of action[3]. Although target-based approaches have therefore accelerated drug discovery, target identification has remained a problem.

In a longitudinal examination of AstraZeneca's R&D strategies in 2014, the company came up with a "5R framework" (Figure 1.1.4 A) in order to promote more efficient drug discovery and development[4]. The analysis shows that the majority of program failures were due to insufficient *in vivo* efficacy, mainly due to weak or insufficient links between the target protein of a drug candidate and effectiveness for disease biology[4] (Figure 1.1.4 B). The data highlights the importance of thorough target identification and validation studies for successful target-based drug discovery, as well as the necessity for target deconvolution after phenotypic screens.

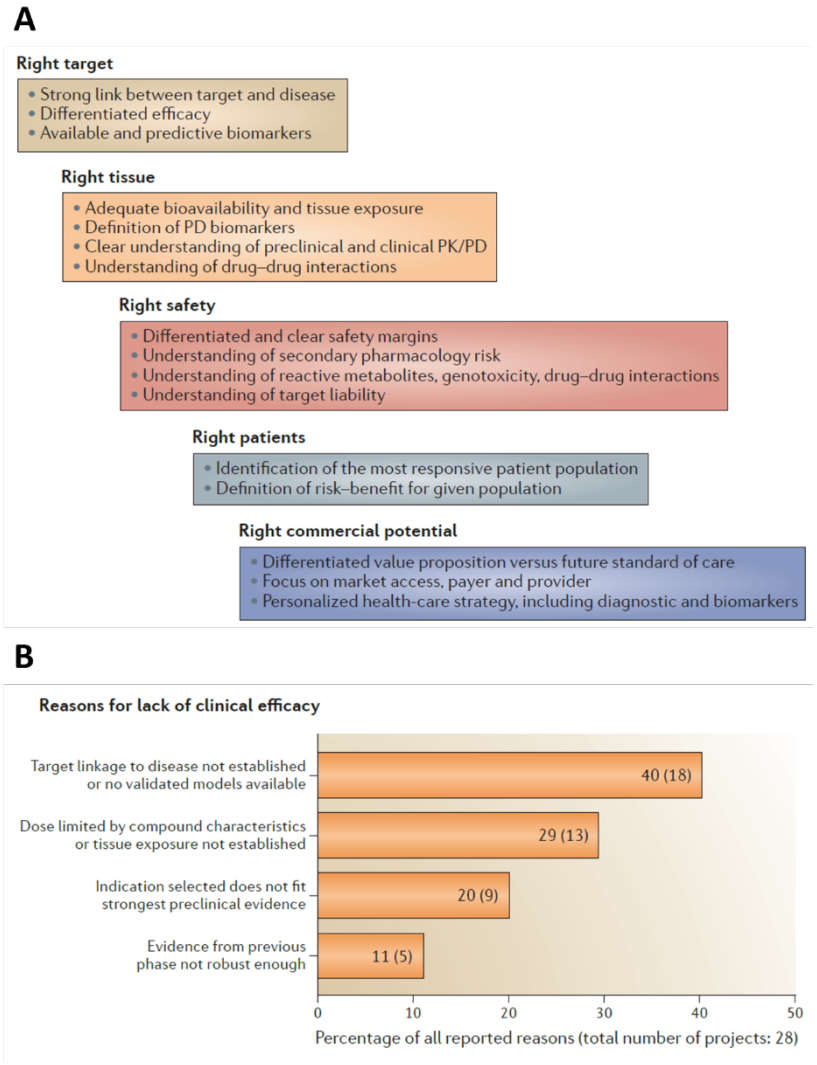


Figure 1.1.4 | Target Identification and Validation are Crucial Steps in Drug Development[a]. **A.** The "5R framework" from AstraZeneca's R&D analysis. Five important considerations were identified, based on issues that led to terminations of clinical trials in the past. **B.** Missing link between target and disease was the most common reason for failures of drug development processes.
[a] Adapted from Cook *et al.*, *Nat. Rev. Drug Discov.* **2014**, *13* (6), 419-431[4].

1.1.3 Approaches in the Drug and Probe Development and Discovery Process

Due to the general complexity of disease biology, continuous advancements in technologies and tools used for its study offer a variety of techniques to aid drug development campaigns. As previously mentioned, target-based strategies are commonly applied drug discovery approaches, which require the development or discovery of potent target binders. Hence, biochemical and cell-based activity screens build an essential foundation for testing potency and selectivity of small molecules in these processes[5].

Commonly, high-throughput screens are used as starting points in the discovery of a novel lead scaffold, which, in some cases, can be expanded to multi-target screens to define polypharmacological profiles of the screened compounds[5]. In addition, combination

screens, in which the synthetic lethality of drug combinations is investigated, as well as screens of already approved drugs, aiming for drug repurposing to accelerate the development process, both add to the drug discovery toolbox[5,6]. More recently, *in silico* screening methods, applying artificial intelligence and machine learning approaches to drug discovery, enabling virtual screens of compound libraries with billions of molecules against a target of interest[7-9]. In case of phenotype-based drug discovery, screens for desired phenotypes in cellular assays are used as starting points for the development of bioactive compounds[2].

Typically, scaffolds of potential lead compounds or probes are designed *de novo*, based on fragment-based approaches[5,10], or *via* drug optimization campaigns starting with screening hits from small molecule libraries[5], including virtual libraries[7]. Thorough structure-activity relationship (SAR) studies further guide lead optimization processes. Ideally, co-crystal structures or docking models of the target of interest in complex with the lead compound are used to elucidate the compounds' binding mode and assist in rational SAR study design[11,12]. Ultimately, after iterative cycles of compound design, *in vitro* or *in cellulo* testing, and optimization phases, the lead compound(s) with desired biochemical and/or cellular potency can be selected for further characterization. Figure 1.1.5 provides an overview of common drug (and probe) discovery and development processes.

Notably, one essential lead compound property, despite on-target potency, is selectivity. Over time, a variety of different chemical proteomic approaches have emerged to aid in the target identification process of small molecules[13], a crucial step in both target- and phenotype-based approaches. Chemical proteomics generally include global profiling techniques *via* identification of trypsin-digested peptides of cell lysates, as well as affinity- or activity-based approaches using beads for protein purification and enrichment steps[13]. These and other strategies for target identification purposes are discussed in more detail in following chapters. Finally, at the stage of hit optimization, ADME(T) (acronym for: absorption, distribution, metabolism, excretion, and toxicity) parameters of a subset of hit compounds are determined and optimized to prepare for *in vivo* application of a candidate[5].

After the discovery, optimization, and characterization steps, a small molecule lead or probe can be obtained. Briefly, while lead molecules serve as the starting point for drug development and compounds that move into clinical trials, chemical probes are used to investigate disease biology or the mode of action of a bioactive compound[14]. While polypharmacological profiles of drug candidates are often acceptable and sometimes even desired for optimal clinical efficacy, for chemical probes, compound selectivity is essential to guarantee optimal traceability of probe-specific effects[14]. That is, probes help in identifying the connection between a target and its physiological function, as well as guide translation of these links into the clinical setting[14].

Naturally, chemical probes can also serve as a starting point for the development of a potential drug candidate. While drug candidates are ideally potent and selective agents to avoid off-target effects *in vivo*, many drugs to date have actually been reported to be multi-targeting agents, highlighting the importance of polypharmacology for the clinic[15]. Because of this, phenotype-focused drug discovery processes generally have the advantage of definitively identifying bioactive agents[16]. Target-focused approaches, on the other hand, identify biochemically active molecules against a specific (purified) target, making the identification of polypharmacology difficult[16]. However, target-based screens can

detect low-weight and fragment-size molecules, which is mostly not feasible for phenotypic screens, allowing the exploration of a larger chemical space[16]. Consequently, a combination of target-based approaches and phenotypic screens poses a valuable consideration to increase the chances of a successful drug and probe development campaign[16].

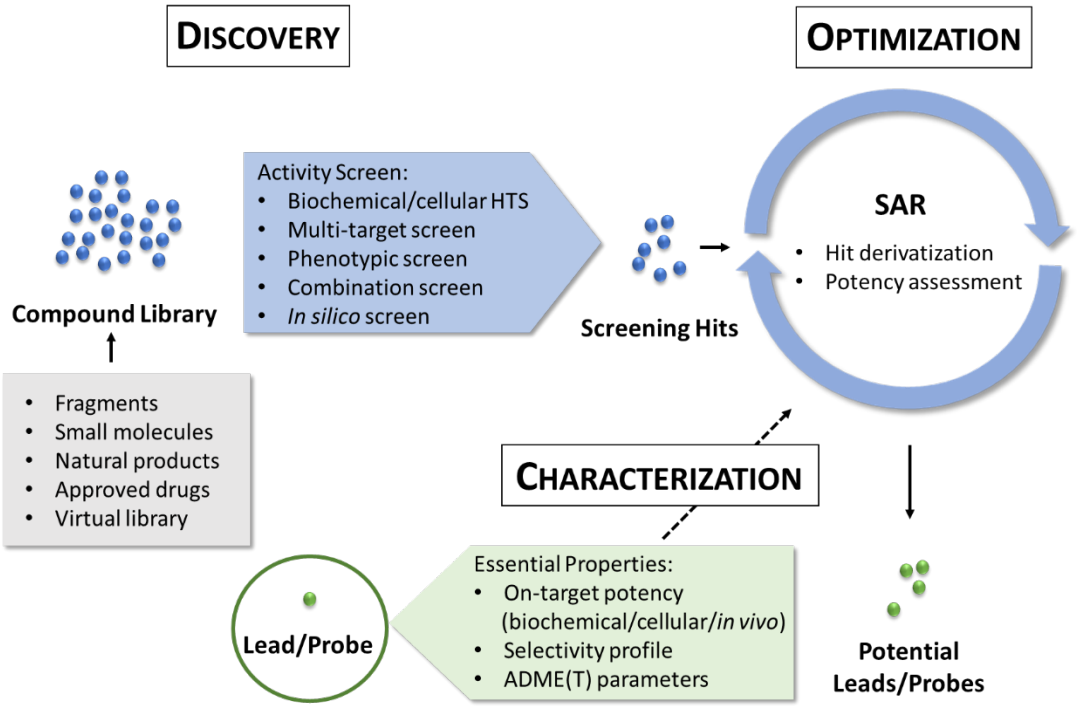


Figure 1.1.5 | General Overview of a Drug/Probe Discovery and Development Process. Hit compounds are often discovered through activity screens of compound libraries. These libraries can consist of fragments, small molecules, natural products, or approved drugs. In the case of *in silico* screens, virtual libraries of billions of compounds can be screened. Activity screens are often based on biochemical or cellular screens (HTS = high-throughput screen), but also include multi-target, phenotypic, and combination screens to uncover synthetic lethal drug combinations. Screening hits are further optimized by thorough SAR studies. Finally, selected potential leads or probes of these optimization processes are then characterized to assess cellular or *in vivo* on-target activity, selectivity profiles, and ADME(T) parameters. The results from this characterization stage can either lead to further optimization processes or to a potential drug candidate, which moves towards preparation for clinical trials, or a probe for the investigation of a target of interest.

1.2 Methods of Target Deconvolution and Validation

Confidence in the association of a biological target and disease biology positively correlates with a higher success rate in drug development processes[4]. Therefore, validation of the drug target(s) continues to be evaluated until the late stages of clinical trials or even after drug approval[17]. In contrast, target identification studies, commonly performed at the beginning of the drug discovery process, either serve to identify new targets for target-based drug discovery[17,18], or deconvolute the target(s) of a bioactive compound following phenotypic-based discovery strategies[19]. For example, identification of key players within a physiological pathway involved in disease biology or the off-target activity of a bioactive compound can lead to the discovery of new targets for potential novel treatments[17,18].

1.2.1 Techniques of Target Deconvolution

Target identification steps can either start from an observed phenotype ("top-down" approach) or with the detection of small molecule-target interactions ("bottom-up" approach)[20] (Figure 1.2.1). Briefly, target deconvolution strategies can be divided into three general methods: chemical biology (direct) approaches, genetic (indirect) approaches, and computational approaches[18] (Figure 1.2.2). While specific methods used are described within their respective chapters of this thesis, a general overview shall be given as part of this introduction.

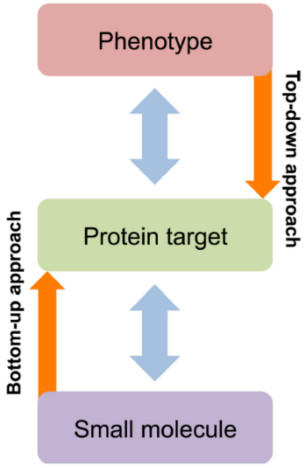


Figure 1.2.1 | "Top-down" and "Bottom-up" Approaches as Target Deconvolution Strategies. The identification of a protein target for drug and probe discovery purposes can be identified through a "top-down" approach, in which an underlying phenotype is observed and leads to the identification of a protein target. In contrast, a "bottom-up" approach is characterized by the direct identification of a protein target of a small molecule. Taken from Titov *et al.*, *Bioorg. Med. Chem.* **2012**, *20* (6), 1902-1909[20].

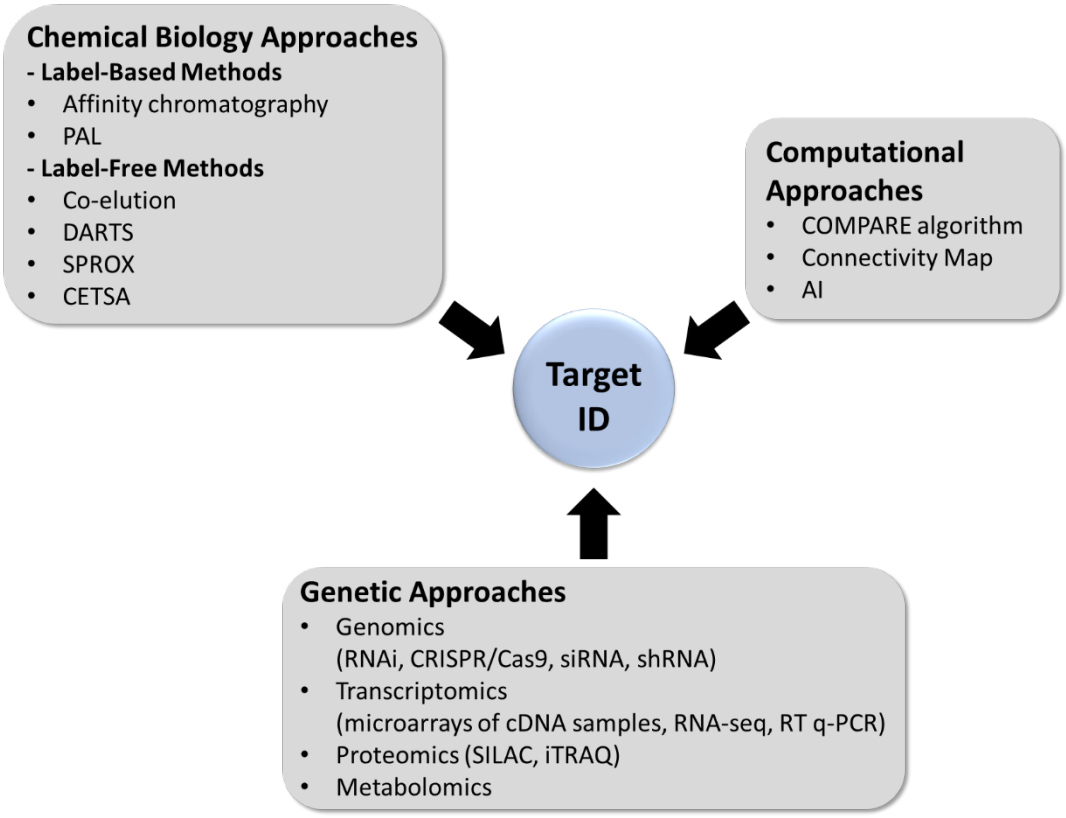


Figure 1.2.2 | Basic Approaches for Target Identification Studies. Methods to uncover target identity (target ID) of a small molecule of interest are generally divided into chemical biology, computational, and genetic approaches. Selected examples of commonly used techniques applied (further mentioned in this chapter) in drug and probe discovery and development are given.

Genetic (indirect) Approaches

In genetic approaches, a small molecule or physiological gene activity is observed before and after genetic manipulation or with and without drug treatment. Changes in phenotype or drug efficacy provide evidence for protein targets of a bioactive molecule or the physiological role of novel targets of interest for subsequent discovery processes. For example, comparative genomic methods, including sequencing of drug resistant strains (genome-wide sequencing), RNA-interference (RNAi)-based selection for drug-sensitive strains, or desensitization of cell clones with overexpressed drug target, give an indication for the mechanism of action of a small molecule[18,21]. More recent methods, such as the CRISPR/Cas9 (CRISPR standing for clustered regularly interspaced short palindromic repeats) technology, small interfering RNA (siRNA), and short hairpin RNA (shRNA), pose additional methods to investigate loss-of-function phenotypes of a protein of interest[21,22].

In contrast, transcriptomic strategies uncover changes on transcription level[21]. One advantage of these methods is the possible read-out of acute drug-treatment effects in a complex environment such as the cell[21]. For example, reverse transcription techniques, in which messenger RNA (mRNA) is transformed into complementary DNA (cDNA), which, due to its higher stability, can be utilized to reveal amplified or down-regulated gene transcripts caused by drug treatment[21]. This method can be found within microarrays of

labeled cDNA samples, RNA-seq experiments, and reverse transcription quantitative polymerase chain reaction (RT q-PCR)[21].

Similar to transcriptomic approaches, liquid chromatography-tandem mass spectrometry (LC-MS/MS) techniques used within proteomic studies provide information about changes in expression levels of proteins[21]. Isotope-labeled proteins in stable isotope labeling with amino acids in cell culture (SILAC) methods, or isobaric tags for relative and absolute quantification (iTRAQ) methods are commonly applied methods for this purpose[21]. Likewise, LC-MS/MS-based metabolomic approaches have emerged in the context of metabolic enzymes, providing valuable information about changes in metabolic substrate to product ratios[21].

Chemical Biology (direct) Approaches

Chemical biology-based target deconvolution approaches are affinity-based methods and provide information on direct interactions of small molecules with their target protein(s)[18]. Generally, chemical biology approaches are divided into direct (label-based) and indirect (label-free) strategies[21].

Affinity chromatography and photoaffinity labeling (PAL) are among the most commonly applied direct approaches and require modification of the investigated small molecule for detection or enrichment purposes[21]. For example, in affinity-based chromatography techniques, a small molecule of interest is either directly fixed on beads or tagged with biotin *via* a linker for target enrichment through pulldown with streptavidin-coated beads[23]. The resulting ligand-target protein complexes are purified, and then detected and analyzed *via* gel electrophoresis and/or mass spectrometry[23]. However, if the ligand-target interaction is weak, these targets cannot be detected with this method[21]. Thus, PAL strategies have been developed to overcome these limitations by introducing reactive warheads to the compound structure, yielding covalent ligand-target complexes, which are then analyzed in ways similar to the chromatographic approaches[21].

In contrast, label-free methods have the advantage that the investigated small molecules do not have to be chemically modified[21]. Besides the chromatographic co-elution method that detects a change in the retention time of a native drug compared to when bound to its target[21], stability-based approaches offer an invaluable strategy for target identification[24]. The underlying concept stems from the observation that ligand-target complexes exhibit different stability properties than the unbound target protein[24]. For instance, drug affinity responsive target stability (DARTS) methods recognize differences in proteolytic stability of ligand-target complexes, while stability of proteins from rates of oxidation (SPROX) methods detect changes in methionine oxidation levels, which correlate with thermodynamic changes within the ligand-target complex[24]. Finally, cellular thermal shift assays (CETSAs) analyze thermodynamic stability of a target of interest incubated with a potential ligand in cell lysate or intact cells[24]. Briefly, melting curves of a target of interest are compared to compound treated samples, which show a curve-shift due to thermodynamic stabilization of the complex upon ligand binding[24]. Combining the CETSA approach with mass spectrometry analysis elevates this technology to a proteome-wide profiling tool[24].

Similar to genetic approaches, affinity-based target identification methods and CETSA methods can be applied as chemical proteomic tools in the context of activity-based protein profiling (ABPP)[25] and thermal proteome profiling (TPP)[24], respectively. Furthermore,

chemical genetic knock-out strategies have emerged with the introduction of protein degrader molecules[26]. Here, small molecule-mediated protein-protein interactions (PPIs) between a target of interest and E3 ubiquitin ligases induce protein degradation *via* the cellular proteasomal machinery[27]. The advantage of this method is an immediate and long-term effect on the target of interest mediated by the degrader molecule[28].

Computational Approaches

Growing databases and collected knowledge on well-characterized compounds and their modes of action allow computer-based prediction tools to help identify potential targets of a molecule of interest[18]. For example, the COMPARE analysis algorithm from the National Cancer Institute (NCI) was derived from comparing toxicity profiles of different compounds, and the Connectivity Map database from the Broad Institute is the result of whole-genome transcript profiling efforts[29].

Naturally, there are several *in silico* methods to model, dock, and evaluate ligand binding using available protein structures or homology models[30]. More recently, artificial intelligence (AI) approaches have entered drug discovery processes[31]. Briefly, machine learning algorithms, including deep learning strategies, use neural network models to train and develop algorithms based on existing data sets[31]. After validation of the resulting algorithm on a test set, predictions for novel data sets can be made[31]. This thus opens up an emerging field within the drug discovery process, in which prediction tools for target identification purposes are being developed[31]. Some of these algorithms interrogate the chemical space of protein ligands by applying methods such as natural language processing (NLP) to simplified molecular input line entry specification (SMILES) of molecules as input data to generate predictions of novel scaffolds[8].

1.2.2 Techniques of Target Validation

Target deconvolution techniques primarily aim to identify targets of a small molecule or proteins causing a specific phenotype[17]. Target validation approaches, however, function as selection tools for choosing appropriate targets for the development of a treatment for a disease, meaning, whether targeting a specific protein actually has the desired effects within the complex environment of a cell or organism[17].

Most techniques for early target validation purposes are genetic approaches and interrogate changes in the mRNA or protein levels within a certain disease model or cellular system[17]. Moreover, knockout and transgenic animal models are highly useful tools in investigating disease biology and verifying a hypothesized target *in vivo*[17]. As this is an important step in the drug development process, it highlights the advantage of phenotype-based discovery processes, since these already include preliminary validation of compound efficacy *in cellulo* or *in vivo* early-on in the discovery process. Notably, as mentioned before, target validation is a process that stretches through all drug and probe development and discovery steps[17].

Figure 1.2.3 gives an overview of further methods used to validate drug targets[32]. Naturally, the target has to be druggable, meaning that a small molecule needs to be able to bind and modulate the target of interest[32]. Chemical probes are often used to investigate cellular and *in vivo* effects of target manipulation, while desired phenotypes should be reproducible in an animal model at acceptable doses[32]. Manipulating a target of interest should further be safe and ideally supported by a link between disease biology and target

gene (expression)[32]. For this, biological tools are highly helpful in creating disease models and genetic studies of a target of interest[32]. Finally, the effects of a potential drug candidate need to be validated in the clinical setting[32].

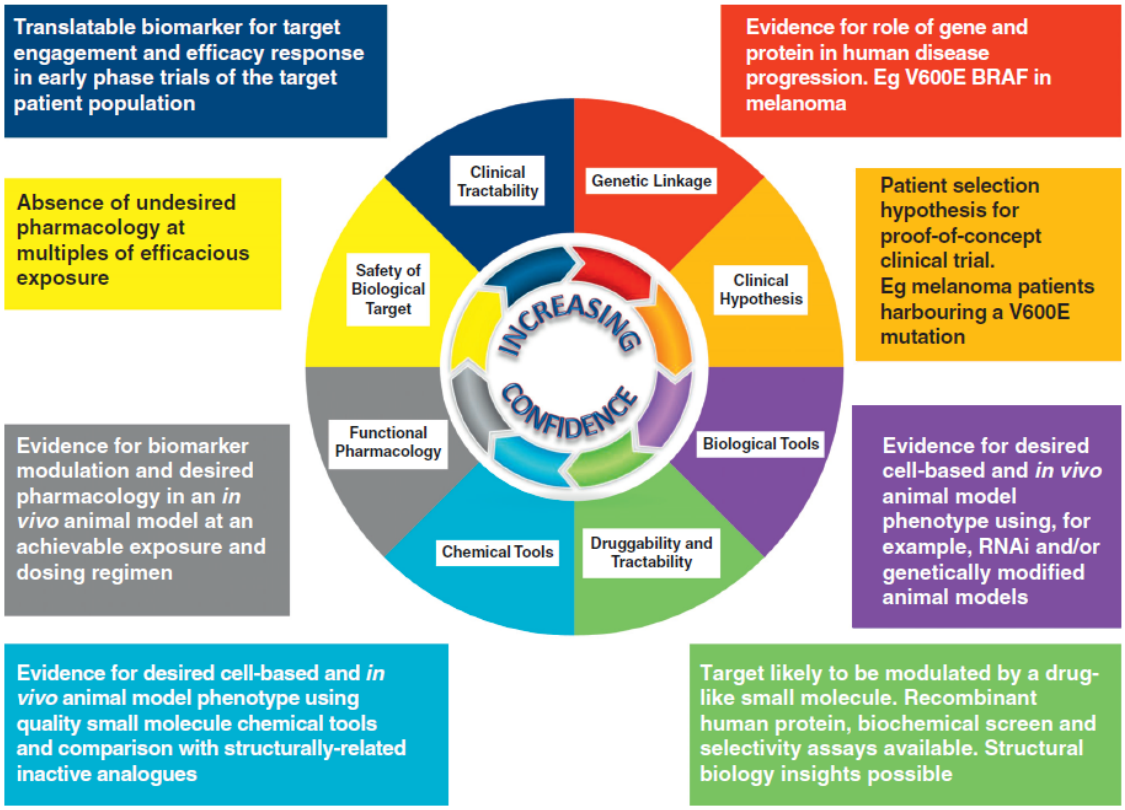


Figure 1.2.3 | Approaches to Target Validation. Target validation methods begin in early stages of the drug discovery process with establishing a strong link between target gene and disease biology *via* biological and chemical tools and continue on through investigations of a target of interest in *in vivo* models, as well as the clinical setting, including patients. Taken from Blagg *et al*., *Curr. Opin. Pharmacol.* **2014**, *17*, 87-100[32].

The presented examples of target deconvolution and validation strategies and techniques emphasize the importance of a strong connection between disease and biological target for the drug discovery and development process. However, the variety and richness in technologies and methods available further highlight that there is no superior strategy to drug discovery. Depending on the disease biology, target structure, chemical probe availability, lead compound scaffold properties and other factors, combining different techniques and strategies is necessary to gain the desired confidence in a drug target in the context of a novel treatment.

2 The Development and Application of Chemical Probes

2.1 Medicinal Chemistry and Chemical Biology Orchestrate Drug Development

While some literature suggests that "Medicinal Chemistry" began with the introduction of the drug receptor theory by John Newport Langley and Paul Ehrlich at the end of the 19th century and the beginning of the 20th century[33,34], other sources argue that the field was initially mentioned after World War II, when first comprehensive SAR studies were performed[35]. In 1998, the International Union of Pure and Applied Chemistry (IUPAC) ultimately defined "Medicinal Chemistry" as a discipline that combines the fields of "biological, medical and pharmaceutical sciences" with chemistry, with the intent to "[invent, discover, design, identify, and prepare] biologically active compounds", including "the study of their metabolism, the interpretation of their mode of action and the construction of [SARs]"[36]. As a highly interdisciplinary approach (Figure 2.1.1)[35], medicinal chemistry plays thus an essential part in the development of novel therapeutics.

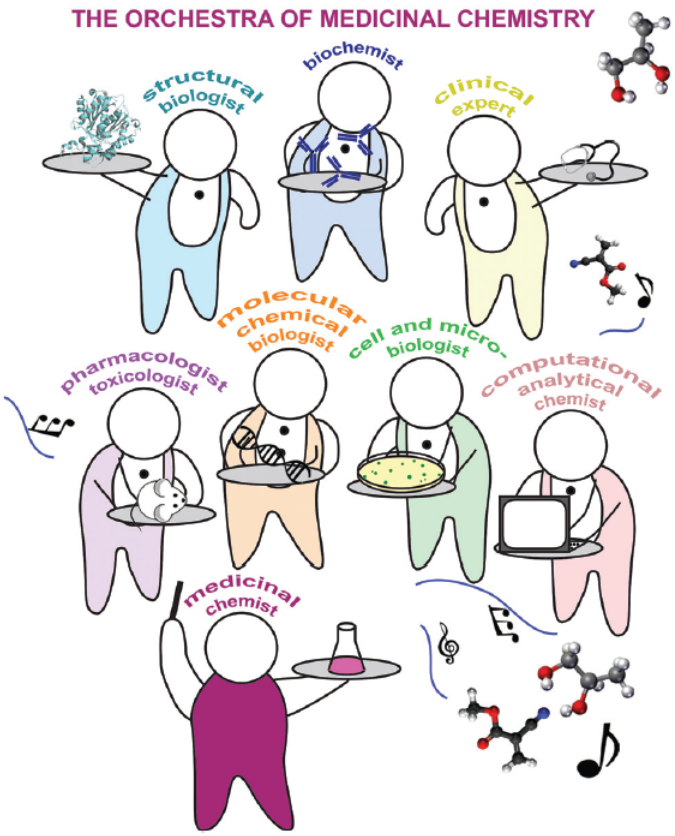


Figure 2.1.1 | "The Orchestra of Medicinal Chemistry". Cartoon of a medicinal chemist orchestrating a diverse range of different disciplines with the aim of creating, optimizing, and characterizing novel bioactive molecules. Taken from Holbrook *et al.*, *Med. Chem. Commun.* **2017**, *8*, 1739-1741[35].

In recent years, chemical biology emerged as another important field in drug discovery, interrogating biological systems with chemical tools and probes[37]. Consequently, one of the main goals of chemical biology is target identification and validation[38,39]. Thus, medicinal chemistry and chemical biology approaches go hand in hand in early drug discovery, with medicinal chemistry enabling drug design, synthesis, and optimization to generate potential drug candidates and probes, which in turn serve as tools to interrogate disease biology in chemical biology studies[37]. Hence, at the heart of both disciplines are small molecules as chemical tools or probes. In contrast to genetic approaches, chemical probes have the advantage of closely mimicking effects of potential novel medicines on cellular or whole organism level[37]. As shown in Chapter 1, this is particularly important for drug development, since efficacy issues in patients is a prominent reason for failure of drug candidates in the clinic[4].

Development of novel technologies and methods, including the appearance of interdisciplinary fields such as chemical biology, has immensely changed the face of medicinal chemistry as a discipline over the years[40]. However, the underlying mission of medicinal chemistry campaigns to advance drug discovery and development processes stays the same, resulting in an ever-evolving field that unites diverse disciplines to efficiently develop therapeutics[40].

2.2 Chemical Probes – Useful Tools in the Drug Discovery and Development Process

2.2.1 Development of Chemical Probes

Historically, chemical probes were developed based on natural products or by structure-based designs in the case of target-focused approaches[41] (Figure 2.2.1). In recent years, screening technologies have changed the landscape of probe discovery and now pose important sources for small molecule probes[41], especially in the context of high-throughput platforms[42]. Since the main purpose of chemical probes is to investigate a target of interest, a defined mode of action and selectivity profile are essential[43]. Several examples of applying “low quality” probes in drug development processes, including promiscuous or “frequent hit” compounds, such as pan-assay interference compounds (PAINS), led to false understandings of disease biology, which underlines the importance of applying well-characterized probes[44].

While high-quality ligands of a target of interest can already exhibit sufficient properties to allow investigation of target function and phenotype, labeled ligand analogs increase the chemical biology toolbox and are commonly used to further analyze cellular target engagement and selectivity profiles of a bioactive molecule, as well as cellular localization of a target of interest or probe[43,45]. Respective labels serve imaging or affinity purposes and consist of either detection tags, such as fluorophores or radio-labeled agents, or affinity ligands, such as biotin[45]. In case of bulky reporter groups, which can cause potential cellular uptake issues, probes can be modified with a bioorthogonal handle, which allows for the introduction of a tag after ligand binding[45]. In addition, the formation of covalent ligand-target complexes can be achieved by the introduction of electrophiles or photoactive warheads into a probe scaffold, further broadening the options for target detection, enrichment, and purification methods[46].

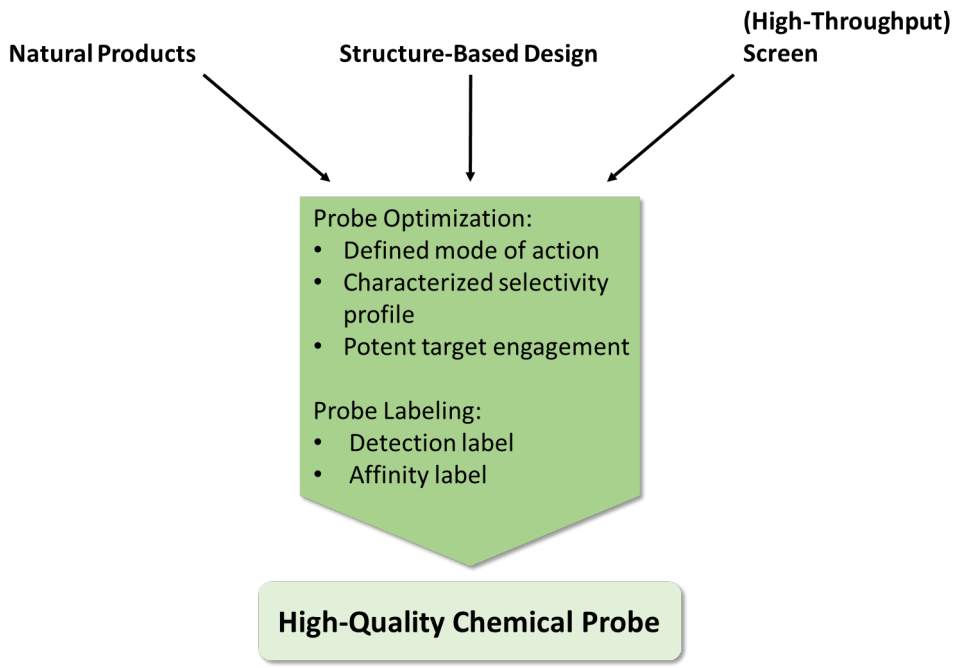


Figure 2.2.1 | Development of a Chemical Probe. Chemical probes are generally derived from natural products, structure-based designs, or screening hits, further optimized, and, in some cases, labeled with a tag.

2.2.2 Applications of Chemical Probes in Drug Discovery and Development

To study complex biological systems, chemical probes and their effects on a cell, tissue, or organism are essential tools in modern research and drug discovery[37] (Figure 2.2.2). While chemical probes reveal cellular effects often immediately and reversibly, complementary genetic approaches are commonly analyzed after days and cause irreversible changes to a system, which highlights the advantages of applying chemical probes[37]. Furthermore, small molecule probes generally bind to a specific subunit or binding site of a protein of interest and may be used to specifically study the catalytic function of the target protein[37]. Thus, within phenotypic screens, chemical probes can be used to uncover novel target proteins in the context of disease biology[37] or interrogate and characterize a known target in the course of target deconvolution and validation studies[32,39].

Besides reversible ligands (discussed in Chapter 5), recent studies introduced covalent probes (discussed in Chapters 5, 6, 7, and 9) or small molecule-based degraders (discussed in Chapter 11) for immediate but long-term modulation of a target of interest[47,48]. Covalent small molecule binders further have the advantage that otherwise "undruggable" targets can be engaged by labeling target-specific nucleophilic structures, such as conserved cysteines[49], which in other cases are targeted to improve the selectivity profile of a probe[46,50]. Similarly, degrader probes tune the selectivity profile of a probe[27], as well as offer new opportunities to modulate otherwise challenging protein targets[28]. Hence, probes for targeted protein degradation have added valuable tools for the investigation of biological systems in drug discovery, while not only targeting the catalytic function of a protein but also potential scaffolding functions[51].

The described unmodified ligands not only serve as probes for the study of a target of interest but further pose potential starting points for drug development. However, initially

identified small molecule ligands are often labeled to enable further characterization and interrogation of a probe's or target protein's modes of action[45]. For example, ABPP strategies apply tagged covalent probes (discussed in Chapters 5, 6, 7, and 9) or photoaffinity probes (discussed in Chapter 10) for target deconvolution and validation purposes, as well as drug screens[25]. Photoaffinity labeling of a probe is a useful technique to form covalent ligand-target complexes for further characterization or labeling[45].

Through ABPP functionalization of a probe, fluorescent tags, radio labels, or affinity reporter groups are introduced and aid in a vast variety of different target and probe characterization processes, including imaging, selectivity profiling, target identification, and engagement studies[45]. Respective probes have hence been applied to chemical proteomics[52] and drug discovery processes[45], but are also useful tools in disease diagnostics and monitoring[53].

Specific applications of chemical probes used within this thesis are discussed in detail in the respective following chapters of the publication section and appendix.

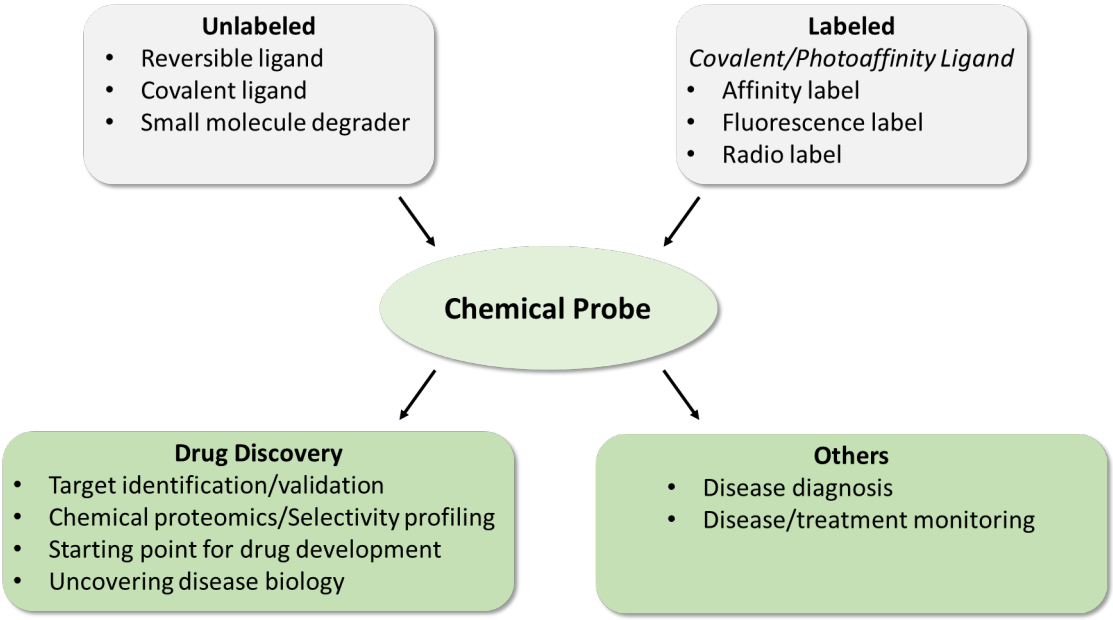


Figure 2.2.2 | Applications of Chemical Probes. Labeled and unlabeled chemical probes are widely used in the drug discovery process to uncover disease biology, identify and validate target proteins, define selectivity profiles of small molecules through chemical proteomics, or are starting points for drug development processes. Other applications of chemical probes include diagnostics and monitoring of diseases and treatments.

3 References

[1] Plenge, R. M. Disciplined approach to drug discovery and early development. *Sci. Transl. Med.* **2016**, *8* (349), 349ps15.

[2] World Health Organization. "The top 10 causes of death.", **2018**, https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. (last checked: 04/03/2020).

[3] Kubota, K.; Funabashi, M.; Ogura, Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. *Biochim. Biophys. Acta Proteins Proteom.* **2019**, *1867* (1), 22-27.

[4] Cook, D.; Brown, D.; Alexander, R.; March, R.; Morgan, P.; Satterthwaite, G.; Pangalos, M. N. Lessons learned from the fate of AstaZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **2014**, *13* (6), 419-431.

[5] Roy, A. Early Probe and Drug Discovery in Academia: A Minireview. *High Throughput* **2018**, *10* (4), 383-394.

[6] Parvathaneni, V.; Kulkarni, N. S.; Muth, A.; Gupta, V. Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discov. Today* **2019**, *24* (10), 2076-2085.

[7] Lavecchia, A. Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discov. Today* **2019**, *24* (10), 2017-2032.

[8] Öztürk, H.; Özgür, A.; Schwaller, P.; Laino, T.; Ozkirimli, E. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discov. Today* **2020**, *25* (4), 689-705.

[9] Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **2019**, *20* (11), 2783.

[10] Erlanson, D. A.; Davis, B. J.; Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chem. Biol.* **2019**, *26* (1), 9-15.

[11] Zheng, H.; Hou, J.; Zimmerman, M. D.; Wlodawer, A.; Minor, W. The future of crystallography in drug discovery. *Expert Opin. Drug Discov.* **2014**, *9* (2), 125-137.

[12] Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current Progress in Structure-Based Rational Drug Design Marks a New Mindset in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2013**, *5* (6), e201302011.

[13] Bantscheff, M.; Drewes, G. Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg. Med. Chem.* **2012**, *20* (6), 1973-1978.

[14] Garbaccio, R. M.; Parmee, E. R. The Impact of Chemical Probes in Drug Discovery: A Pharmaceutical Industry Perspective. *Cell Chem. Biol.* **2016**, *23* (1), 10-17.

[15] Reddy, A. S.; Zhang, S. Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* **2013**, *6* (1), 41-47.

[16] Heilker, R.; Lessel, U.; Bischoff, D. The power of combining phenotypic and target-focused drug discovery. *Drug Discov. Today* **2019**, *24* (2), 526-532.

[17] Hooft van Huijsduijnen, R.; Rommel, C. Decompartmentalizing target validation – thinking outside the *pipeline* boxes. *J. Mol. Med.* **2006**, *84* (10), 802-813.

[18] Schenone, M.; Dančík, V.; Wagner, B. K.; Clemons, P. A. Target identification and mechanisms of action in chemical biology and drug discovery. *Nat. Chem. Biol.* **2013**, *9* (4), 232-240.

[19] Ziegler, S.; Pries, V.; Hedberg, C.; Waldmann, H. Target Identification for Small Bioactive Molecules: Finding the Needle in the Haystack. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (10), 2744-2792.

[20] Titov, D. V.; Liu, J. O. Identification and validation of protein targets of bioactive small molecules. *Bioorg. Med. Chem.* **2012**, *20* (6), 1902-1909.

[21] Tulloch, L. B.; Menzies, S. K.; Coron, R. P.; Roberts, M. D.; Florence, G. J.; Smith, T. K. Direct and Indirect Approaches to Identify Drug Modes of Action. *IUBMB Life* **2018**, *70* (1), 9-22.

[22] Schirle, M.; Jenkins, J. L. Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discov. Today* **2016**, *21* (1), 82-89.

[23] Lee, H.; Lee, J. W. Target identification for biologically active small molecules using chemical biology approaches. *Arch. Pharm. Res.* **2016**, *39* (9), 1193-1201.

[24] Chang, J.; Kim, Y.; Kwon, H. J. Advances in identification and validation of protein targets of natural products without chemical modification. *Nat. Prod. Rep.* **2016**, *33* (5), 719-730.

[25] Xu, J.; Li, X.; Ding, K.; Li, Z. Applications of Activity-Based Protein Profiling (ABPP) and Bioimaging in Drug Discovery. *Chem. Asian J.* **2020**, *15* (1), 34-41.

[26] Cromm, P. M.; Crews, C. M. Targeted Protein Degradation: from Chemical Biology to Drug Discovery. *Cell Chem. Biol.* **2017**, *24* (9), 1181-1190.

[27] Chopra, R.; Sadok, A.; Collins, I. A critical evaluation of the approaches to targeted protein degradation for drug discovery. *Drug Discov. Today Technol.* **2019**, *31*, 5-13.

[28] Gu, S.; Cui, D.; Chen, X.; Xiong, X.; Zhao, Y. PROTACs: An Emerging Targeting Technique for Protein Degradation in Drug Discovery. *BioEssays* **2018**, *40* (4), e1700247.

[29] Tashiro, E.; Imoto, M. Target identification of bioactive compounds. *Bioorg. Med. Chem.* **2012**, *20* (6), 1910-1921.

[30] Nero, T. L.; Parker, M. W.; Morton, C. J. Protein structure and computational drug discovery. *Biochem. Soc. Trans.* **2018**, *46* (5), 1367-1379.

[31] Lavecchia, A. Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discov. Today* **2019**, *24* (10), 2017-2032.

[32] Blagg, J.; Workman, P. Chemical biology approaches to target validation in cancer. *Curr. Opin. Pharmacol.* **2014**, *17*, 87-100.

[33] Maehle, A.-H.; Prüll, C.-R.; Halliwell, R. F. The emergence of the drug receptor theory. *Nat. Rev. Drug Discov.* **2002**, *1* (8), 637-641.

[34] Khan, M. O. F..; Deimling, M. J.; Philip, A. Medicinal Chemistry and the Pharmacy Curriculum. *Am. J. Pharm. Educ.* **2011**, *75* (8), 161.

[35] Holbrook, S. Y. L.; Garneau-Tsodikova, S. What is medicinal chemistry? – Demystifying a rapidly evolving discipline! *MedChemComm.* **2017**, *8*, 1739-1741.

[36] Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of terms used in medicinal chemistry. *Pure & Appl. Chem.* **1998**, *70* (5), 1129-1143.

[37] Plowright, A. T.; Ottmann, C.; Arkin, M.; Auberson, Y. P.; Timmerman, H.; Waldmann, H. Joining Forces: The Chemical Biology-Medicinal Chemistry Continuum. *Cell Chem. Biol.* **2017**, *24* (9), 1058-1065.

[38] Zhu, Y.; Xiao, T.; Lei, S.; Zhou, F.; Wang, M.-W. Application of chemical biology in target identification and drug discovery. *Arch. Pharm. Res.* **2015**, *38* (9), 1642-1650.

[39] Jung, H. J.; Kwon, H. J. Target deconvolution of bioactive small molecules: the heart of chemical biology and drug discovery. *Arch. Pharm. Res.* **2015**, *38* (9), 1627-1641.

[40] Campbell, I. B.; Macdonald, S. J. F.; Procopiou, P. A. Medicinal chemistry in drug discovery in big pharma: past, present and future. *Drug Discov. Today* **2018**, *23* (2), 219-234.

[41] Gerry, C. J.; Schreiber, S. L. Chemical probes and drug leads from advances in synthetic planning and methodology. *Nat. Rev. Drug Discov.* **2018**, *17* (5), 333-352.

[42] Inglese, J.; Johnson, R. L.; Simeonov, A.; Xia, M.; Zheng, W.; Austin, C. P.; Auld, D. S. High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol.* **2007**, *3* (8), 466-479.

[43] Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J.; Bountra, C.; Brennan, P E.; Brown, P. J.; Bunnage, M. E.; Buser-Doepner, C.; Campbell, R. M.; Carter, A. J.; Cohen, P.; Copeland, R. A.; Cravatt, B.; Dahlin, J. L.; Dhanak, D.; Edwards, A. M.; Frederiksen, M.; Frye, S. V.; Gray, N.; Grimshaw, C. E.; Hepworth, D.; Howe, T.; Huber, K. V. M.; Jin, J.; Knapp, S.; Kotz, J. D.; Kruger, R. G.; Lowe, D.; Mader, M. M.; Marsden, B.; Mueller-Fahmow, A.; Müller, S.; O'Hagan, R. C.; Overington, J. P.; Owen, D. R.; Rosenberg, S. H.; Roth, B.; Ross, R.; Schapira, M.; Schreiber, S. L.; Shoichet, B.; Sundström, M.; Superti-Furga, G.; Taunton, J.; Toledo-Sherman, L.; Walpole, C.; Walters, M. A.; Willson, T. M.; Workman, P.; Young, R. N.; Zuercher, W. J. The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015**, *11* (8), 536-541.

[44] Blagg, J.; Workman, P. Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell* **2017**, *32* (2), 268-270.

[45] Prevet, H.; Collins, I. Labelled chemical probes for demonstrating direct target engagement in living systems. *Future Med. Chem.* **2019**, *11* (10), 1195-1224.

[46] Dalton, S. E.; Campos, S. Covalent Small Molecules as Enabling Platforms for Drug Discovery. *ChemBioChem.* **2020**, *21* (8), 1080-1100.

[47] Ghosh, A. K.; Samanta, I.; Mondal, A.; Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem.* **2019**, *14* (9), 889-906.

[48] Maniaci, C.; Ciulli, A. Bifunctional chemical probes inducing protein-protein interactions. *Curr. Opin. Chem. Biol.* **2019**, *52*, 145-156.

[49] Maurais, A. J.; Weerapana, E. Reactive-cysteine profiling for drug discovery. *Curr. Opin. Chem. Biol.* **2019**, *50*, 29-36.

[50] Mah, R.; Thomas, J. R.; Shafer, C. M. Drug discovery considerations in the development of covalent inhibitors. *Bioorg. Med. Chem. Lett.* **2014**, *24* (1), 33-39.

[51] Scheepstra, M.; Hekking, K. F. W.; van Hijfte, L.; Folmer, R. H. A. Bivalent Ligands for Protein Degradation in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 160-176.

[52] Zhang, H.; Yan, W.; Aebersold, R. Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 66-75.

[53] Gao, M.; Yu, F.; Lv, C.; Choo, J.; Chen, L. Fluorescent chemical probes for accurate tumor diagnosis and targeting therapy. *Chem. Soc. Rev.* **2017**, *46* (8), 2237-2271.

Part II

PUBLICATIONS

4 Aims and Scope

Medicinal chemistry and chemical biology tools are essential parts of drug development and discovery processes. Especially during early steps in the development of a novel therapeutic, chemical probes play a critical part in target identification and validation, can help to further study disease biology, and pose potential starting points for the development of new small molecule drugs. Considering the complexity of human physiology in health and disease, continual advancement of improved chemical and biological tools for the investigation of cellular processes is needed. While there are numerous research campaigns for various diseases and physiological mechanisms, cancerous and infectious diseases still pose prominent threats to human health.

In this thesis, the development of chemical probes as part of drug discovery and development is described. The presented compounds were developed with the aim to provide useful chemical tools for further investigations of a target's or inhibitor's mode of action. To achieve this goal, medicinal chemistry techniques and tools from chemical biology were married to discover, optimize, and characterize novel tool compounds.

In Chapter 5, chemical probes for the study of phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) – a lipid kinase subfamily, whose cellular modes of action in health and disease are still poorly understood due to a lack of available, potent, and selective kinase binders – are introduced. In the first part, the application of a desthiobiotin-labeled probe, derived from an in-house developed, covalent PI5P4K binder, is shown (Publication A; Chapter 5.2). Here, the primary aim in context of this thesis was the synthesis of the labeled probe to support characterization efforts of the covalent parent compound, which could then be applied to investigate PI5P4Ks involvement in autophagy.

In the second and third parts, medicinal chemistry campaigns aiming to improve the presented covalent chemical probe of Chapter 5.2 (Publication B; Chapter 5.3) as well as to develop a novel reversible PI5P4K inhibitor based on a screening hit (Publication C; Chapter 5.4) are presented. These studies include the rational design and synthesis of small compound libraries for the performance of SAR studies, as well as the application of chemical biology tools to investigate the compounds' efficacies in cell-based assays. This required the establishment of a CETSA set-up for the testing of selected subsets of compounds for *in cellulo* analysis. The publications also include crystallographic and mass spectrometry-based protein labeling analysis, as well as kinome-wide selectivity profiling for further characterization of the developed PI5P4K-binding probes.

In Chapter 6, an example of the development of a biotin-labeled covalent kinase inhibitor is given (Publication D). Based on an in-house developed, covalent cyclin-dependent kinase 12/13 (CDK12/13) inhibitor, the aim was to design and synthesize a biotin-labeled probe, which could then be used for further analysis and characterization of this CDK-inhibitor scaffold and confirmed its selectivity towards CDK12 and CDK13 in the cellular context. Because of the thorough characterization of the CDK12/13 inhibitor, the compound could be applied to elucidate CDK12/13's function in transcription regulation.

Concluding the main part of this thesis, the huge benefits of developing a desthiobiotin-labeled chemical probe based off a covalent protein inhibitor is shown for an irreversible inhibitor of the peptidyl-prolyl *cis*/*trans* isomerase never-in-mitosis, gene A (NIMA)-interacting 1 (Pin1) protein in Chapter 7 (Publication E). While there is a lot of literature found on potential Pin1 functions, poor quality chemical probes led to misunderstandings

in the interpretation of Pin1's modes of action in the past, further highlighting the importance of high-quality chemical probes for proper target validation studies. Here, the goal in context of this thesis was to support the chemistry efforts within the probe development, as well as to aid with the biochemical and cellular characterization of this covalent Pin1 inhibitor.

While the presented published studies already cover a great variety of application strategies for a diverse set of chemical probes, including biotin/desthiobiotin-labeled, or reversible and covalent unlabeled protein inhibitors, as well as label-free *in cellulo* characterization methods such as CETSA, further ongoing studies are presented in the appendix (Chapters 9-11).

In Chapter 9, the development of another desthiobiotin-based probe of a covalent, fragment-derived Pin1 inhibitor is described (pre-print; the article has been *accepted* in *Nature Chemical Biology* while this thesis was reviewed). For this project, a desthiobiotin-labeled version of the covalent parent compound was designed, synthesized, and its Pin1-labeling activity was then characterized, applying the streptavidin-coated bead-based pulldown technology. The developed probe proved highly valuable in the validation of the fragment as a selective Pin1 inhibitor, including its efficacy *in vivo*, which ultimately made it possible to study Pin1 function in cancer development.

In Chapter 10, the development of a photo-affinity probe for the study of nucleoside-derived antibiotics is presented. To accomplish this, the scaffold of a group of natural product-based antibiotics, the muraymycins, was modified with a photo-affinity label, leading to a potentially useful tool to investigate the mechanism by which these nucleoside antibiotics inhibit their known bacterial target MraY in the context of infectious diseases, as well as to uncover potential human off-targets. In addition, a reference probe without the photo-affinity label was synthesized, which was essential to confirm activity of the designed compound on MraY, providing *proof of concept* for this project.

In final Chapter 11, another kind of protein modulator strategy is introduced. Here, attempted strategies in the development of bivalent protein degraders, which are based off small molecule inhibitors for PI5P4K and Pin1, are described. This relatively novel technique of target modulation through degradation provides an interesting mode of action for probing enzyme function, including the development of novel therapeutics.

In summary, the main aim of this thesis was the investigation of the development, characterization, and subsequent application of different chemical probes in modern drug development and discovery. Ranging from labeled and unlabeled, covalent and reversible inhibitors, to protein degraders and photo-affinity probes, the presented studies utilize techniques from medicinal chemistry and chemical biology to optimize and characterize the discussed probes, which further enabled confirmation of their suitability for the study of their protein targets or as novel inhibitor scaffolds. Eventually, the characterized bioactive compounds could then be applied to interrogate their targets' function in the cellular system and *in vivo*, providing valuable information about disease biology and respective target validation.

5 Discovery and Development of Chemical Probes for the Study of Lipid Kinases

5.1 Introduction

5.1.1 Label-based and Label-free Methods for Target Identification Studies in Drug Development

During the development of small molecule protein binders as chemical probes or potential drug candidates, target identification studies are critical for the deconvolution of the compounds' mode of action. While general chemical biology methods for target identification were discussed in the General Introduction (Chapter 1.2), in this chapter, the application of label-based strategies (e.g. biotin- or desthiobiotin-labeled probes) and label-free techniques (such as CETSA) is discussed in detail in the context of developing both reversible and covalent chemical probes.

Biotin- and Desthiobiotin-labeled Probes for Target Identification Studies
Biotin-tagged covalent small molecules are part of the so-called activity-based probes (ABPs), which, together with affinity-based probes (A*f*BPs), are important chemical tools for ABPP strategies in drug discovery[1]. While ABPs intrinsically form covalent ligand-target complexes, A*f*BPs include a photoaffinity warhead, which is transformed to a reactive moiety upon UV-irradiation[1]. ABPP strategies generally combine small molecule binders with a reporter tag to study interactions between small molecules and their protein targets[1]. These reporter tags can either be moieties that enable affinity purification, as in case of biotin/desthiobiotin, or tags that enable detection, such as fluorescent or radioactive groups[1]. While general ABPP strategies are further discussed in Chapter 10 of this thesis, this chapter focuses on affinity-based strategies using ABPs for target identification purposes of covalent protein binders.

Biotinylated small molecules and macromolecular probes and their wide variety of applications have transformed drug discovery and development for the past 40 years[2]. Biotin is a strongly bound target of avidin (isolated from egg-white)[2], streptavidin (isolated from *Streptomyces avidinii*)[3], and analogs, which form highly stable tetrameric target-ligand complexes with K_D values of around 10^{-15} M[2] (Figure 5.1.1). By 1990, several methods involving (strept)avidin-biotin complex formation had already been established, including affinity chromatography, gene probes, diagnostics, immunoassays, microscopy methods, flow cytometry, and others[2] (Figure 5.1.2). Due to non-specific binding issues with avidin, streptavidin is commonly preferred[2]. Notably, genetically altered versions of avidins and streptavidins were deployed to investigate the mechanisms of complex formation with biotin, as well as to tailor the properties of the complexes to enable further applications of the (strept)avidin-biotin technology[3,4].

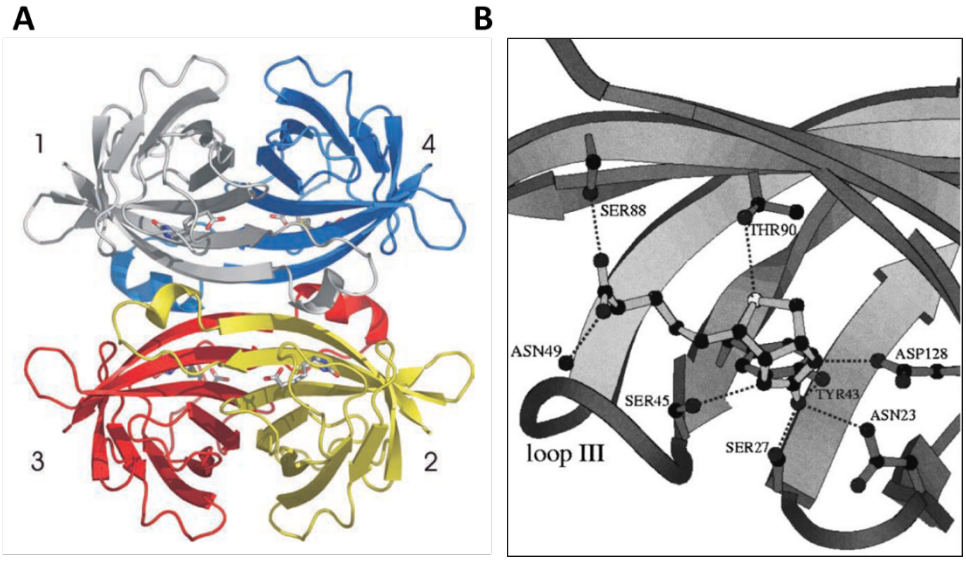


Figure 5.1.1 | The Biotin-Streptavidin Complex. A. The tetramer complex of streptavidin with bound biotin (gray sticks) is shown. Streptavidin monomers are highlighted in different colors (gray, blue, red, and yellow ribbons). Adapted from Laitinen *et al.*, *Cell Mol. Life Sci.* **2006**, *63* (24), 2992-3017[4]. **B.** The hydrogen bond network of biotin bound to streptavidin is highlighted. Adapted from Stayton *et al.*, *Biomol. Eng.* **1999**, *16* (1-4), 39-44[5].

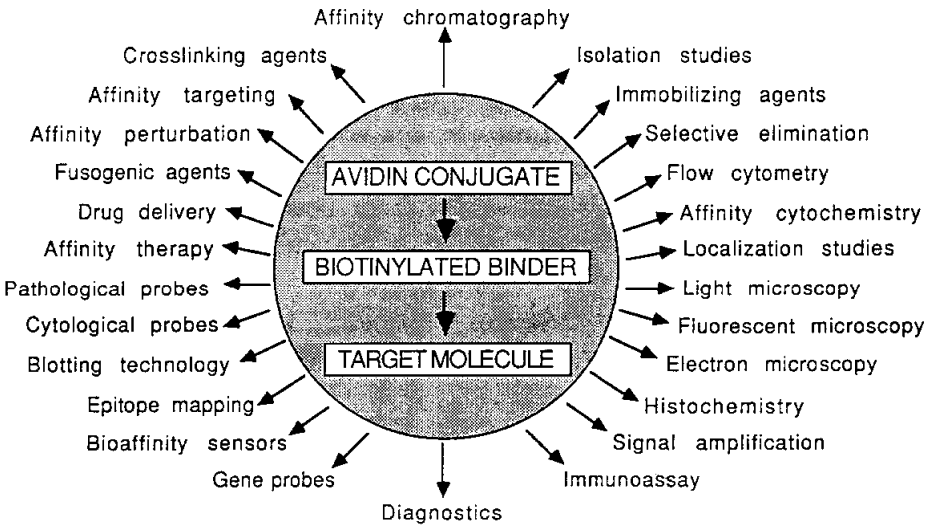


Figure 5.1.2 | Overview of Major Applications of Avidin-biotin Complex Strategies Already Available in 1990. Taken from Bayer *et al.*, *J. Chromatogr.* **1990**, *510*, 3-11[2].

One essential application of the streptavidin-biotin technology is the target identification of bioactive small molecules. Here, biotin tagged ABPs are commonly used for affinity-based target isolation *via* streptavidin-coated beads, enabling target identification of candidate proteins by Western blot analysis or more unbiased approaches utilizing mass spectrometry[6]. In the first step, an ABP is incubated with a protein mixture, e.g. a cell lysate[7]. Since ABPs are commonly equipped with an electrophilic warhead, targets are usually covalently bound to the probe[7]. Using streptavidin beads, the probe-target complexes are separated from unbound proteins and can be ultimately analyzed and

identified[7] (Figure 5.1.3), which makes this method fruitful for identifying potential protein targets. This approach can be further refined by performing competitive pulldown assays with untagged probes[7]. That is, by applying the described strategy to samples pre-incubated with an untagged compound, read-out of competition experiments can identify targets of a small molecule of interest[7]. This pulldown technique is therefore a useful tool to determine the selectivity and potency of a bioactive molecule, as well as to confirm target engagement of a probe in cells[1,8].

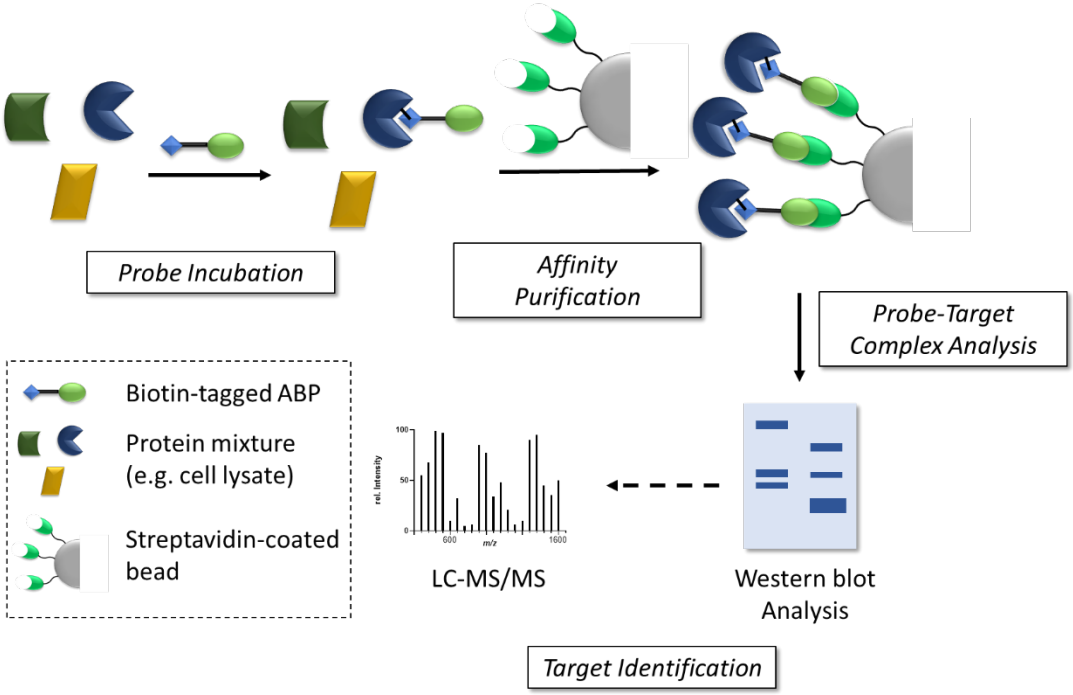
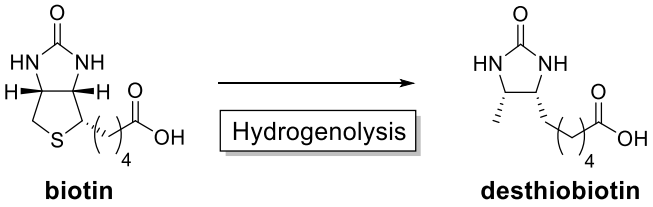


Figure 5.1.3 | General Mechanism of Affinity-based Target Isolation and Identification Using the Biotin-Streptavidin Technology. A mixture of proteins (e.g. cell lysate) is incubated with a biotin-tagged ABP. Since ABPs are commonly derived from covalent bioactive molecules, the probe binds irreversibly to its target protein (depicted as black line). The probe-target complexes are isolated using streptavidin-coated beads, which bind to the biotin of the complexes. The isolated target proteins are further analyzed by Western blot, which can be coupled with mass spectrometry analysis.

While a variety of streptavidin analogs were engineered to study and tailor streptavidin complex formation with biotin, also different biotin analogs, such as desthiobiotin[9], were investigated. As a result of structural analysis studies of biotin, desthiobiotin was formed after hydrogenolytic cleavage of the organic sulfide in biotin[9] (Scheme 5.1.1). Interestingly, because the binding affinity of the biotin-streptavidin complex is so strong, the bond between biotin and streptavidin can be described as essentially irreversible; in contrast, the desthiobiotin-streptavidin binding affinity is weaker, and the bond therefore reversible under physiological conditions[10]. In some applications, this feature of desthiobiotin is actually taken advantage of, as the harsh dissociation conditions needed to separate the strong biotin-streptavidin complexes often results in target inactivation, which leads to problems in target isolation studies[10]. One technology that applies this is the proteome-wide covalent inhibitor target-site identification (CITe-Id)[11] technology, which is described in Chapter 7.

Scheme 5.1.1 | Structural Relationship of Biotin and Desthiobiotin



Cellular Thermal Shift Assay – A Label-free Method to Study Cellular Target Engagement

The premise of thermal shift assays (TSAs) is that binding of a ligand to a protein will alter the thermal stability of this protein[12] (Figure 5.1.4). While the ligand-bound and -unbound, folded and unfolded versions of a protein exist in an equilibrium, the formation of aggregates from unfolded proteins is considered to be an irreversible process[12]. Moreover, ligands usually only bind properly folded proteins[12]. Therefore, binding of a ligand to a protein shifts its equilibrium towards its folded state, which positively correlates with the concentration of the ligand[12]. Ligand binding thus causes a shift in the melting point (T_m) of a protein, which has led to the development of a variety of different thermal shift assays, ranging from high-throughput platforms to highly accurate techniques, which are capable of not only describing ligand binding mechanisms but also the kinetic properties of a ligand-protein complex[13].

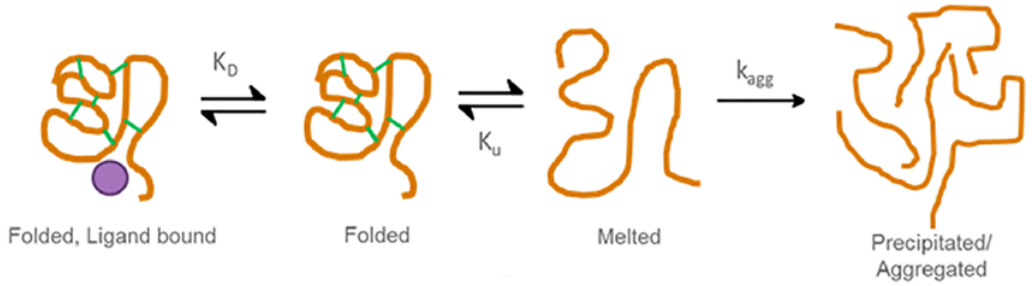


Figure 5.1.4 | Overview of Important Protein States to Consider for Thermal Shift Assays. A protein-ligand complex and the unbound, folded protein are reversible states described by the equilibrium constant K_D. Folded and unfolded (melted) protein states are interchangeable as well and are described by K_u. When unfolded protein forms aggregates, the commonly irreversible phenomenon is described by the rate constant k_{agg}. Taken from Redhead *et al*., *Biochemistry* **2017**, *56* (47), 6187-6199[12].

In 2013, Molina and colleagues applied the TSA concept to cells to develop CETSA[14]. While TSA methods commonly analyze ligand binding for purified proteins, classical CETSA determines melting curves for proteins within the context of a cellular system treated with a small molecule ligand of interest[14]. Aggregated and precipitated proteins (Figure 5.1.5 A) are separated from the soluble fraction of the respective cell lysate, which can then be analyzed by Western blot, in which band intensities are quantified and plotted against the temperature to determine the melting point (Figure 5.1.5 B)[14]. This enables the detection of small molecule binding of a target protein in the context of complex environments such as within cells or tissues[14]. Since the stabilizing effect of a ligand depends on ligand concentration, isothermal concentration-response experiments (iso-

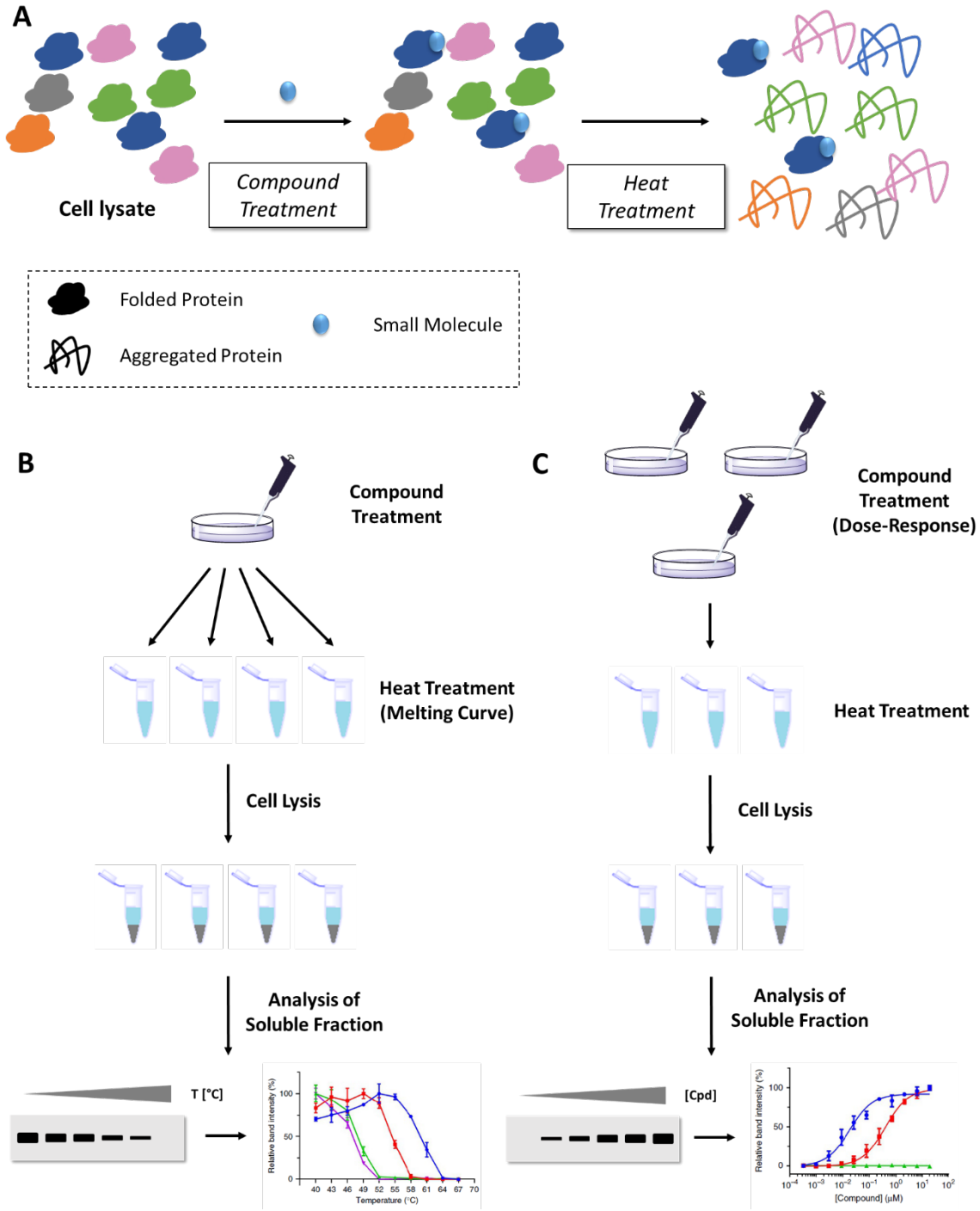


Figure 5.1.5 | Experimental Set-up of Cellular Thermal Shift Assays. A. A protein mixture, e.g. cell lysate, is treated with a bioactive small molecule. Bound protein targets are more heat-resistant due to the stabilizing effect of the ligand, while unbound proteins will aggregate and precipitate after heat treatment. **B.** Shown is the general experimental procedure to determine the melting curve of a protein of interest. Briefly, cells are incubated with a small molecule or DMSO. The cells are aliquoted, and the aliquots treated at increasing temperatures. After lysis of the samples, the soluble fractions are analyzed by Western blot[a]. **C.** Shown is the general experimental procedure of an isothermal CETSA set-up. Cells are treated with different concentration of a compound or DMSO. The cells are then incubated at a specific temperature and the soluble fractions of the cell lysates are analyzed by Western blot[a].

[a] Plotted result graphics are adapted from Jafari *et al.*, *Nat. Protoc.* **2014**, *9* (9), 2100-2122[16].

thermal dose-response fingerprint; $ITDRF_{CETSA}$[15]) have been developed as another tool for the investigation of target engagement[16] (Figure 5.1.5 C).

Subsequently, novel variations of the CETSA method have been established and have further advanced this technique. For example, coupling CETSA with mass spectrometry (MS) enables TSA-based proteomic profiling of a compound of interest to elucidate its potential binding partners and its mode of action[15,17]. CETSA techniques have further been applied to metabolomic studies, PPI analysis, post-translational modification investigations[18], and the analysis of membrane-based proteins[19]. Finally, high-throughput versions of CETSAs have been used for drug screens in early drug discovery[20], highlighting the broad range of applicability of CETSA techniques to drug discovery and development processes[21].

While the CETSA methodology is technically fairly simple and can be established under general laboratory conditions as there are no specific reagents needed to perform the assay and controlled heating periods can be performed using commonly available thermocyclers, this technique does have some disadvantages. For instance, the equilibrium for protein binding of small molecule ligands is temperature-dependent and adjusts during the heating pulse, which can lead to inaccurate results[22,23]. It has also been reported that some ligand-target interactions do not lead to an increase in thermal target stability, due to domain-specific effects, general target protein properties, or very high melting temperatures of the target of interest[24]. Notably, while ligand concentration influences target stability, target concentration within a cellular system can also lead to differences in outcome depending on the cellular system used in a CETSA[25]. Finally, general assay conditions, including incubation temperature, as well as heating and cooling periods, further alter the assay results and are main contributors to fluctuations between assay repeats[25]. Nevertheless, the relative tendencies of effects of small molecule ligands on target protein stability stay the same, making CETSA methodologies a broadly applicable and useful technique[25].

In summary, although CETSA results should be considered in context and validation with other assay results[25], this method is a straightforward technique for analyzing the cellular target engagement of bioactive small molecules, without needing to chemically alter the molecular structure of the compound of interest, as is necessary for ABPP-based methods described above. Moreover, CETSA allows for direct analysis of the mode of action of a small molecule in the complex environment of a cell, which is usually only feasible *via* indirect methods, such as studies of down-stream target effects[14].

5.1.2 Targeting Strategies in the Treatment of Cancer – A Brief Overview

Besides the fact that indications of cancerous diseases were already found in dinosaur fossils as old as 70 million years, the first documentations of human cancers date back to approximately 3,500 years ago, in which cases of breast cancer in Egypt were described[26]. The basis of the historic perception and treatment of cancerous diseases was primarily established by observations and theories of Hippocrates (~400 B.C.)[26]. His association of cancer and the physiology of a crab[27] led to the nomenclature of “karkinos” and “karkinoma” for benign and malignant tumors, respectively[26]. In the 2nd century, Claudius Galen further refined Hippocrates’ work and hypothesized the origin of cancerous abnormalities as the result of an imbalance of either black or yellow bile[27]; a theory which persisted until long after the end of the middle ages[26]. Even though several connections between environmental factors and the occurrence of cancerous conditions had been

observed since the early 18th century[26,28], it was only until recently that cellular mechanisms of cancer started to be uncovered. As cancer rose to the second leading cause of death worldwide with ~9.6 million deaths in 2018 (~17% of overall deaths in 2018)[29], the dire need of advancements in cancer treatment regiments and thus our understanding of cancer biology becomes evident.

Hallmarks of Cancer

Cancerous diseases have been generally perceived as masses of uncontrolled and excessive growth in the past, while the underlying molecular mechanisms remained a mystery[26,27,30]. However, at the end of the 20th century, cancer research had immensely expanded our knowledge of the origins and development of tumors[31]. Within 20 years, it was found that a vast variety of individual diseases are covered by the umbrella-term "cancer" and that a tumor often originates from only one initial mutated cell[31]. Genetic studies further identified two classes of genes that are able to initiate tumor growth and development upon mutation: proto-oncogenes (driving cancer growth) and tumor suppressor genes (inhibiting cancer growth)[31].

In 2000, these and other findings were summarized in Hanahan and Weinberg's much-cited publication outlining the six "Hallmarks of Cancer"[32] (Figure 5.1.6). Briefly, it was proposed that cancerous cells are characterized by six universal differences to healthy cells, which can be applied to any tissue in the human body[32]. The most obvious characteristics of cancer cells is cell growth that is independent from extracellular stimuli, as well as a developed resistance towards intracellular growth inhibiting signals, senescence, and apoptosis cascades[32]. In order to survive, cancer cells thus require an increased supply of nutrition, which depends on the development of an additional blood vessel system[32]. In the final stage, a primary tumor invades other tissues, resulting in metastasis, which initiates the formation of new tumor sites[32]. Due to the advances made in the understanding of cell biology, several cellular signaling pathways have been thus discovered to be associated with these different hallmarks of cancer[32] (Figure 5.1.7).

After eleven years, Hanahan and Weinberg revisited their model and added four more hallmarks of cancer, while providing more detailed explanations of potential mechanisms for each hallmark[33] (Figure 5.1.6):

(1) "*Sustaining Proliferative Signaling*" describes the ability of cancer cells to promote their own growth[33]. Several mechanisms have been described, including the production of growth inducing signaling molecules and respective receptors, the production of signaling agents to stimulate healthy cells to produce growth stimulating ligands, and the upregulation or mutation of respective receptors[33]. Intracellular activation of growth stimulating pathways, which are regulated by numerous signaling cascades, poses another option[33].

(2) "*Evading Growth Suppressors*" is a contributing feature of cancer cells to yield unhindered cell growth[33]. While there are several triggers and regulators inducing cellular growth cascades, there are always respective counter mechanisms to inhibit these pathways and promote growth arrest[33]. Blockage or mutation of such regulators supports continuous growth signaling[33]. Tumor suppressors also work through communication with other cells to trigger growth inhibition caused by densely grown cell populations[33]. It has been shown that tumor cells can disable these mechanisms to avoid cell contact-based growth regulation[33].

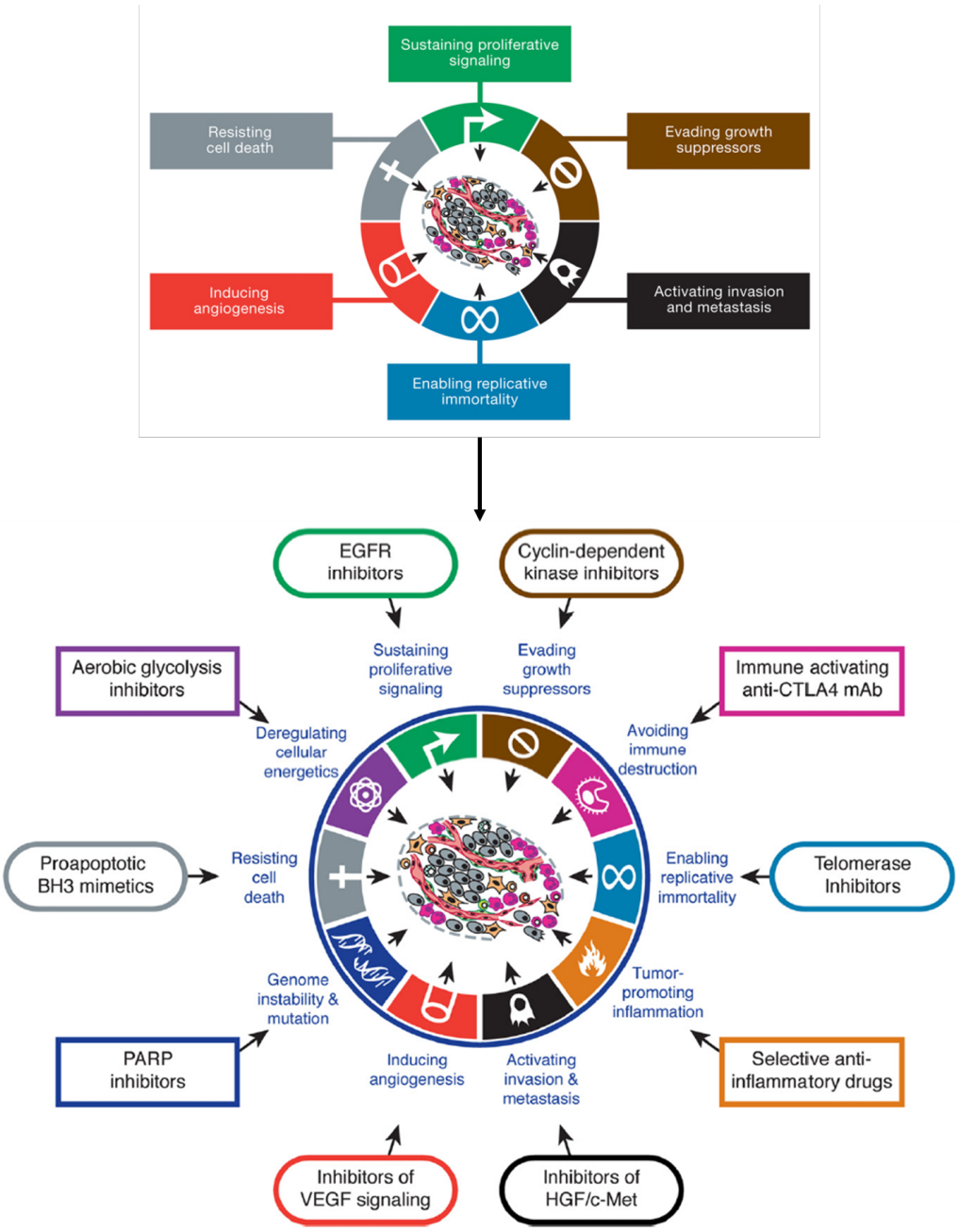


Figure 5.1.6 | Hallmarks of Cancer. Healthy human cells can mutate to develop into cancer cells, which are characterized by specific hallmarks. In 2000, six characteristics of cancer were defined (top graphic)[a]: self-controlled proliferation signaling, decreased growth suppressing factors and cell death, immortalization of the cell, increased angiogenesis, and further spread *via* local invasion or metastasis. In 2011, four additional hallmarks were added (bottom graphic)[a]: genomic instability to enable mutation, inflammatory tumor environment to further support cancer development, evasion of immune responses, and adapted metabolism of the cancer cell. Respective targeting strategies to treat individual hallmarks are depicted in outlined boxes next to the respective hallmark feature.
[a] Adapted from Hanahan and Weinberg, *Cell* **2011**, *144* (5), 646-674[33].

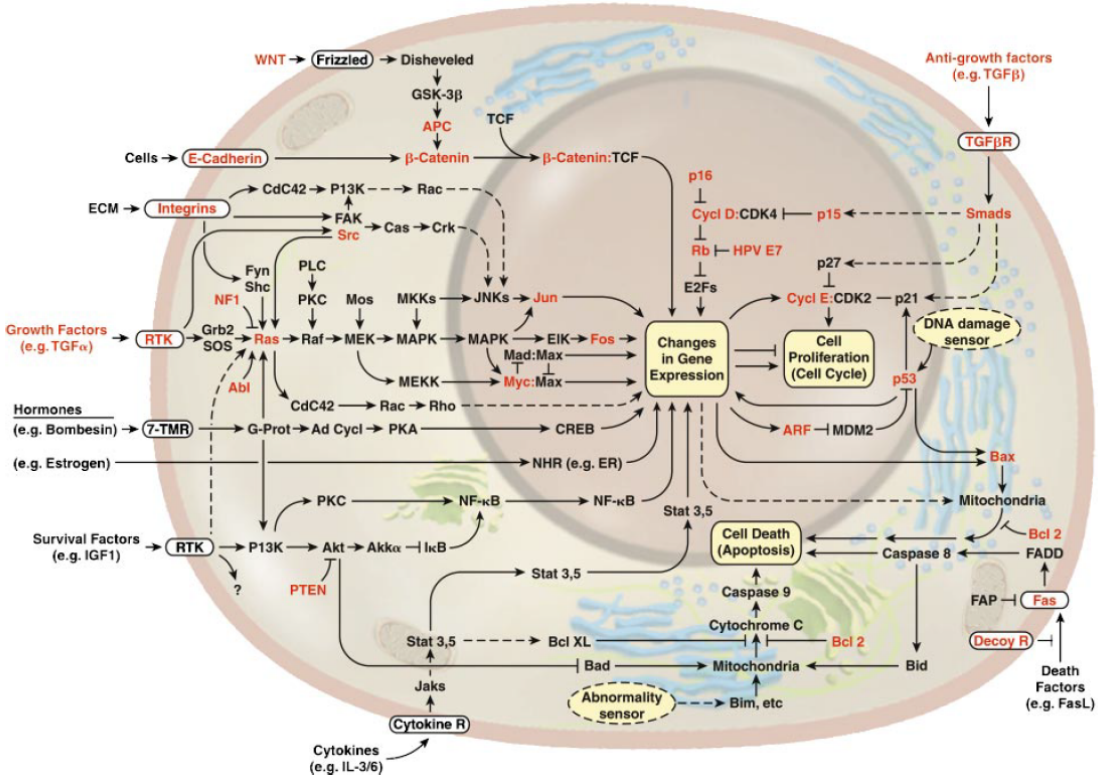


Figure 5.1.7 | Overview of Essential Cellular Signaling Pathways in Cancer. Selected signaling pathways that play an important role in cancer development are shown. Growth factors and hormones are essential proliferation driving factors, acting mainly through intracellular Ras/Raf and the mitogen-activated protein kinase (MAPK) pathways[32]. Survival and death factors, as well as cytokines, regulate the viability of a cell[32]. Phosphatidylinositol 3-kinase (PI3K) and the respective antagonist phosphatase and tensin homolog (PTEN), the signal transducer and activator of transcription 3/5 (STAT3/5) pathway, and caspase-mediated apoptosis are key players in regulating the intracellular mechanisms. As a growth-inhibition factor, transforming growth factor beta (TGF-β) is an essential promotor of cytostasis[32]. Other growth-regulating mechanisms work through contacts of extracellular matrix (ECM) or other cell contacts, which control cell motility, e.g. *via* the β-catenin pathway[32]. The overview underlines the interconnectivity of the pathways orchestrating the different hallmarks of cancer, as well as the importance of the microenvironment for the regulation of each of the described cascades[32]. Taken from Hanahan and Weinberg, *Cell* **2000**, *100* (1), 57-70[32].

(3) "*Activating Invasion and Metastasis*" describes the local and distant spread of a tumor, respectively[33]. Generally, regulating cell surface proteins as communicators of cell population organization is one of the main mechanisms through which cancer cells enable these processes[33]. However, a comprehensive understanding of invasion and metastasis is still missing, while a variety of contributing pathways and cascades are currently being uncovered and investigated[33].

(4) "*Enabling Replicative Immortality*" refers to cancer cells and their strategies to avoid entering senescence[33]. In healthy cells, a natural response to overstimulation of growth signaling is the initiation of a dormant state of the cell[33]. Hence, careful control of growth stimulating pathways to gain maximal replication rates, while avoiding the shift towards cellular senescence, is crucial for the development of cancer cells[33]. Another strategy to enable "unlimited" cell growth is by expressing

the DNA polymerase telomerase in order to counteract natural DNA tail degradation[33]. Notably, under certain circumstances, lack in telomerase activity has been linked to increased DNA mutations, actually enhancing tumor development[33].

(5) "*Inducing Angiogenesis*" in order to accommodate an increase in demand for nutrients to sustain high replication rates, as well as an accumulation of cellular waste, describes another important step in the development of a tumor[33]. Typically, cancer cells continuously activate blood vessel development *via* the secretion of respective signaling molecules[33].

(6) "*Resisting Cell Death*" is an important characteristic of cancer[33]. Regulated cell death – or apoptosis – is an essential cellular mechanism, which is commonly triggered by cellular stress[33]. Cancer cells are able to circumvent initiation of this cascade by regulating expression of initiators and/or suppressors of this pathway[33].

Newly added hallmarks of cancer (Figure 5.1.6):

(7) "*Genome Instability and Mutation*" describes a major prerequisite of a cell to acquire and develop necessary mutations leading to the cancerous state of a cell[33]. This is achieved by an insufficient machinery to oversee genomic maintenance but also includes epigenetic control of the genome[33].

(8) "*Tumor-promoting Inflammation*" summarizes the overall effect of the immune cell infiltrated environment of cancer cells within a tumor on several other hallmarks of cancer[33]. Said effects resemble inflammatory processes and reach from the supply of growth and survival factors and angiogenesis-inducing agents to an increase in reactive oxygen species (ROS), which in turn promote genomic instability[33].

(9) "*Deregulating Cellular Energetics*" poses another logical result from upregulated cell growth and the consequently higher demand for energy and the production of cellular components, such as DNA nucleotides, proteins, organelles, and others[33]. One successful strategy of cancer cells to regulate their energy production is *via* the so-called "Warburg effect", which describes the ability of cancer cells to perform glycolysis even under aerobic conditions[33]. This allows for the generation of several cellular macromolecules needed for upregulated replication rates[33]. In order to make up for lower energy outcomes of this process, cancer cells in turn upregulate transporters to increase uptake of necessary nutritional molecules, such as glucose[33]. In summary, cancer cells, especially in the complex environment of a tumor, develop balanced energy and nutrition regulating systems to afford increased growth rates[33].

(10) "*Avoiding Immune Destruction*" is a necessary characteristic of cancer cells due to the presence of a variety of immune cells as part of the tumor microenvironment (Figure 5.1.8), described in point (8)[33]. Two important mechanisms of cancer cells to evade an immune response include the secretion of immunosuppressive factors and the recruitment of immunosuppressive cells of inflammatory processes[33].

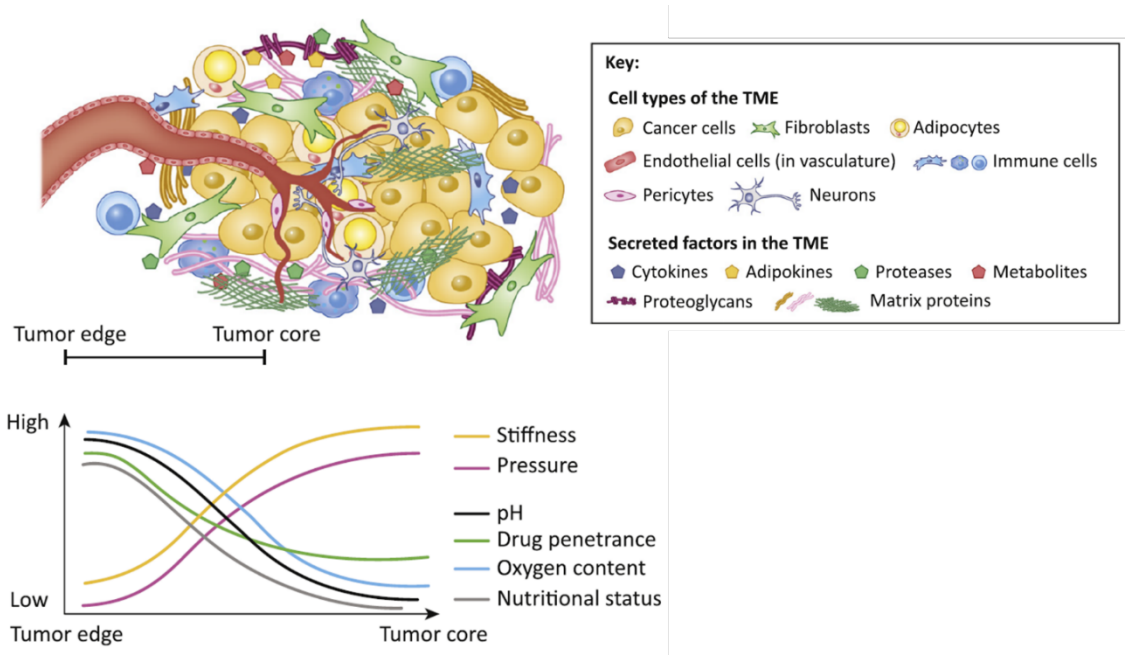


Figure 5.1.8 | The Tumor Microenvironment. The tumor microenvironment is a complex interplay of various cell types. Cancer cells at the tumor core face different conditions than cells at the edges of a tumor, which leads to different subtypes of cancerous cells throughout a tumor mass. Adapted from Kozlova *et al.*, *Trends Pharmacol. Sci.* **2020**, *41* (3), 183-198[34].

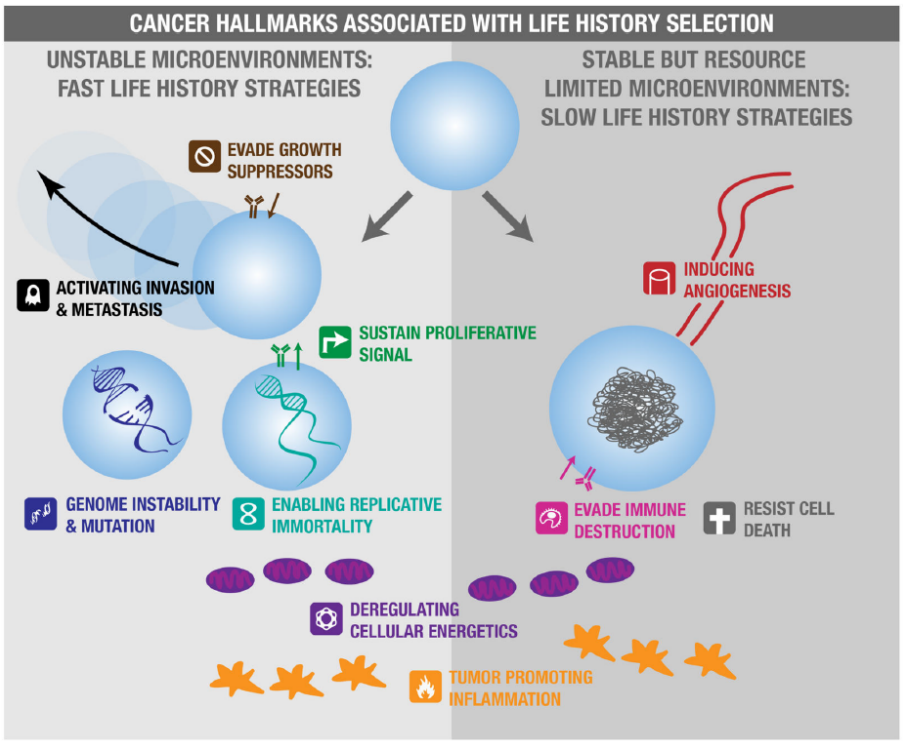


Figure 5.1.9 | Evolutionary Regulation of Cancer Hallmarks. The hallmarks of cancer can be viewed as regulatory consequences of environmental conditions a cancer cell is faced with. While unstable microenvironments trigger fast cell development strategies, environments, which are stable but lack resources, require slow developmental mechanisms. Taken from Aktipis *et al.*, *Nat. Rev. Cancer* **2013**, *13* (12), 883-892[35].

In summary, the analysis of cancer hallmarks elucidated many important intracellular pathways, as well as extracellular stimuli, that are essential for cancer development[32,33]. However, this overview also emphasizes the complexity of the regulatory mechanisms of cancer (Figure 5.1.7). Especially in recent years, the importance of the tumor microenvironment was uncovered[34] (Figure 5.1.8), further complicating a comprehensive model of cancer regulating pathways. Even within one tumor, multiple different types and stages of cancerous cells can be found[35]. Furthermore, some literature suggests that this is a phenomenon of evolutionary development, depending on the conditions a cancer cell is facing[35]. Hereby, based on the ten hallmarks proposed by Hanahan and Weinberg[33], unstable microenvironments are associated with fast cell growth and therefore increased mutation rates, cellular immortality, and self-sustained proliferation, as well as invasion, and disabling growth suppressing signals[35]. Environments of low resources, but stable conditions therefore favor slow life strategies, including induced angiogenesis, and suppression of an immune response as well as cell death[35]. Both conditions require the cancer cell to adapt their energy metabolism and are based on an inflammatory tumor environment[35] (Figure 5.1.9).

Early Chemotherapeutics – A Short Introduction

Historically, cancer treatment was just as much a mystery as were the underlying mechanisms that caused cancer[26,27]. From magical rituals to toxic, heavy metal-based formulations, many options had been tried in the attempt to control and cure this puzzling disease[26,27]. As a result, surgical removal of a tumor was and still is one of the most widely applied treatments for solid tumors[26,27,36]. Over the years, radiation therapy and first chemotherapeutics have been added to the initial treatment regimen of cancerous diseases and are commonly used in combination to increase therapeutic success[36]. However, these treatments exhibit severe toxicity effects due to their lack in selectivity for cancer cells[36]. Thus, tailored therapies specific for targeting cancer cells have been the goal of latest research endeavors, leading to the development of immunotherapies, antibody-directed therapies, small molecule-based targeted therapeutics, and proteolysis targeting chimera (PROTACS)[36].

This chapter focuses on small molecule-based (targeted) cancer therapies and serves as a brief overview of common targeting strategies to engage cancerous cells. After a brief introduction to the traditional chemotherapeutic agents used in cancer treatment, some examples of novel targets that are currently being investigated will be discussed. Notably, this chapter does not give a comprehensive overview on current targets for cancer treatment, which would go beyond the scope of this thesis.

Early chemotherapeutics (Figure 5.1.10) are highly unspecific agents that are primarily toxic to fast dividing cells, such as cancer cells[37]. The expression “chemotherapy” was introduced by Paul Ehrlich in the beginning of the 20th century, when he was working on anti-infectives (also see Chapter 10.1) and refers to “the use of chemicals to treat disease”[37]. One of the earliest chemotherapeutics used for treating cancer were alkylating agents such as derivatives of nitrogen mustards or cyclophosphamide[37]. Another important class of early chemotherapeutic agents are antimetabolites[37]. While Sidney Farber (Dana-Farber Cancer Institute, Boston) investigated folic acid analogs to treat pediatric leukemia patients already in the 1940s, methotrexate (folic acid antagonist), 6-

mercaptopurine (inhibits nucleic acid synthesis), 5-fluorouracil (inhibits nucleic acid synthesis), and others have been approved for several malignant diseases in the course of the following decades[37].

The DNA-crosslinking agent cisplatin was approved in the late 1970s and (including further analogs) adds to the landscape of small molecule chemotherapeutics in the mid to late 20th century[37]. While cisplatin's effects on cell growth were first identified in bacteria, also other antibiotic agents were later added to cancer-treating chemotherapeutics, including actinomycin D (as a result of Waksman's efforts in the 1940s; also see Chapter 10.1), bleomycin, and anthracycline antibiotics[37]. Furthermore, plant-derived cytotoxins, such as podophyllins, vinca alkaloids (e.g. vinblastine and vincristine), taxanes (e.g. paclitaxel), and semi-synthetic analogs of camptothecin (e.g. irinotecan) have been introduced as cancer therapeutics[37]. Table 5.1.1 summarizes some examples of early chemotherapeutic agents for the treatment of cancer mentioned in this chapter and highlights that most of these molecules cause DNA damage or target proliferation mechanisms, making fast proliferating cells most vulnerable to these treatments. While this includes cancer cells, a variety of non-cancerous cells, such as blood cells, hair follicle cells, and others, are affected as well, leading to the observed side effects of chemotherapy[36].

In order to decrease off-target effects of chemotherapeutics, targeted therapies arose[36,37]. For example, hormone-dependencies of cancers, such as androgen-dependent prostate cancers, were observed, which led to the development of the antiandrogenic therapy concept in the 1940s[37]. Similarly, antiestrogenic therapies for estrogen-dependent breast cancers were established in the 1960s[37]. Furthermore, even though the basic idea of targeting angiogenesis was already introduced by Judah Folkman in the 1970s, and monoclonal antibodies were envisioned to treat cancers *via* immunomodulatory effects in the late 1990s, most improvements in targeted therapies were only seen within the past 20 years[37].

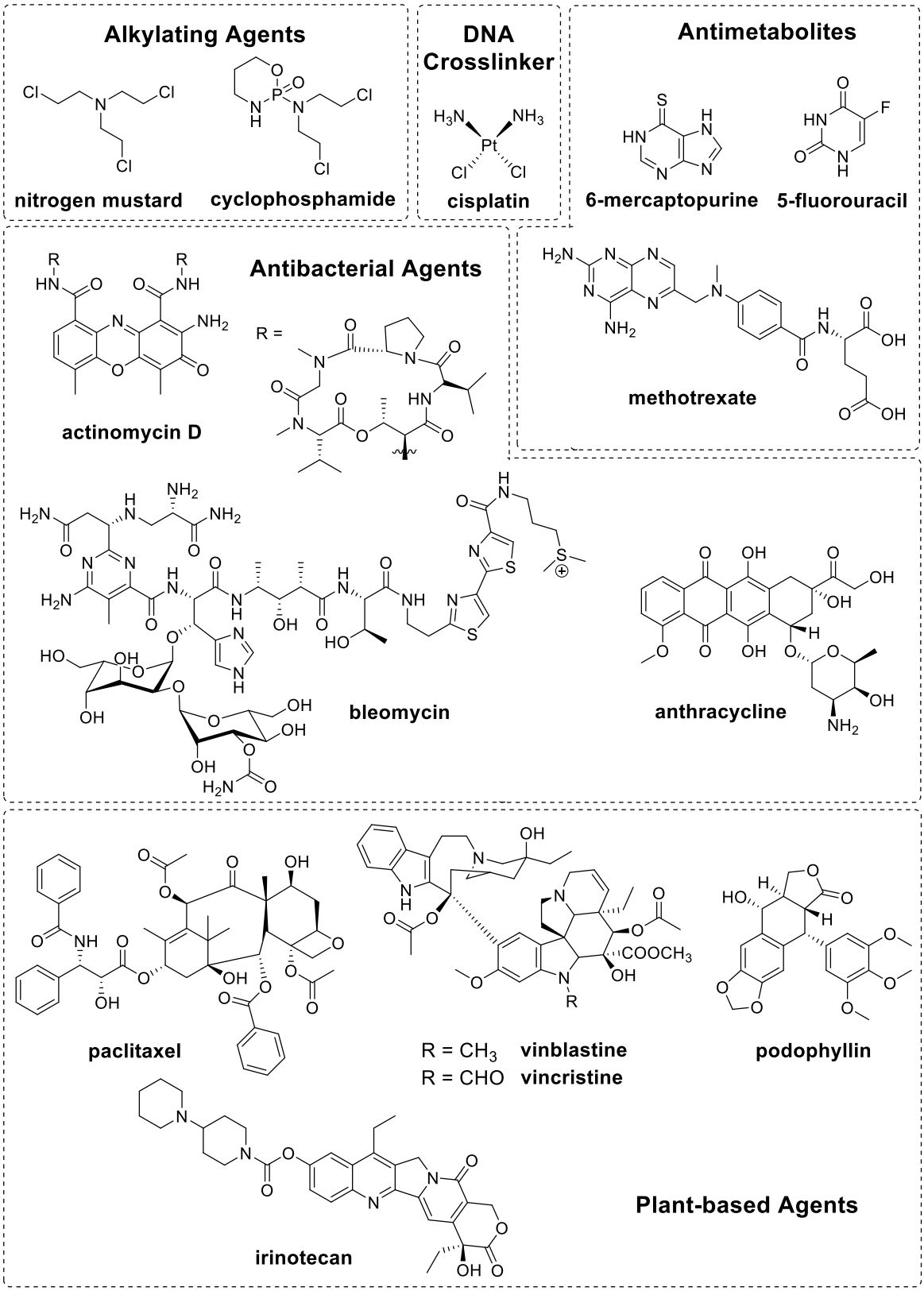


Figure 5.1.10 | Chemical Structures of Early Chemotherapeutics Used for Cancer Treatment. Chemotherapeutic agents of the mid to late 20th century are characterized by exhibiting cytotoxic effects *via* DNA damaging mechanisms or unselective targeting of proliferation cascades. Synthetic chemotherapeutics include alkylating agents, antimetabolites, and DNA closslinkers. Other early chemotherapeutics were derived from antibacterial compounds or isolated from plant-based materials.

Table 5.1.1 | Early Chemotherapeutics in the Treatment of Cancer

Therapeutic Agent	Target[a]	Mechanism of Action	Cancer Hallmark	Ref.
Nitrogen mustards	*nonspecific*	alkylating agent	Proliferation	[37]
Cyclophosphamide	*nonspecific*	alkylating agent	Proliferation	[37]
Methotrexate	DHFR	folic acid antagonist	Proliferation	[37]
6-Mercaptopurine	PRPP	nucleic acid antagonist	Proliferation	[37,38]
5-Fluorouracil	TS, RNA	nucleic acid antagonist	Proliferation	[37,39]
Cisplatin	DNA (purine bases)	DNA-crosslinking agent	Proliferation	[37,40]
Actinomycin D	DNA	transcription inhibition	Proliferation	[37,41]
Bleomycin	RNA, DNA	RNA/DNA cleavage	Proliferation	[37]
Anthracyclines	topoisomerase II	DNA cleavage	Proliferation	[37]
Podophyllins	topoisomerase II	DNA damage	Proliferation	[37]
Vinca alkaloids	β-tubulin	spindle formation inhibition	Proliferation	[37]
Paclitaxel	microtubules	microtubule complex stabilization	Proliferation	[37]
Irinotecan	topoisomerase I	DNA cleavage	Proliferation	[37]

[a] Abbreviations: DHFR, dihydrofolate reductase; PRPP, phosphoribosyl pyrophosphate; TS, thymidylate synthase.

Small Molecule-based Targeted Cancer Therapies – A Novel Approach

After the identification of DNA and its structure in the 1950s, uncovering of DNA mutations and their implications in cancer inevitably followed[42]. For example, a variety of different DNA repair mechanisms, as well as DNA damage response (DDR) pathways were identified[42]. DNA damage and mutations caused by environmental or intracellular factors are important prerequisites for the initiation of cancer and are recognized as an important hallmark of cancerous diseases[33]. It was found that these cancer initiating mutations commonly require a downregulated DDR for the development of a mature cancer cell, while dysregulation in DNA repair processes support initial cancer forming mutations (Figure 5.1.11)[42]. As several traditional chemotherapeutic agents as well as radiation therapy cause DNA damage[36], the understanding of these DNA repair mechanisms and signaling pathways led to a better understanding of the development of cancer resistances to these treatments[43]. That is, while inaccurate DNA damage repair and DDR signaling induce cancer cell formation initially, these pathways are necessary regulators to stabilize

the genome for the survival of cancer cells due to their increased cellular stresses[44] and further protect the cells from damage by cytotoxins, such as anticancer treatments[43].

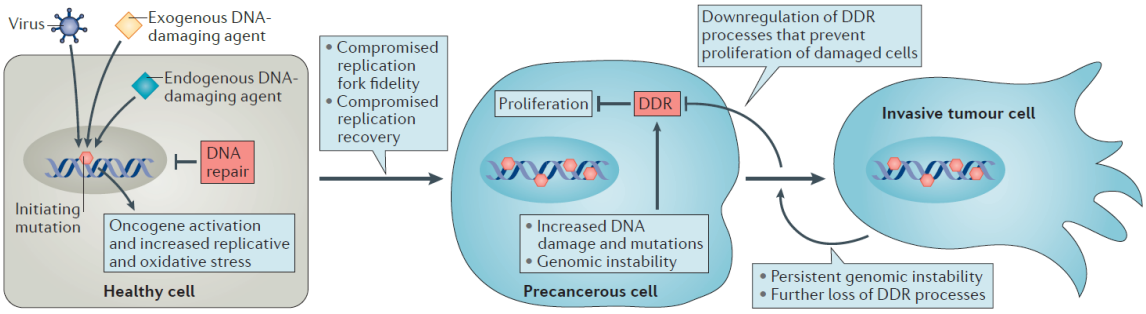


Figure 5.1.11 | The DNA Damage Repair and Response Mechanisms and the Development of Cancer. Initial cancer-causing mutations are commonly detected and regulated by DNA repair mechanisms. If these mechanisms fail to repair the DNA, DDR pathways are initiated to stop proliferation. However, continuous DNA damage/mutations can lead to dysregulated DDR signaling, which leads to the development of a mature cancer cell with increased proliferation. Taken from Jeggo *et al.*, *Nat. Rev. Cancer* **2016**, *16* (1), 35-42[42].

Structurally impaired DNA, including strand breaks, replication lesions, or gaps, triggers a diverse network of DNA repair mechanisms, such as mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), translesion synthesis (TLS), homologous recombination (HR), and others[44]. BER, being mainly involved in the repair of damaged nucleobases, is the most commonly applied DNA repair mechanism in human cells and is a key player in the DNA damage repair mechanisms induced by traditional chemotherapeutics[45]. Thus, inhibition of the BER pathway sensitizes cancer cells to the cellular stress response, as well as common cancer treatments, including DNA damaging chemotherapeutics and radiation therapy[45].

One key player in the BER signaling pathway is poly-(adenosine diphosphate (ADP)-ribose) polymerase 1 (PARP1), which detects single-stranded DNA breaks and tags adjacent proteins with poly(ADP)-ribose (PAR) chains for further initiation of DNA repair mechanisms[46]. Thus, several substrate-competitive PARP1 inhibitors, which are structurally derived from nicotinamide adenine dinucleotide (NAD^+), have been developed and applied as treatments for a vast variety of different cancers in the clinic[46,47]. Since tumor suppressors breast cancer type 1/2 susceptibility proteins (BRCA1/2) are involved in the HR DNA repair pathway, synthetic lethal interactions with PARP inhibitors were identified, leading to the approval of olaparib (PARP1 inhibitor) for *BRCA*-deficient ovarian cancers[43].

Within the DDR, ataxia telangiectasia and Rad3-related protein (ATR), ataxia telangiectasia mutated (ATM), and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) are essential regulators[48] and therefore interesting targets for targeted cancer therapy. The kinase ATR is activated by replication-induced cellular stress signals, ensuring genomic stability during replication *via* phosphorylation of several downstream targets involved in cell cycle regulation, DNA repair mechanisms, and replication stress[48,49]. As a result, first ATR inhibitors are currently being tested in clinical trials[49]. Similarly, ATM inhibitors as well as inhibitors of ATR and ATM down-stream targets, including cell cycle checkpoint kinases 1 and 2 (CHK1/2), respectively, are currently being

investigated[50]. Part of the same kinase family (PI3K-related kinase family; PIKK family) as ATR and ATM is the DNA-PK, which is involved in the repair mechanisms of double-stranded DNA breaks, cell cycle regulation and telomere maintenance[51]. Thus, small molecule-based DNA-PK inhibitors can be found in several clinical trials for the treatment of different cancers[50].

As described, treatments inducing replication stress or inhibiting the cancer cell's defense mechanisms to protect itself from replication stress are promising targets for targeted cancer treatments. Consequently, further strategies targeting or inducing replication stress have been explored[52]. A common origin for replication stress is a lack in resources for DNA replication, leading to an increase in single-stranded DNA (ssDNA) caused by helicase activity[52]. Briefly, ssDNA is normally stabilized by replication protein A (RPA), while an excess in ssDNA can lead to a "replication catastrophe" and ultimately cell death[52]. Besides the already discussed ATR-CHK1, ATM-CHK2, as well as DNA-PK pathways, another main regulator of the replication stress response is the kinase WEE1[52]. WEE1 is further downstream of the ATR/ATM pathways and inhibits CDK1-mediated mitosis entry in case of persistent DNA damage signaling[53]. However, there are only a few WEE1 inhibitors studied in preclinical and clinical trials, suggesting that further development of WEE1-targeting chemical tools has to be done to comprehensively elucidate the potential of targeting this key regulator *in vivo*[53]. Additionally, other potential targets involved in these signaling pathways such as Schlafen 11 (SLFN11)[54] and the Y-family of TLS polymerases[55] are being discussed.

While WEE1 executes cell cycle control *via* CDK1 phosphorylation due to DDR signaling, other cell cycle regulating pathways have also been identified as promising targets for cancer therapy[56]. Key regulators of cell cycle progression are CDKs, especially the isoforms CDK1/2/4/6[56], which are discussed in Chapter 6 of this thesis. Other potential targets include the Polo-like kinases (PLKs)[56,57] and the Aurora kinases[56]. For example, PLK1 has been identified to regulate CDK1 function *via* cell division cycle 25C (CDC25C) activation, which in turn activates CDK1 and degrades WEE1, while also being involved in other cell cycle regulating mechanisms such as centrosome maturation[56]. Similarly, Aurora kinases are key players in coordinating mitosis *via* organizing and controlling centrosome function, spindle development, and chromosome movement during cell division processes[58]. Inhibitors of PLK1 and Aurora A have thus been developed and are being investigated in clinical trials[56].

In contrast to these strategies interfering with mitosis entry and execution, therapies targeting mitotic exit leading into interphase have been discussed[59]. For example, traditional chemotherapeutics, such as paclitaxel or vinca alkaloids (Table 5.1.1 and Figure 5.1.10), disrupt mitosis *via* interference of spindle and tubulin function, thus causing cell cycle arrest[59]. However, cancer cells commonly circumvent these mechanisms using different regulating strategies, including ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C), cyclin B, mitotic kinases and phosphatases (including Aurora kinases and PLK1, but also protein phosphatase 2A (PP2A)), as well as kinesins and microtubule-binding proteins[59], providing another set of potentially novel cancer therapeutic targets.

Besides replication stress caused by environmental factors or DNA damaging chemotherapeutics, cancer cells are commonly faced with other cellular stress-inducing phenomena, leading to a variety of stress response-related therapeutic options for cancer

treatments[60]. For instance, upregulated aerobic metabolism pathways also lead to an increase in ROS, and thus oxidative stress, while a growing need for metabolic products induces metabolic stress, leading to changes in metabolic regulations – one of the hallmarks of cancer[60]. These metabolic systems are stirred by several different mechanisms and pathways, including adenosine monophosphate (AMP)-activated protein kinase (AMPK)[60], mechanistic target of rapamycin (mTOR) complex[61], hypoxia-inducible factors 1-α (HIF-1α)[62], and several others, providing a vast variety of potential therapeutic targets[60]. Furthermore, redox-related stress signals are also connected to endoplasmic reticulum (ER) stress processes, a consequence of alterations in protein homeostasis[63]. Consequently, ER stress regulates a number of different cellular effects, including cell death, differentiation, and intercellular communication processes[64]. While the heat-shock protein chaperone system (heat shock protein 90 (HSP90)), the unfolded protein response pathway (including targets such as protein kinase RNA-like ER kinase (PERK), inositol-requiring enzyme 1 (IRE1), and activating transcription factor 6 (ATF6)), the proteasome and de-ubiquitinating enzymes (DUBs), the aggresome pathway (histone deacetylase 6 (HDAC6)), and autophagy control the protein life-cycle, disruption of these pathways pose interesting targets to induce ER stress as a therapeutic strategy in cancer treatment[63].

As the tumor grows, physical stresses caused by the interaction of a cell with the environment are communicated through integrins, ion channels, and catenins (mechanosensors)[60]. A multitude of intracellular signaling cascades transforms these extracellular triggers into cellular effects, including the focal adhesion kinase (FAK), Src, PI3K, Rho/Rho kinase, and others, in order to alter the cytoskeleton in accordance with the environmental conditions[60]. Intracellular pathways triggered by FAK or stress sensors, such as integrins, are therefore attractive targets for cancer therapeutics[60,65].

A general theme of stress signaling cascades, as well as most cancer therapeutics, is the ultimate consequence to lead a cancer cell into cell death (apoptosis)[66] (Figure 5.1.12). Consequently, cancers are masters in evading these growth regulating mechanisms, which is thus one of the hallmarks of cancer[33]. This in turn spurred intensive studies of the apoptotic pathways as potential tumor therapeutic targets themselves[66]. Generally, while there are numerous ways of triggering apoptosis, the process is mainly controlled by caspases, which are activated by cytochrome *c*[66]. The central cellular mechanism of apoptosis, once activated, is then the breakdown and degradation of the nucleus and the contained DNA[66]. Briefly, intrinsic stimuli lead to cytochrome *c* release from mitochondria that is controlled by proteins of the B-cell CLL/lymphoma 2 (BCL-2) family[66]. That is, BCL-2 interacting mediator of cell death (BIM), BCL-2 homolog 3 (BH3) interacting domain death agonist (BID), and BCL-2 binding component 3 (BBC3) activate BCL-2 associated X protein (BAX) and BCL2-antagonist/killer1 (BAK1), which induce apoptosis *via* formation of pores in the mitochondrial membrane, whereas BCL-2, BCL-X_L, BCL-W, BCL-2-A1, and myeloid cell leukemia sequence 1 (MCL1) inhibit the pathway by binding of the activators[66]. Interestingly, BID is also involved in inducing the apoptotic pathway caused by extrinsic signaling cascades, which are communicated *via* death receptors such as Fas, tumor necrosis factor (TNF) receptors (TNFRs), and TNF-related apoptosis-inducing ligand (TRAIL) receptors[66].

Ultimately, cancer cells developed several mechanisms to interfere with these cascades, many of which are now established oncogenic pathways and targeted by a vast number of cancer therapeutics in clinic and in development[66], including AKT and the MAPK

pathway[67], extracellular signal-regulated kinase (ERK)[68] and upstream effectors Ras, Raf, and MAPK kinase (MEK)[69], mTOR[70], and others[66,71,72]. Furthermore, inhibitors of the BCL-2 family proteins, e.g. BAX, BIM, and MCL1, have been developed[66].

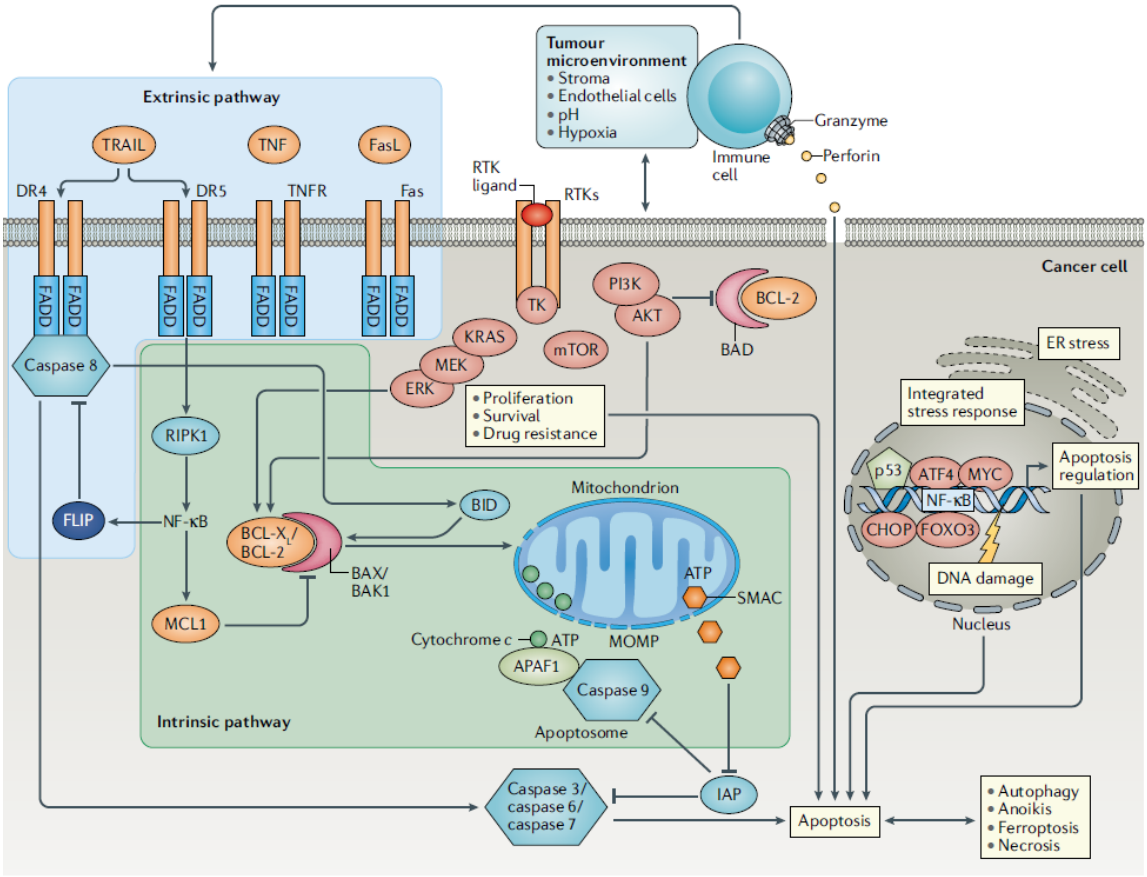


Figure 5.1.12 | Apoptosis-inducing Signaling Pathways. Apoptosis is commonly induced by BAX/BAK1-mediated cytochrome *c* release from mitochondria, which triggers caspase activity resulting in apoptosis. BAX/BAK1 are part of the BCL-2 protein family and are inhibited by other BCL-2-related proteins. These mechanisms are either stimulated through intrinsic signaling pathways, including extracellular signal-regulated kinase (ERK), AKT, or mTOR related pathways, or *via* extrinsic stimuli communicated *via* TRAIL, TNF, or Fas receptors. The DDR or cellular stress responses as well as immune cells from the tumor microenvironment are also common initiators of apoptosis. Taken from Carneiro *et al.*, *Nat. Rev. Clin. Oncol.* **2020**, *17* (7), 395-417[66].

Many of the described processes eventually initiate transcription factors. These factors are DNA-binding proteins, which orchestrate transcription of their target genes *via* recruitment of RNA polymerase II (RNAPII)[73]. Since these key regulators of gene expression are involved in a variety of different diseases, including cancer[73], it was therefore hypothesized that targeting transcription factors would be the most straightforward way to treat cancer[74]. That is, upstream signaling pathways are commonly highly complex and interconnected, which complicates targeting of these pathways due to potential pathway redundancies or a multitude of affected downstream targets that could lead to unintended side-effects[74]. Transcription factors are either steroid receptors, including estrogen and androgen receptors (ER and AR, respectively), nuclear proteins, such as Jun, or cytoplasmic proteins, like the STATs[74]. While inhibitors of steroid receptors have been

successfully applied in the clinic[75], other transcription factors have proven more challenging to target[74].

Although most transcription factors lack a distinct targetable binding pocket for small molecule modulation, different strategies have been developed to control transcription factor function, involving PROTACs, PPI inhibitors, transcription factor expression and degradation regulators, and DNA-binders[75]. Some examples of important transcription factor signaling cascades in cancer include:

(1) *STAT* which are linked with Janus kinases (JAKs) signaling and are regulated by extracellular stimuli such as cytokines, hormones, and growth factors[76]. Small molecules targeting the Src homology 2 (SH2) domain of STATs interfere with phosphorylation and thus activation of these transcription factors[76].

(2) Nuclear factor κB (*NF-κB*) is part of the response to inflammation triggered by pathogens, interleukin 1 (IL-1), or tumor necrosis factor-α (TNF-α)[74]. Strategies to interrupt NF-κB-DNA binding have been developed[77].

(3) *β-Catenin* is activated by wingless-related integration site (Wnt) protein[78] signaling (cell differentiation/development) *via* Frizzled receptors, which inhibit β-catenin inhibiting kinases[74]. Different PPI inhibition as well as degradation strategies have been investigated to target this transcription factor[79].

(4) *Notch* receptors undergo conformational changes and cleavage *via* a disintegrin and metalloproteinase (ADAM) 10/17[80] upon ligand binding and are important communicators of cell-cell interactions[74]. Strategies to inhibit Notch signaling involve prevention of DNA-binding, as well as some degradation-inducing compounds[77].

(5) Glioma-associated oncogene (*GLI*) is a nuclear transcription factor of the Hedgehog pathway, which is regulated by extracellular stimuli (cell-cell communication) and induced *via* the activation of patched (ptch) or smoothened (smo) receptors[81]. Inhibitors targeting the DNA-binding site of GLI proteins have been developed[77].

(6) The tumor protein 53 (*p53*) transcription factor is commonly downregulated by the mouse double minute 2 (MDM2) E3 ubiquitin ligase, which is inhibited by cellular stress signaling pathways leading to upregulated p53[82]. In cancer, the *TP53* gene is most frequently mutated to inhibit its tumor suppressing effect[82]. Hence, direct targeting strategies include small molecules that are restoring activity of mutant p53[83].

(7) Forkhead box (*FOX*) proteins are a large family of transcription factors, which are regulated by the PI3K-AKT and the Ras-Raf-MEK-ERK pathways[84]. Inhibitors interfering with FOX-DNA binding have been reported[77].

(8) *MYC* proteins are transcription factors for numerous target proteins by formation of a heterodimer with MYC associated factor X (MAX)[85]. While transcription of MYC proteins was shown to be induced by the bromodomain-containing 4 (BRD4) protein[85], MYC activity is further regulated by a variety of mechanisms, e.g. the Ras-MEK-ERK pathway, CDKs, glycogen synthase kinase 3 (GSK3)[86], PLK1, Aurora kinases, and others[87]. While mainly indirect targeting strategies have been reported, some peptide inhibitors of the complex formation with MAX are being investigated in clinical trials[85].

Even though transcription factors were thought to be straightforward targets for cancer treatments[74], transcription factors are actually regulated by a variety of different processes, including other transcription factors, themselves. For example, transcription of Notch2 can be initiated by STAT3[80], and while NF-κB triggers p53 expression, it also regulates p53 activity *via* MDM2 and is itself regulated by p53[88]. Besides transcription factors, transcription of target genes is also regulated by epigenetic mechanisms such as DNA methylation, histone modifications, and others[89], whose regulators, including HDAC and DNA methyltransferases (DNMTs), are attractive targets for cancer therapy[90].

Protein expression is further regulated by translational mechanisms. An important role in initiating translation play eukaryotic translation initiation factors (eIFs), which build a large complex with mRNAs and recruit ribosomes for protein translation[91]. One of the main intermediate complexes responsible for ribosome recruitment is the eIF4F complex and is regulated by mTOR complex 1 (mTORC1), which initiates translation *via* various mechanisms[91]. Furthermore, MAPK-interacting kinases (MNKs) downstream of the MEK-ERK and the p38 MAPK pathways are involved in the eIF4F complex formation, providing several pathways for targeted strategies to intervene in upregulated translational processes in cancer[91]. For instance, mTOR and PI3K (upstream of mTOR) inhibitors, MEK and MNK inhibitors, as well as eIF complex-targeting agents have been reported[91].

While the described signaling cascades are obviously main regulators of cell proliferation, hence important targets in cancer therapy, there are multitudinous other options that have been or are currently explored as anticancer targets. For example, similar to the antimetabolites of traditional chemotherapeutics, efforts in developing agents that target cancer-specific alterations in metabolism cascades have been undertaken[92,93]. Moreover, targeting the tumor microenvironment[94], senescence[95], alternative splicing events[96], inflammation processes[97], the nucleolus[98], angiogenesis[99], and many other approaches to targeting cancer can be found in the literature.

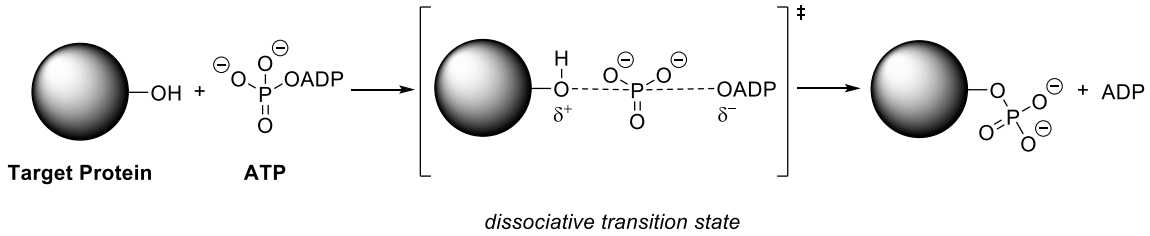
Even though this chapter provides only a brief overview of selected mechanisms that are interesting targets for cancer chemotherapeutics, it demonstrates the complexity and connectivity of regulatory pathways in human healthy and cancerous cells. The overview also highlights the advantage of multitargeting polypharmacology or synthetic lethal approaches in cancer therapy over single agent treatment[100].

5.1.3 Targeting Kinases with Small Molecules

Kinases are enzymes that catalyze the transfer of a phosphoryl group (PO_3^{2-}) from adenosine triphosphate (ATP) to the hydroxy group of a target structure *via* a dissociative transition state, which commonly involves stabilizing Mg^{2+} or Mn^{2+} cations[101,102] (Scheme 5.1.2). For protein kinases, target structures are generally comprised of serine/threonine (Ser/Thr) or tyrosine (Tyr) phosphorylation sites on other proteins[101,102], as well as intramolecular structures[103]. Protein phosphorylation and dephosphorylation are essential post-translational mechanisms to regulate protein stability, function, cellular localization, and interaction with other proteins and affect a vast majority of the human proteome[104]. Hence, kinases have been implicated in almost all cellular processes in health and disease[102]. For example, kinase activity has been linked to the DDR[105], extracellular signaling cascades (e.g. *via* receptor kinases)[106], the secretory pathway[107], cellular stress signaling, cell metabolism[108], apoptosis, cell-cycle regulation, autophagy[109,110], and other functions. Thus, dysregulated kinase activity is involved in the development of numerous

diseases, including cardiovascular conditions[111,112], neurological diseases, such as Parkinson's[113,114], and cancer[115], and poses therefore an attractive target for potential therapeutics.

Scheme 5.1.2 | Proposed Phosphorylation Mechanism of Protein Kinases[a]



[a] Adapted from Wang *et al*., *Methods Enzymol.* **2014**, *548*, 1-21[101].

Structural Features of Kinases

Based on sequence similarity, the over 500 kinases of the human kinome are generally divided into eukaryotic protein kinases (ePKs), which include almost 500 kinases, and atypical protein kinases (aPKs), e.g. lipid kinases[116]. Despite differences in their primary structure, almost all kinases share similar structural features[116] (Figure 5.1.13). The overall structure of kinases is characterized by a β-sheet-rich N-lobe and an α-helices-rich C-lobe, which are connected *via* a short, disordered linker[116]. The ATP-binding pocket is commonly situated at the intersection between both domains[116]. In addition, kinases share several characteristic motifs:

(1) The *hinge* region within the linker between the N- and the C-lobe, which directs ATP binding as well as binding of most kinase inhibitors[116]. In certain kinases, the hinge region hosts a specific amino acid at the entrance of a hydrophobic pocket commonly accessed by kinase inhibitors, which is hence referred to as the "*gatekeeper*"[117]. Gatekeeper mutations blocking access to this pocket thus lead to inhibitor-resistant kinases[117].

(2) The (glycine-rich) *P-loop* of the catalytic domain, which is involved in phosphate binding of ATP[117]. While the P-loop fulfills the same function across the kinome, aPKs do not exhibit the characteristic increase in glycines within this motif, as is common among ePKs[116].

(3) The *DFG* (Asp-Phe-Gly; aspartic acid – phenylalanine – glycine) motif, which plays an important role in the activation of a kinase[116]. The Asp of the DFG sequence is responsible for coordinating the bivalent metal cation (Mg^{2+}/Mn^{2+}) in its "in" conformation, while a kinase is considered inactive with a DFG-"out" conformation, meaning that the motif faces towards the outside of the ATP-binding pocket[117].

(4) The *αC-helix* at the N-terminus, which contributes to ATP coordination within the binding pocket[117]. It carries a glutamate, which forms a salt bridge with the catalytic lysine within the N-lobe in active kinases but faces away from the lysine in inactive kinase conformations[117].

(5) The *HRD* (His-Arg-Asp; histidine – arginine – aspartic acid) motif is involved in catalyzing the phosphorylation process and is located within the C-lobe[116]. In aPKs, the motif is found reversed (DRH motif), while the function stays the same[116].

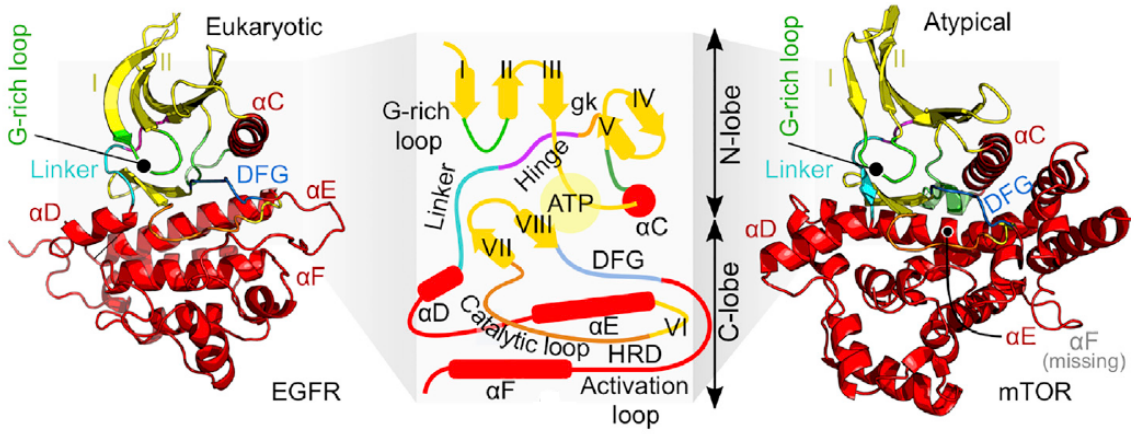


Figure 5.1.13 | Structural Comparison of ePKs and aPKs. Despite a lack of sequence homology, ePKs (eukaryotic; here: epidermal growth factor receptor (EGFR)) and aPKs (atypical; here: mTOR) are comprised of an N- and C-lobe, which form the ATP binding pocket at their intersection. Furthermore, kinases share common structural motifs, including a hinge region at the ATP binding site, a catalytic DFG motif, (glycine-rich) P-loop, HRD motif, and αC-helix. Adapted from Kanev *et al*., *Trends Pharmacol. Sci.* **2019**, *40* (11), 818-832[116].

Small Molecule-based Kinase Inhibitors – An Introduction

As described above, kinases are involved in the regulation of several cell signaling pathways and are thus exciting targets for a variety of diseases. Using small molecule modulators, kinase signaling can be inhibited *in vivo* and has been successfully applied in the clinic leading to 59 clinically approved kinase inhibitors within the past 20 years[118,119]. Considering the conserved structural features of kinases, ATP-competitive small molecule inhibitors exhibit three common binding modes[120] (Figure 5.1.14). For instance, type 1 inhibitors bind to the active kinase conformation and comprise the majority of known kinase inhibitors[120]. While this conformation is generally characterized by a DFG motif facing towards the ATP binding site (DFG-'in'), it is also important that the αC-helix is facing the ATP binding pocket as well[117]. The subtly different type 1.5 inhibitors (a subgroup of type 1 inhibitors) bind to the inactive DFG-'in' but αC-helix-'out' kinase conformation[121]. In contrast, kinase inhibitors that exhibit binding affinity towards the inactive DFG-'out' characterized kinase conformations are designated type 2 inhibitors[120].

Further mode of actions for small molecule kinase inhibitors include allosteric kinase inhibition[122]. Kinase modulators binding to allosteric pockets of the kinase domain adjacent to or further away from the ATP-binding pocket are inhibitors of type 3 and 4, respectively[122]. Some kinase inhibitors target two motifs within the kinase domain and belong to the type 5 kinase inhibitors[123]. In recent years, two more inhibitor types have been defined, which describe small molecules affecting protein regions outside the kinase domain[122]. For instance, kinase inhibitors of type 6 bind to allosteric sites outside the kinase domain, including pseudokinase domains (pseudokinases are defined as kinase-like proteins that do not exhibit catalytic activity, thus mainly performing scaffolding, cellular localization, and complex formation functions[124]), while type 7 inhibitors are defined as small molecules which effect receptor tyrosine kinase activity by binding extracellular protein domains[122].

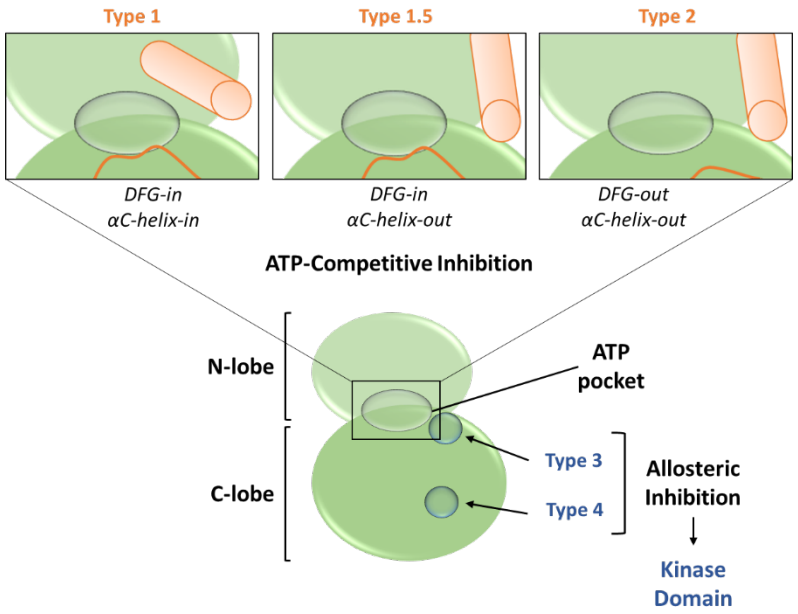


Figure 5.1.14 | Overview of Main Kinase Inhibitor Types. Since kinases generally share common structural features, kinase inhibitors are categorized by their targeted binding site. ATP-competitive inhibitors are divided into three types, where type 1 inhibitors bind to the kinase active conformation (DFG-'in' and αC-helix-in), type 2 inhibitors bind to the kinase inactive conformation (DFG-'out' and αC-helix-out). Some kinase inhibitors target the inactive kinase conformation with a DFG-'in' motif, which are therefore referred to as type 1.5 kinase inhibitors. Alternative kinase inhibitor binding pockets are situated either adjacent to (type 3) or further away from (type 4) the ATP binding site. Notably, these allosteric binding sites are part of the kinase domain.

Even though the great majority of approved kinase inhibitors bind their target proteins reversibly, the development of irreversible inhibitors which form a covalent bond with the kinase is an active area of research, culminating in the FDA (U.S. Food and Drug Administration) approval of several drugs, such as ibrutinib and osimertinib[118]. Covalent inhibitor-kinase complexes are achieved by the introduction of an electrophilic warhead, including acrylamide derivatives, haloketones, or sulfonyl-based structures (Figure 5.1.15), which react with nucleophiles such as cysteines (Cys), lysines (Lys), Asp[125], and others. Since most reported covalent kinase inhibitors, including the first FDA-approved agents, carry acrylamide derivatives as their reactive warheads[125], irreversible inhibitors commonly bind kinases *via* a Michael-addition mechanism[126] (Scheme 5.1.3).

Covalent kinase inhibitors are generally derived from reversible kinase binders, which are within reach of a nucleophilic amino acid, e.g. a conserved Cys[126]. Using structure-guided design, rational positioning of a covalent warhead moiety within the reversible scaffold can yield an irreversible kinase binder[126]. Notably, this approach has some advantages over reversible kinase inhibition. For example, while the additional structural feature of the kinase needed to result in potent kinase binding of the small molecule often improves selectivity, the irreversible nature of a covalently binding compound also commonly results in prolonged inhibitory effects and a general increase in potency compared to the reversible analogs[126]. However, the reactivity of the warheads, as well as the covalent binding mechanism itself, can lead to unwanted side effects[127], which might also be consequences of non-kinase targets[128].

Generally, kinase inhibitors are essential tools in investigating kinases and their contribution to cell and disease biology as well as in validating kinases as potential targets for novel therapeutics[126]. While small-molecule kinase inhibitors were historically most commonly developed for the treatment of cancerous diseases, recent advances have been made in the fields of immune and degenerative diseases, leading to novel applications and drug discovery campaigns for kinase inhibitors[118].

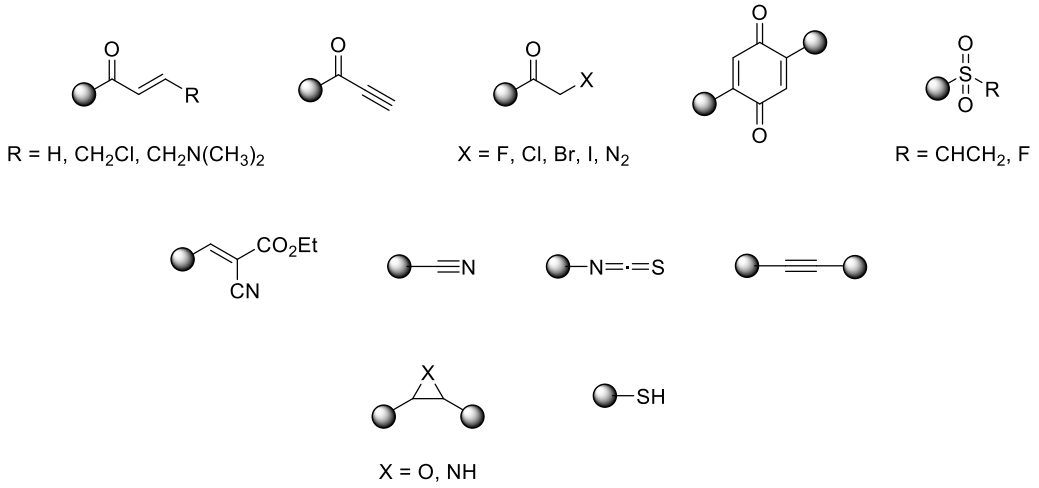
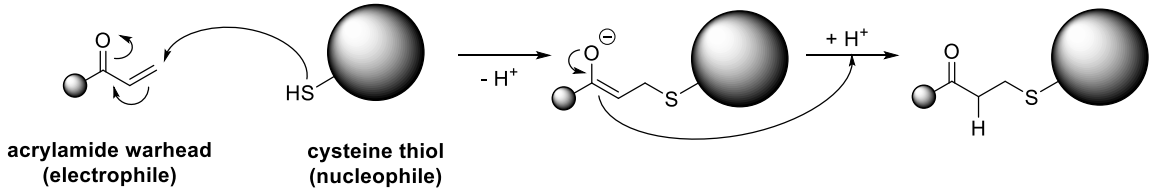


Figure 5.1.15 | Common Electrophilic Warheads of Covalent Kinase Inhibitors. An overview of typical warhead structures found among covalent kinase inhibitors. Adapted from Liu *et al.*, *Chem. Biol.* **2013**, *20* (2), 146-159[126].

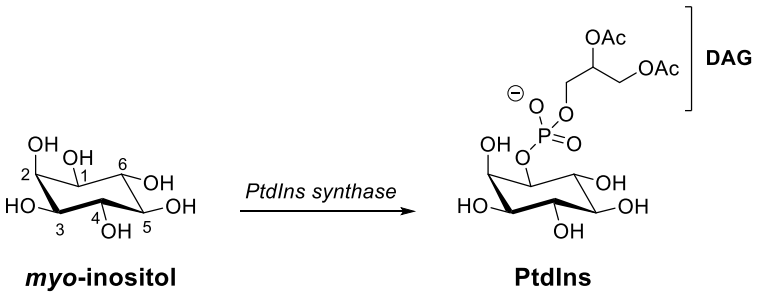
Scheme 5.1.3 | Michael-addition Mechanism of Covalent Kinase Inhibitors



5.1.4 The Lipid Kinase Phosphatidylinositol 5-Phosphate 4-Kinase (PI5P4K)

Intra- and extracellular membranes pose essential platforms for signaling cascades in human cells[129]. That is, membrane lipids are central regulators in the recruitment and allosteric manipulation of soluble proteins, and orchestrate cellular processes such as membrane trafficking[129]. The key lipid molecules of these mechanisms are phosphatidylinositols (PtdIns)[129], which are comprised of diacylglycerol (DAG) and an inositol moiety connected *via* a phosphate group[130] (Scheme 5.1.4). Inositol (here: D-*myo*-inositol) is generated from glucose-6-phosphate and is coupled with the lipid component by PtdIns synthase[131]. Interestingly, while inositol and its derivatives are ubiquitously found in eukaryotes, only very few bacteria take advantage of these lipid-based signaling molecules[131]. The solvent exposed inositol unit of PtdIns is further decorated with phosphates by a network of lipid kinases, which in concert with phosphatases control the phosphorylation pattern thus the function of these signaling molecules[130].

Scheme 5.1.4 | Structure and Formation of Phosphatidylinositol (PtdIns)



Phosphatidylinositol Phosphates (PIPs) – Lipid-based Regulators of Cell Function

The first observation of the signaling function of PtdIns phosphates (PIPs) was in the late 1970s/early 1980s, when the receptor-initiated cleavage of phosphatidylinositol-4,5-bisphosphate (PI-4,5-P_2) by phospholipase C into the second messengers DAG and inositol-1,4,5-trisphosphate (IP_3), with the latter triggering intracellular Ca^{2+}-release, was investigated[132,133]. In the following decades, a vast variety of roles for the different kinds of PIPs, in addition to serving as precursors of other second messengers, were uncovered[133]. This was especially owed to the discovery of PIP-binding domains in the early 2000s, providing a mode of action for various PIP functions as they control protein complex formation, proximity, or even allosteric activation of target proteins[134,135].

In human cells, PtdIns is typically found phosphorylated at positions 3, 4, and 5, resulting in seven distinct PIPs[136] (Figure 5.1.16). Since their function is based on cellular localization and recruitment of proteins for downstream effects, an important feature to understand PIP function is the clustering of PIPs in specific cellular membrane compartments[136] (Figure 5.1.17) and is summarized in Table 5.1.2. For example, phosphatidylinositol 3-phosphate (PI-3-P) is mostly found in the membranes of early endosomes as well as autophagosomes at the ER[136]. PI-3-P is then converted to phosphatidylinositol-3,5-bisphosphate (PI-3,5-P_2), marking late endosomes, lysosomes, and their trafficking back to the Golgi[136]. At the Golgi, phosphatidylinositol 4-phosphate (PI-4-P) is commonly found to regulate vesicle trafficking to the plasma membrane[136]. Here, PI-4,5-P_2 dominates a multitude of processes, including clathrin-mediated endocytosis together with phosphatidylinositol-3,4-bisphosphate (PI-3,4-P_2)[136]. Phosphatidylinositol-3,4,5-trisphosphate (PIP_3) is also located at the plasma membrane and is involved in processes such as early phagosome formation or clathrin-independent endocytosis[136].

PIPs are bound by proteins to direct and regulate their function at specific membrane sites within the cell, with over 50 protein kinases exhibiting respective PIP-binding domains[129]. One such domain, the Pleckstrin homology (PH) domain, was identified in the context of PI-4,5-P_2-binding by phospholipase C[135]. PH domains have thus been intensely studied and can be found in protein kinases, guanine nucleotide exchange factors, guanosine triphosphate hydrolase (GTPase)-activating proteins, and a variety of other proteins, while PIP-selectivity of PH domains can vary[134]. Interestingly, the structurally similar phosphotyrosine binding (PTB) domains have been also shown to bind PI-4,5-P_2, e.g. in the context of the clathrin adaptor protein disabled homolog 1 (Dab1)[134]. FERM (acronym for: 4.1 Protein, erzin/radixin/moiesin) domains pose another example of PH-like protein domains and were first identified in proteins mediating the connection of the cytoskeleton

with the plasma membrane[134]. While these PIP-binding domains can show selectivity for different PIPs, FYVE (acronym for: Fab-1 (1-phosphatidylinositol 3-phosphate 5-kinase), YGL023 (fungal protein), Vps27 (vascuolar protein sorting-associated protein 27; yeast), and EEA1 (early endosome antigen 1)) and phagocytic oxidase (phox) homology (PX) domains have been identified as PI-3-P-specific protein structures[134]. PI-4,5-P_2-specific domains include the AP180 N-terminal homology (ANTH) and the epsin N-terminal homology (ENTH) domains and play a role in clathrin adapter proteins[134].

Table 5.1.2 | Overview of PIPs and Their Cellular Locations[a]

PIP	Abundance	Location	Function
PtdIns	10-20 mol% of phospholipids	ER (biosynthesis), Golgi, early endosomes, plasma membrane	(PIP precursor)
PI-3-P	~0.1-0.5% of PtdIns	early endosomes, autophagosomes	early endosome formation
PI-4-P	~2-5% of PtdIns	Golgi, endosomes, (plasma membrane)	Golgi trafficking, endosomes
PI-5-P	~0.02-0.05% of PtdIns	*unclear* (early endosomes, Golgi, plasma membrane, nucleus)	cytoskeleton regulation, stress signaling
PI-3,4-P_2	~0.04-0.25% of PtdIns (~2-5% of PtdIns after stimulated production)	ER, endosomes, plasma membrane	clathrin-mediated endocytosis
PI-4,5-P_2	~2-5% of PtdIns	plasma membrane (Golgi, endosomes)	clathrin-mediated & clathrin-independent endocytosis
PI-3,5-P_2	~0.02-0.05% of PtdIns	late endosomes/lysosomes	endosome maturation/lysosome formation, trafficking of endosomes to Golgi
PIP_3	~0.02-0.25% of PtdIns	plasma membrane (enriched at site of chemotactic stimuli), phagosomes	clathrin-independent endocytosis, phagosome formation

[a] This information was assembled from Hammond *et al*., *Biochim. Biophys. Acta* **2015**, *1851* (6), 746-758[135] and Billcliff *et al*., *Biochem. J.* **2014**, *461* (2), 159-175[136].

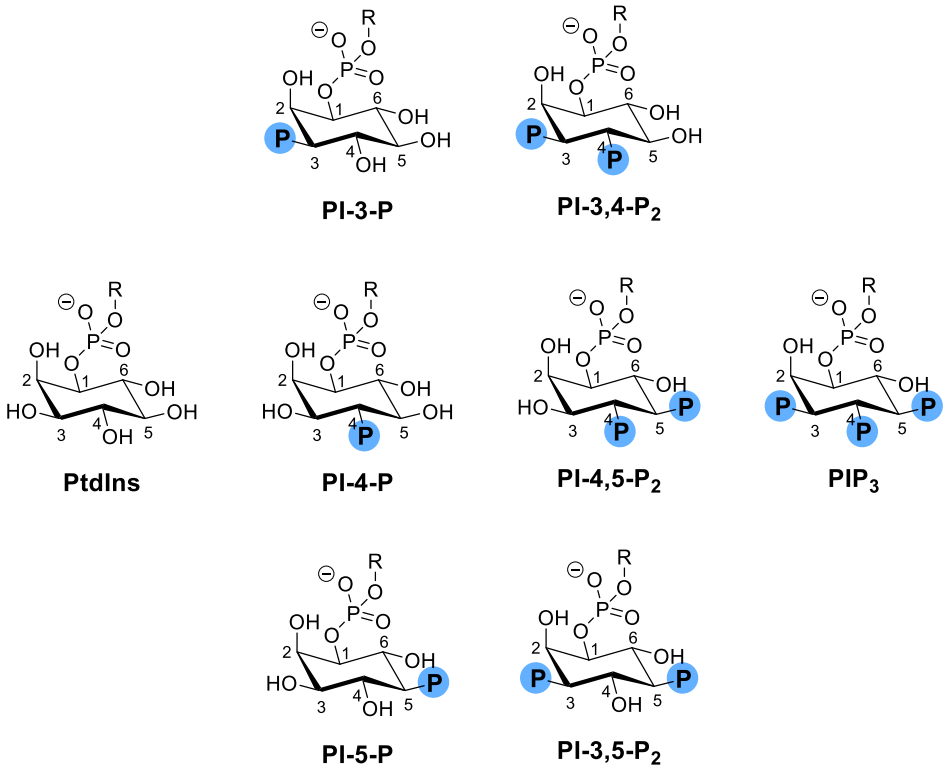


Figure 5.1.16 | Phosphorylation Patterns of PIPs. Shown are the structures of the seven distinct PIPs found in human cells. Phosphate groups are indicated with a blue circled "P".

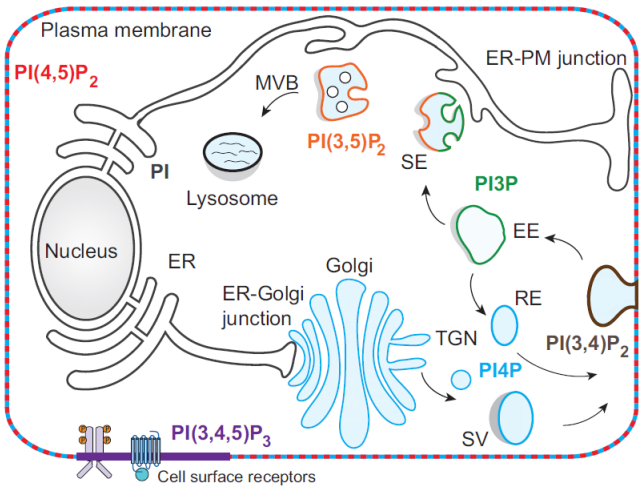


Figure 5.1.17 | Cellular Localization of PIPs. Due to their specific functions, local accumulations of PIPs at specific cellular membrane compartments are observed. For example, PI-4-P (light blue) is mainly found at the Golgi and secretory vesicles (SVs). PI-3,4-P_2 (black) mediates endocytotic processes. In the following, PI-3-P (green) is found in early endosomes and PI-3,5-P_2 (orange) facilitates maturation to lysosomes. At the plasma membrane, PI-4,5-P_2 (red) is produced primarily from PI-4-P and can be further phosphorylated to form PIP_3 (purple; here: PI-3,4,5-P_3). Adapted from Burke, *Mol. Cell* **2018**, *71* (5), 653-673[137].

Phosphatidylinositol (Phosphate) Kinases and Phosphatases Regulate PIP Function

While PIPs are bound by their effector proteins, including ion channels, scaffolding, cytoskeletal, and membrane trafficking-mediating proteins, the main regulators of PIP function and distribution are PtdIns (phosphate) kinases and phosphatases[130]. PtdIns kinases have been known for over 50 years; however, most of our knowledge on the significance of PtdIns kinases and their activity has only been generated within the past 30 years[138].

In human cells, 19 kinases and 28 phosphatases orchestrating PIP function have been identified[130] (Figure 5.1.18). Even though this chapter focuses on PIP kinases, phosphatases and their functions are briefly presented for context. Structurally, the PtdIns (phosphate) kinases are divided into three main groups: PI3Ks and type III phosphatidylinositol 4-kinases (PI4Ks), sharing an α-helices-rich region, type II PI4Ks, which are lacking these and other specific domains apart from the kinase domains, and PIP kinases (PIPKs), exhibiting PIPK-specific dimerization domains (except for PIKfyve)[137] (Figure 5.1.19). Some key features of PtdIns kinases are summarized in Table 5.1.3.

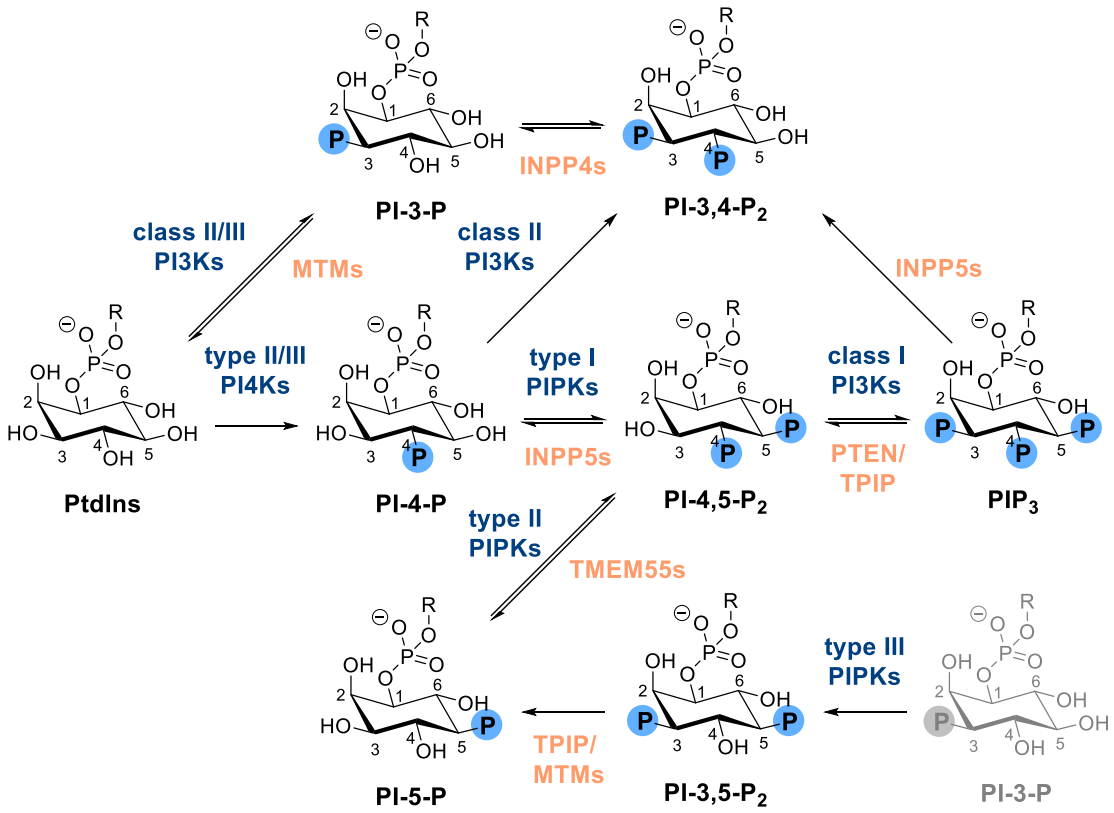


Figure 5.1.18 | The PtdIns Network. Overview of important PtdIns kinases and phosphatases regulating the biosynthesis of PIPs. Phosphate groups are indicated with a blue circled "P". Kinases are shown in dark blue, while phosphatases are depicted in orange.

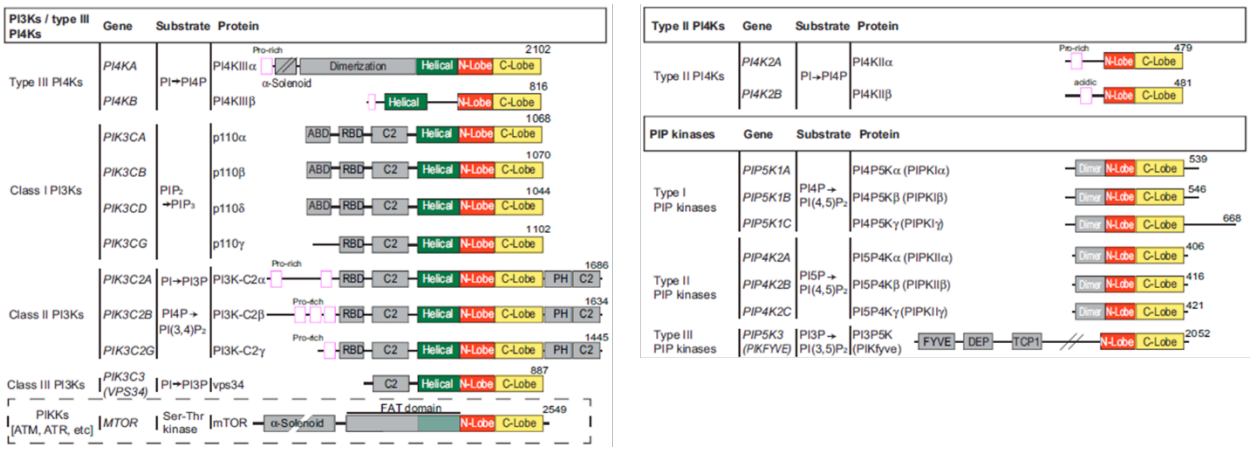


Figure 5.1.19 | Structural Relationship of PtdIns Kinases. PtdIns kinases can be divided into three main groups based on their structural features. PI3Ks and type III PI4Ks share a common core structure, including an α-helix-rich domain. Interestingly, mTOR is part of the PIKK family. Type II PI4Ks do not share this "helical" domain, thus pose a separate group. Finally, PIPKs (with the exception of PIKfyve) exhibit a characteristic dimerization domain, thus forming the third group of PtdIns kinases. Adapted from Burke, *Mol. Cell* **2018**, *71* (5), 653-673[137].

Table 5.1.3 | Mammalian PtdIns Kinases[a]

PtdIns Kinase	Expression[b]	Function[c]
GROUP 1		
Class IA PI3Ks		
PI3Kα (p110α)	ubiquitous	angiogenesis, Akt pathway activation, insulin signaling
PI3Kβ (p110β)	ubiquitous	Akt pathway activation, insulin signaling
PI3Kδ (p110δ)	immune cells	TCR/BCR signaling, leukocyte chemotaxis, mast cell function
Class IB PI3Ks		
PI3Kγ (p110γ)	immune cells	leukocyte and mast cell function
Class II PI3Ks		
PIK3C2α	ubiquitous	clathrin-med. membrane trafficking, (insulin signaling)
PIK3C2β	ubiquitous	cell migration, clathrin-med. endocytosis
PIK3C2γ	prim. liver	unclear
Class III PI3Ks		
PIK3C3	ubiquitous	endosome/lysosome function, autophagy, innate immune resp.

PtdIns Kinase	Expression[b]	Function[c]	
Type III PI4Ks			
PI4K IIIα (PI4KA)	prim. brain	Ca^{2+}-signaling (plasma membrane)	
PI4K IIIβ (PI4KB)	ubiquitous	membrane trafficking (Golgi)	
GROUP 2			
Type II PI4Ks			
PI4K IIα (PI4K2A)	ubiquitous	vesicle trafficking, (EGFR turnover, phagocytosis)	
PI4K IIβ (PI4K2B)	ubiquitous	TCR signaling, (phagocytosis)	
GROUP 3			
Type I PIPKs			
PI4P5Kα (PIPK Iα)	ubiquitous	chemotaxis	clathrin-med. endocytosis (regular: α; stimulated: β, γ), actin dynamics, Ca^{2+}-signaling
PI4P5Kβ (PIPK Iβ)	ubiquitous	(nuclear stress response)	
PI4P5Kγ (PIPK Iγ)	prim. brain	-	
Type II PIPKs			
PI5P4Kα	ubiquitous	PI-5-P conversion (general)	
PI5P4Kβ	ubiquitous	PI-5-P conversion (nucleus), insulin signaling	
PI5P4Kγ	prim. kidney	unclear	
Type III PIPKs			
PIKfyve (PIPK III)	prim. muscle fat cells	membrane dynamics, insulin signaling	

[a] Based on information from Sasaki *et al*., *Prog. Lipid Res.* **2009**, *48* (6), 307-343[130].
[b] prim., primarily.
[c] BCR, B-cell receptor; TCR, T-cell receptor.

The PI3K family is a well-studied PtdIns kinase family and can be divided into three main classes[137]. Class I PI3Ks are primarily responsible for the formation of PIP_3 from PI-4,5-P_2 at the plasma membrane, stimulating different intracellular pathways involved in cell growth, metabolism, and survival, while also being responsible for immune cell function[137]. The class I PI3Ks are further divided into two subgroups based on their respective regulatory subunits[137]. PI3K subclass IA is comprised of the isoforms PI3Kα, β, and δ, while PI3Kγ is sole member of the IB subclass[137]. Generally, class I PI3Ks are stimulated *via* receptor tyrosine kinases, GTPases (Ras-based), and G protein-coupled receptor signaling[137]. Notably, instead of Ras, PI3Kβ is activated by the Rho GTPases[137].

Furthermore, while PI3Kα and β are ubiquitously expressed, PI3Kγ and δ are the main class I PI3Ks in immune cells[137].

Class II PI3Ks are represented by the isoforms PIK3C2α, β, and γ and are mainly converting PtdIns into PI-3-P as well as PI-4-P into PI-3,4-P_2[137]. Interestingly, class II PI3Ks contain a PX domain, recruiting them to PI-4,5-P_2-rich membranes[137]. There, they function as regulators of clathrin-mediated endocytosis, cilium function, as well as insulin signaling[137]. It was shown that PI-4,5-P_2 and clathrin act in concert to activate PIK3C2α to mediate endocytosis[137], while the function of PIK3C2β and γ are still poorly understood[130]. Similar to the class II PI3Ks, class III PI3Ks catalyze the conversion of PtdIns into PI-3-P and consist of the protein PIK3C3[137]. Due to its ability to form different complexes with varying membrane-binding domains, PIK3C3 is either involved in autophagosome formation or endosome trafficking[137].

Besides the PI3Ks, the first group of PtdIns kinases also includes type III PI4Ks, which are comprised of the two isoforms PI4K IIIα and β[137]. PI4Ks produce PI-4-P from PtdIns with PI4K IIIα being mainly located at the plasma membrane, providing PI-4-P for PI-4,5-P_2 synthesis, and PI4K IIIβ regulating lipid distribution, membrane trafficking, cytokinesis, and lysosomal functions at the Golgi[137]. Interestingly, these PtdIns kinases are commonly activated by viruses in course of viral replication[137].

The second group of structurally distinct PtdIns kinases are type II PI4Ks, including the isoforms PI4K IIα and β[137]. Similar to the type III PI4Ks these enzymes catalyze the formation of PI-4-P from PtdIns[137]. However, due to their structure including a lipidation site, type II PI4Ks are permanently associated with intracellular membranes and regulate endosomal exocytosis, EGFR turnover, and some signaling pathways such as the Wnt cascade[137].

The third group of PtdIns kinases are comprised by PIPKs and can be further divided into three subgroups based on their substrate lipids[137]. Type I PIPKs transform PI-4-P to PI-4,5-P_2 and are hence referred to as PI4P5Ks, including the isoforms PI4P5Kα, β, and γ[137]. Similarly, type II PIPKs form PI-4,5-P_2 from PI-5-P and are thus termed PI5P4Ks, with isoforms α, β, and γ[137]. Notably, PI-4,5-P_2 is mainly produced *via* type I PIPKs, while type II PIPKs are thought to act mainly through controlling the cellular PI-5-P pool or producing small, local PI-4,5-P_2 clusters[137]. Type I PIPK activity is stirred by a variety of different stimuli, e.g. GTPases, which are mostly directed to activate specific PI4P5K isoforms[137]. Even though type I and type II PIPKs have been implicated to regulate cellular processes, such as migration, adhesion, cell division, and polarity, with type II PIPKs also being involved in autophagy mechanisms and cellular metabolism, especially PI5P4K (type II PIPKs) function is still mostly unclear[137]. Notably, the type II PIPK isoforms are able to form homo- and heterodimers with each other, and while PI5P4Kα and β (PI5P4Kβ preferring GTP as a substrate instead of ATP) exhibit decent to strong kinase activity, PI5P4Kγ shows no detectable kinase activity[130,137]. Finally, type III PIPKs are comprised by the enzyme PIKfyve (or also: phosphatidylinositol 3-phosphate 5-kinase; PI3P5K), which is characterized by a FYVE domain[130]. FYVE domain-dependent PI-3-P binding leads to production of PI-3,5-P_2 and has been associated with membrane dynamics and insulin signaling[130].

Due to the many functions of PtdIns kinases in regulating cellular signaling cascades and pathways, dysregulation of their function has been involved in many diseases[137]. For instance, Ras-activated PI3Kα and its role in tumorigenesis and angiogenesis processes has

been target for a variety of oncogenic therapies, while also other class I PI3Ks have been implicated as key drivers in cancer, immunological conditions, and inflammation[137]. Type III PI4Ks pose potential targets for viral infections, while type II PI4Ks might play a role in the development of neurodegenerative diseases[137].

Notably, while PtdIns kinases produce PIPs and regulate their function, a set of PIP phosphatases are also involved in these processes and need to be considered to fully understand the overall cellular lipid network[130] (Figure 5.1.18). For example, the well-studied antagonist of class I PI3Ks is the phosphoinositide 3-phosphatase PTEN, which dephosphorylates PIP_3 to form PI-4,5-P_2[130]. The transmembrane phosphoinositide 3-phosphatase and tensin homolog 2 (TPIP) is able to dephosphorylate PIP_3 and PI-3,5-P_2 to form PI-4,5-P_2 and PI-5-P, respectively, while the family of myotubularins (MTMs) are mainly responsible for the dephosphorylation of the 3-position of PI-3-P and PI-3,5-P_2[130]. Inositol polyphosphate-4-phosphatases (INPP4s) and transmembrane protein 55s (TMEM55s) are phosphoinositide 4-phosphatases and turn PI-3,4-P_2 into PI-3-P and PI-4,5-P_2 into PI-5-P, respectively[130]. Phosphoinositide 5-phosphatase activity is found among type II, III, and IV inositol polyphosphate-5-phosphatases (INPP5s), including enzymes such as synaptojanins (SYNJs), OCRL1 (from lowe oculocerebrorenal syndrome protein), Ski-interacting protein (SKIP), and SH2 domain-containing inositol phosphatases 1/2 (SHIP1/2), primarily being responsible for the formation of PI-4-P from PI-4,5-P_2 and PI-3,4-P_2 from PIP_3[130].

The Phosphatidylinositol 5-Phosphate 4-Kinases (PI5P4Ks)

The type II PIPKs – in the following referred to as PI5P4Ks – were identified in the mid to late 1990s[139]. Even though PI5P4Ks were initially believed to exhibit type I PIPK (PIP 5-kinase) activity, differences in their catalytic features resulted in the differentiation between type I and II PIPKs[139]. In 1997, Rameh *et al.* showed that the PIP 5-kinase activity of PI5P4Ks was falsely assumed due to impure PI-4-P batches used to determine PIPK activity[140]. The group of Lewis C. Cantley found that, while PI5P4Ks indeed form the same product (PI-4,5-P_2) as type I PIPKs, the actual substrates of the PI5P4Ks were the PI-5-P impurities within the PI-4-P lipid samples[140]. Hence, due to the differences in substrate specificity, PI5P4Ks were re-labelled from type II PIP 5-kinases to PIP 4-kinases[140]. By the late 1990s, three PI5P4K isoforms (α, β, and γ) had been identified, while their cellular functions remained mostly unclear[139].

In 1998, Rao and coworkers solved the crystal structure of the PI5P4K isoform β, revealing key features of these lipid kinases[141]. Importantly, the group found that homodimerization of the kinase forms a basic, flat, disc-shaped structure, which could provide an explanation for the membrane association of PI5P4Ks[141] (Figure 5.1.20). Despite the overall lack in sequence homology, further structural studies revealed that PI5P4Ks share common features of protein kinases including the catalytic lysine, the catalytic aspartates of the HRD and DFG motives, as well as a glycine-rich (P-)loop[138]. Moreover, similar to protein kinases, PI5P4Ks also exhibit a substrate-determining specificity loop at the PIP-binding site, which is conserved among the PI5P4K isoforms[138]. In case of PIPKs, these specificity loops are also one determining factor for cellular localization of these kinases[138]. Interestingly, immunoprecipitation studies indicate that the PI5P4K isoforms are additionally able to form heterodimers with each other, e.g. PI5P4Kβ with α[142], and PI5P4Kγ with α[143]. This data further supports the hypothesis that

heterodimerization of the PI5P4K isoforms directs PI5P4K activity within the cell (further described below)[137,142,143].

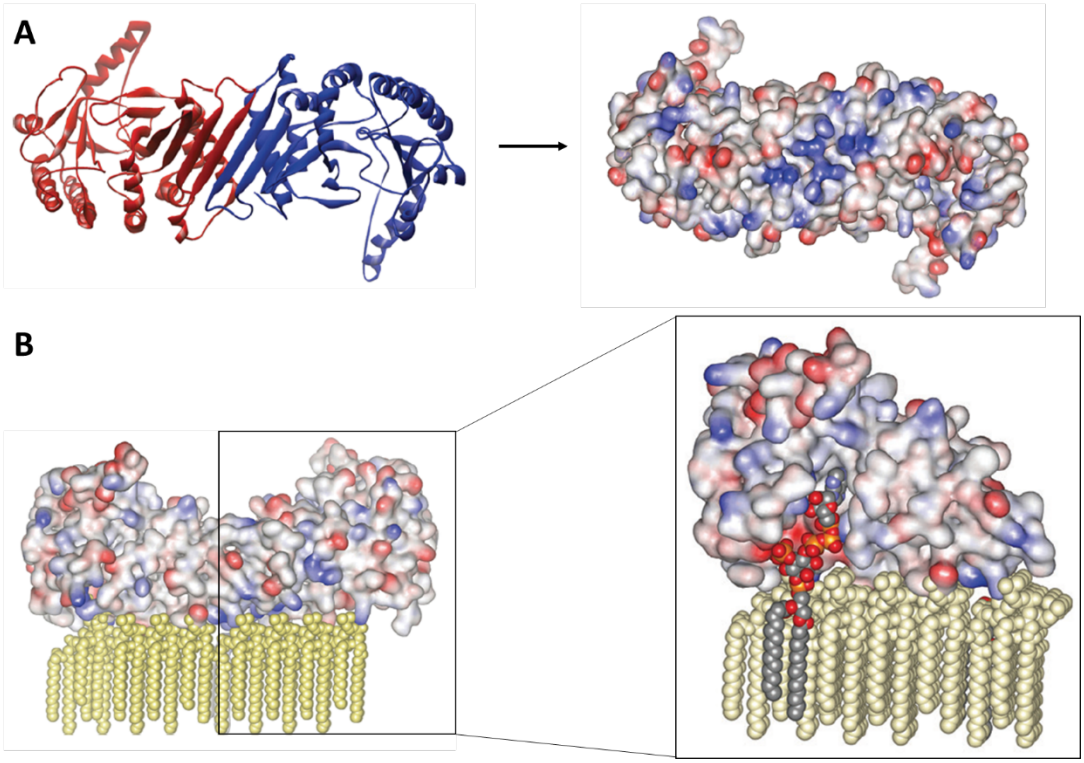


Figure 5.1.20 | Structure of the Homodimer of PI5P4Kβ[a]. A. Crystal structure of a homodimer of PI5P4Kβ (monomers indicated in red and blue). While the left side shows the ribbon structure of the homodimer, the right side shows the space filling structure of the same complex. Notably, the depiction highlights the membrane-binding site of the enzyme dimer with a basic (blue) patch in its center. **B.** A docking model of the cocrystal structure of A bound to a phospholipid layer representing a cellular membrane. On the right, ATP is docked into the binding site demonstrating substrate binding of the PI5P4K lipid kinase.
[a] Graphics adapted from Heck *et al*., *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42* (1), 15-39[138].

While PI5P4Ks generally catalyze the same phosphorylation reaction, enzymatic studies found that the PI5P4Kα isoform exhibits significantly higher enzymatic rates than the other two isoforms *in vitro*, with an approximately 2,000-fold higher kinase activity than PI5P4Kβ[144]. Surprisingly, almost no kinase activity was detected for PI5P4Kγ[145], underlining the importance of heterodimerization for the understanding of PI5P4Ks' function[140]. For example, while the most active PI5P4K isoform, PI5P4Kα, is commonly found in the cytosol, PI5P4Kγ is located at membranes of cellular vesicles and PI5P4Kβ is able to transition between cytosolic and nuclear locations[146] *via* a specific α-helical structural motif unique for the PI5P4Kβ isoform[147]. Thus, PI5P4Kα being able to heterodimerize with PI5P4Kβ and γ provides an overall explanation as to why PI5P4Kα is sometimes found to be located mostly in the nucleus or cytosol, where it can be found in soluble fractions or membrane-bound, depending on the study design[146]. Further studies suggested that heterodimerization may also serve to regulate the activity of the complexes, demonstrating that maximal activity of PI5P4Kβ occurs when PI5P4Kβ associates with

PI5P4Kα in its kinase active conformation, compared to a complex including its inactive conformation[142].

Because the lipid substrate of PI5P4Ks, PI-5-P, is only found in small pools at the plasma membrane or intracellular compartments, most of the cellular PI-4,5-P_2 lipid is produced *via* the PI-4-P intermediate by PI4P5Ks (type I PIPKs)[144]. While PI5P4K function could still technically control local PI-4,5-P_2 production, another intriguing hypothesis suggests that PI5P4Ks' main role involves the regulation of PI-5-Ps[146]. In the nucleus, PI-5-P was found to orchestrate the function of the tumor suppressor inhibitor of growth protein 2 (ING2), which mediates p53-induced cell death as a result of DNA damage[146]. It was further found that different cellular stress factors increase nuclear PI-5-P activity[146]. For instance, while PI5P4K activity, especially nuclear PI5P4Kβ activity, reduces the local PI-5-P pool, p38 MAPK or Pin1 (see Chapter 7) are able to inhibit PI5P4Ks, thus increase PI-5-P signaling[146]. Moreover, nuclear PI-5-P function has been associated with stress-induced or viral infection-mediated transcription events as well as epigenetic DNA-modifications[146]. PI-5-P has been further linked to other cellular processes including Akt signaling at the plasma membrane and cytoskeletal motility[146]. PI-5-P-induced Akt activation is proposed to be mediated *via* its inhibitory activity on PTEN as well as through activation of a variety of other phosphatases, including SHIP2, MTMs, and Akt phosphatases[146]. Furthermore, PI-5-P is able to regulate insulin-mediated translocation of the glucose transporter type 4 (GLUT4) as well as F-actin stress fiber disassembly[146].

Considering that PI5P4Ks are thus regulating PI-5-P and local pools of PI-4,5-P_2 and that PI5P4K isoforms can trigger localization and function of the formed heterodimers, PI5P4Ks have been shown to have additional functions within a cell. For example, Sumita and colleagues demonstrated that PI5P4Kβ has a unique feature of preferring GTP over ATP as a phosphate-donor[148] (Figure 5.1.21). The group showed that PI5P4Kβ acts as a GTP-sensor, contributing to orchestrate metabolic processes within the cell and potentially translating GTP levels into PI-5-P signaling for further downstream effects[148].

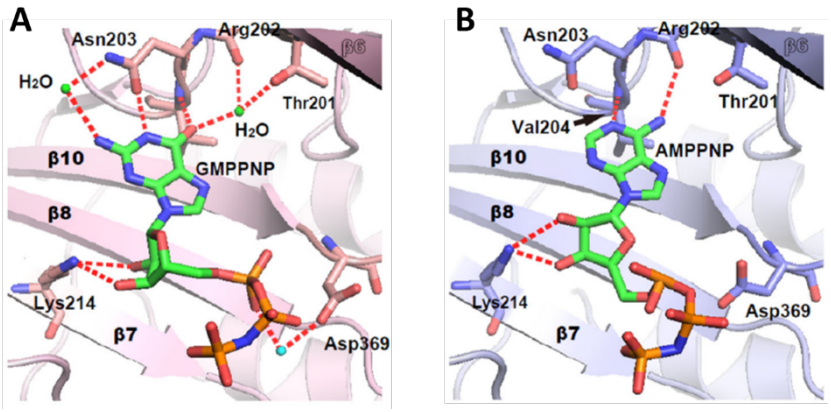


Figure 5.1.21 | Co-crystal Structures of PI5P4Kβ in Complex with GTP- and ATP-analogs. PI5P4Kβ was co-crystallized with either a GTP- (**A**) or an ATP-analog (**B**). Analysis of nucleotide binding pocket revealed an elaborate hydrogen bond network for GTP, while ATP shows less favorable interactions. Adapted from Sumita *et al*., *Mol. Cell* **2016**, *61* (2), 187-198[148].

In summary, this overview of PI5P4Ks' modes of action emphasizes the complexity of the lipid network and how the different PI5P4K isoforms can direct and regulate their kinase

function. To further complicate the understanding of PI5P4K activity, a recent study published by Lewis C. Cantley and his group identified a mechanism by which PI5P4Ks directly target PI4P5Ks and influence their PIP 5-kinase function independently of PI5P4Ks' kinase activity[149]. Thus, research and proper tools/tool compounds are necessary to further investigate the influence of PI5P4K catalytic and allosteric functions in regulating cellular processes.

Even though their function and regulatory processes are still not fully understood, PI5P4Ks have been linked to a variety of physiological processes (Figure 5.1.22). For instance, PI5P4Kα was found to be activated by photo-induced ROS signaling *via* its phosphorylation by receptor tyrosine kinases, thus implicating a role of PI5P4Kα in retina function[150]. Knockdown studies in zebrafish suggest that PI5P4Kα is further essential for embryonic development[151]. Recent studies have also linked PI5P4Kα function to lysosomal regulation of cholesterol trafficking, which is potentially facilitated by local PI-4,5-P_2 pools at the lysosome-peroxisome interface[152]. Notably, PI5P4Kα and β activity have been linked to autophagy and fusion of autophagosomes with lysosomes under nutrition-starved (metabolic stress) conditions *in vivo*, likely due to PI-4,5-P_2 production at respective membranes[153].

Lamia *et al.* reported that PI5P4Kβ$^{-/-}$ mice exhibit increased sensitivity towards insulin signaling, while generally being of lower body weight than control mice[154]. Surprisingly, it was discovered that PI5P4Kβ, while catalyzing PI-4,5-P_2 synthesis, actually promotes PIP_3 dephosphorylation at the 5-position, thus leading to a decrease in PIP_3-mediated Akt activation following insulin signaling[155]. Recent studies elucidated that these insulin-dependent mechanisms are mTORC1-independent; however, require PI5P4Kβ catalytic activity pointing towards a PI-5-P-mediated mode of action[156]. In contrast, another recent study suggests that the hypersensitivity towards insulin in PI5P4Kβ$^{-/-}$ mice is due to allosteric effects of PI5P4Ks[149], that form complexes with PI4P5Ks and therefore inhibit PIP_3 formation, which in turn downregulates class I PI3K signaling such as the Akt pathway[149]. Other studies implicated PI5P4Kβ activity in the nucleus with cellular stress cascades[157], including mechanical stress-induced healing processes in muscle cells[158] as well as vitamin D-facilitated nuclear PI-4,5-P_2 production leading to E-cadherin expression[159].

PI5P4Kγ-specific cellular functions include coordination of vesicle transport, especially in kidney cells, where PI5P4Kγ is found to be highly expressed[160]. Interestingly, the allosteric PI4P5K inhibitory activity of PI5P4Ks could explain how PI5P4Kγ maintains high PI-4-P levels at cellular vesicles by preventing conversion of this phospholipid to PI-4,5-P_2[149]. Moreover, PI5P4Kγ function has been implicated in Rho activation and mitotic spindle formation[161], which was recently confirmed in a study showing that PI5P4Kγ localizes at the spindle pole, producing local pools of PI-4,5-P_2 and supporting stability of microtubules[162]. PI5P4Kγ has been further found to regulate a variety of signaling pathways including mTORC1 signaling in nutrition-deprived cells[163], which is potentially also involved in regulatory processes of the immune system[164], as well as Notch signaling *via* induction of receptor recycling mechanisms at tubulovesicular membranes[165]. Interestingly, it was found that PI5P4Kγ and Notch are both negatively regulated by the local E3 ligase deltex-1 (DTX1) leading to degradation of the receptor and the lipid kinase[165].

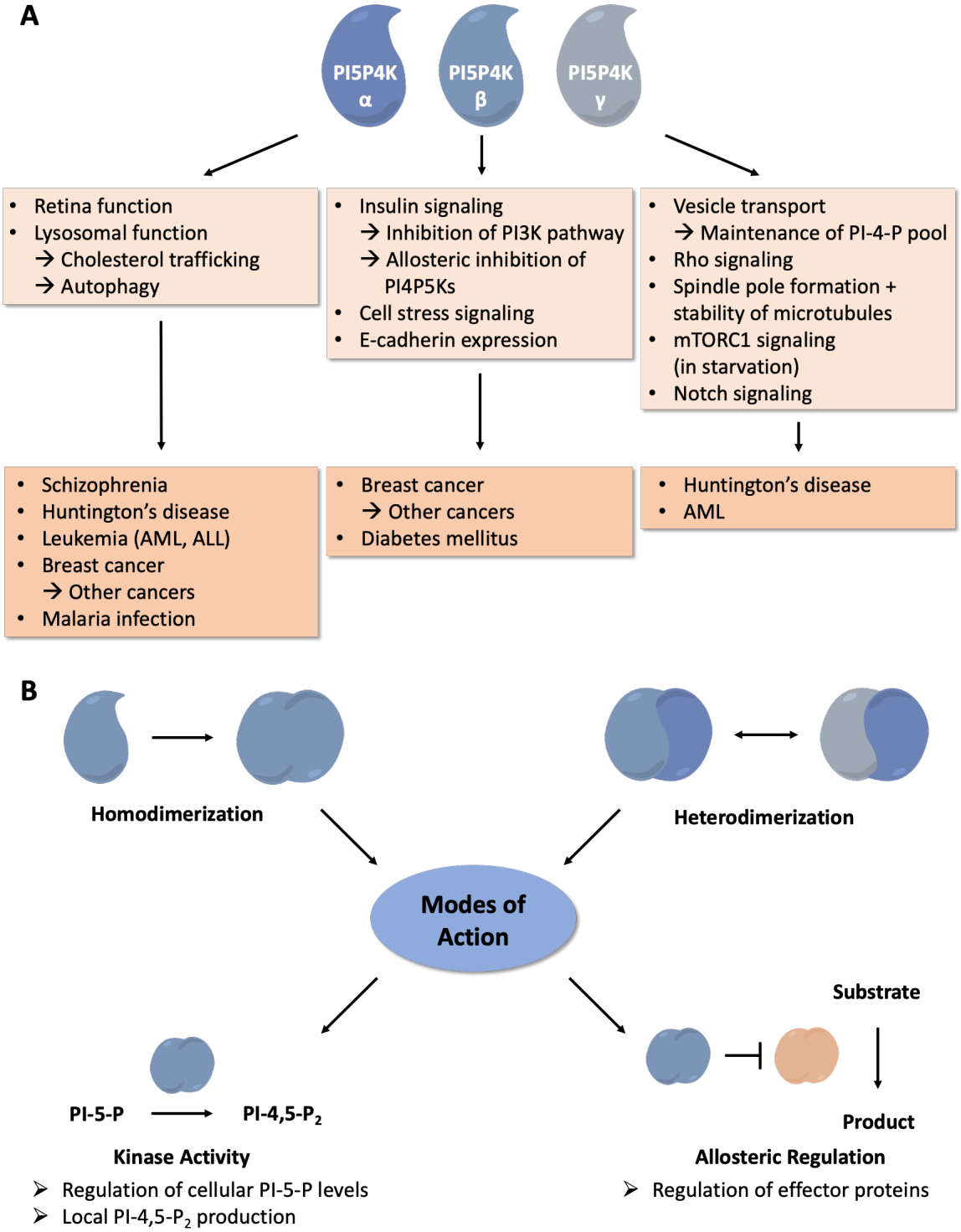


Figure 5.1.22 | PI5P4Ks in Health and Disease. A. Overview of PI5P4K function in cellular processes (light orange) and implications in human diseases (dark orange). **B.** PI5P4Ks are able to homo- and heterodimerize. The dimers can associate with cellular membranes, while heterodimerization of the isoforms is stirring and regulating PI5P4K function. Besides their kinase activity, PI5P4Ks are also able to control effector proteins *via* allosteric mechanisms.

Dysregulation of PI5P4K function has thus been associated with the development of a variety of diseases (Figure 5.1.22 A). Genetic variations of the *PIP4K2A* gene (for PI5P4Kα) have been linked to schizophrenia[166] as well as responsiveness to antipsychotic treatment[167]. Similarly, PI5P4Kγ has been investigated as a potential target for the treatment of Huntington's disease due to its implications in autophagy processes[168]. PI5P4Kα expression has been further linked to leukemia, including acute lymphoid leukemia (ALL) and acute myeloid leukemia (AML)[169]. Interestingly, some studies argue that low expression of PI5P4Kα in leukemia cells is associated with poor prognosis, which could be a result of the consequently high cellular PI3K/Akt and p38 MAPK signaling[169]. This tumor suppressor activity of PI5P4Kα was further described in glioblastoma lacking PTEN-facilitated inhibition of the PI3K pathway[170]. However, genetic knockdown studies provided evidence that *PIP4K2A* is an essential gene in AML[171] and that (childhood) ALL prevalence is associated with polymorphisms of *PIP4K2A*[172]. Furthermore, a recent study found that not only low PI5P4Kα levels but also downregulated PI5P4Kγ (not PI5P4Kβ) in AML patients generally correlated with a positive clinical outcome[173]. Single nucleotide polymorphisms (SNPs) in *PIP4K2A* leading to PI5P4Kα overexpression were thus found in a variety of different cancers and were associated with lower overall survival rates, while in certain cancers the opposite was the case[174]. Surprisingly, PI5P4Kα was also found to play a role in transcription initiation *via* direct mRNA binding in the context of infections with the malaria-causing parasite *Plasmodium falciparum*[175].

Similar to PI5P4Kα, PI5P4Kβ has been found overexpressed in cancers, such as breast cancer, promoting cancer cell proliferation[176]. Some studies suggested that PI5P4Kβ is able to sustain cell growth in nutrition-deprived environments, e.g. tumors, *via* its GTP-sensing ability to regulate cellular metabolism[177]. However, PI5P4Kβ is also discussed as a tumor suppressor by promoting transcription of E-cadherin and low PI5P4Kβ expression correlating with a decrease in survival in breast cancer patients[178]. A recent study further suggests that the tumor suppressing long non-coding RNA PXN-AS1 act by promoting PI5P4Kβ expression[179].

Notably, Emerling and coworkers reported in 2013 that both PI5P4Kα and β become essential proteins in p53-deprived breast cancer cells[180]. The group showed *via* genetic knockout studies that the combination of p53 deletion with *PIP4K2B*$^{-/-}$ (for PI5P4Kβ) was lethal *in vivo*, while a *PIP4K2A*$^{-/-}$ and *PIP4K2B*$^{+/-}$ combination led to a significant decrease in tumor formation in p53-null mice, potentially due to resulting issues in glucose metabolism and elevated ROS levels[180]. These results are highly intriguing since p53 mutations or deletion mutations are found in over 70% of human cancers, thus suggesting an interesting potential in targeting PI5P4Ks in cancer[181]. Consequently, PI5P4K inhibitors pose essential tools in interrogating PI5P4K function in cancer progression and would further provide essential information for the development of potential novel cancer treatments.

The identification of first small molecule PI5P4K inhibitors (Figure 5.1.23) was reported by Davis *et al.* in 2013[182]. The group presented a novel high-throughput platform enabling the screening of chemical libraries for PI5P4K inhibitor scaffolds[182]. As a result, the tyrphostin analog I-OMe-AG-538 was identified, exhibiting ATP-competitive inhibition of PI5P4Kα with an IC_{50} of 1 μM[182], while further studies demonstrated pan-PI5P4K activity of the inhibitor[183]. Shortly after, a PI5P4Kβ inhibitor, SAR088, was identified in a PI5P4Kβ isoform-specific high-throughput screen[184]. The compound showed selectivity

within a panel of 35 kinases and inhibited PI5P4Kβ with an IC_{50} of 2.18 μM, exhibiting a 9-fold selectivity over PI5P4Kα[184]. Interestingly, the compound showed *in vivo* efficacy in an insulin-resistant type 2 Diabetes mellitus mouse model[184], providing another potential therapeutic angle for selective PI5P4Kβ inhibitors.

In 2015, Clark and colleagues reported the identification of the first PI5P4Kγ-selective small molecule inhibitor (NIH-12848), which targets the PI-5-P-binding pocket of the lipid kinase, thus exhibiting an over 100-fold selectivity over the PI5P4Kα and β isoforms[185]. To determine IC_{50} values of the compound, the group employed an activity assay using a kinase active PI5P4Kγ mutant to increase its intrinsically low catalytic activity, leading to an apparent IC_{50} of 1 μM[185]. The tool compound was further applied to investigate PI5P4Kγ function in kidney cells, where the small molecule was able to inhibit Na^+/K^+-ATPase redistribution, mirroring effects seen with RNAi knockdown of PI5P4Kγ[185]. More recently, Kitagawa *et al.* reported the discovery of a131, which showed cancer cell-selective growth inhibition[186]. The group identified pan-PI5P4K inhibitor activity of the compound and proposed a unique, dual mode of action, in which a131 causes a temporary growth arrest of normal cells *via* upregulation of *PIK3IP1* (a negative regulator of the PI3K pathway), while Ras-activated tumor cells override this signal by activating PI3K signaling, thus causing mitotic catastrophe and ultimately cell death[186]. This novel mechanism therefore elucidated a potential cellular link between the Ras/MEK/ERK and PI3K/Akt/mTOR pathways[186].

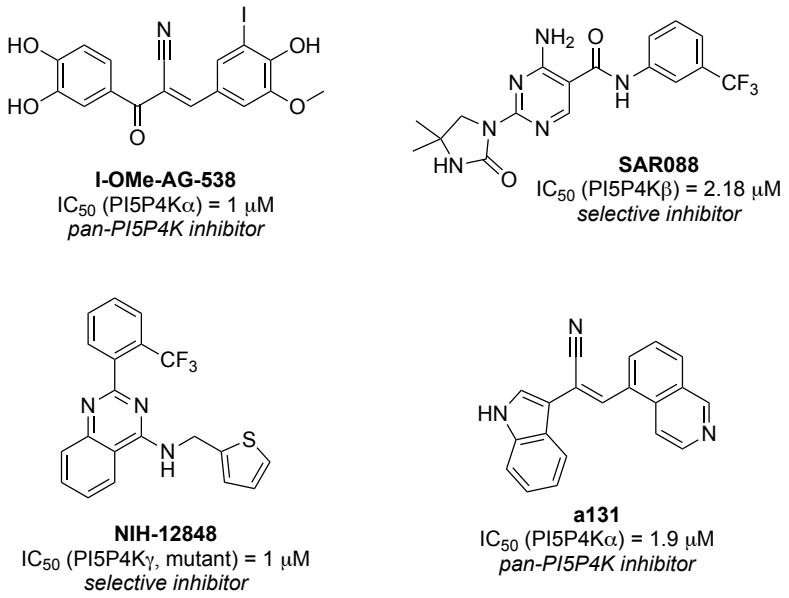


Figure 5.1.23 | Reported PI5P4K Inhibitors. Structures of reported selective and pan-PI5P4K inhibitors.

Taken together, PI5P4Ks pose intriguing novel targets in a variety of diseases including neurological conditions, diabetes, and cancer. However, PI5P4K function and regulation is still poorly understood, highlighting the need of highly potent and selective chemical tools to further investigate these lipid kinases and their function in human health and disease.

5.2 Publication A: *Targeting the PI5P4K Lipid Kinase Family in Cancer Using Novel Covalent Inhibitors*

5.2.1 Specific Aims

THZ-P1-2 was found to be a potent pan-PI5P4K inhibitor as the result of an internal medicinal chemistry campaign, which followed-up on the PI5P4K off-target activity of the THZ-P1-2 analog JNK-IN-7. When joining this effort in the development of a selective and potent PI5P4K-targeting probe, the aim was the synthesis of the DTB-labeled analog of the parent compound (THZ-P1-2-DTB) to enable further characterization and validation of the PI5P4K inhibitor.

5.2.2 Main Outcomes and Context

The presented study showed that THZ-P1-2 covalently and potently binds and inhibits PI5P4K activity in biochemical and cellular assays, with THZ-P1-2-DTB as an essential labeled probe of the investigated inhibitor. THZ-P1-2 further exhibited decent selectivity across the kinome, while treatment of PI5P4K-dependent AML and ALL cell lines with the covalent inhibitor resulted in dose-dependent antiproliferative effects. Finally, the probe was shown to disrupt autophagy *via* PI5P4K-mediated mechanisms.

As a result, the synthesis of the DTB-labeled analog of THZ-P1-2 enabled the direct investigation of the cellular target engagement and selectivity of the parent compound *via* bead-based pulldown experiments. Thus, the characterized probe enabled further studies of the biological targets – as in this case, PI5P4K function in AML/ALL cells, including their role in autophagy. The presented study nicely shows the value of desthiobiotin-labeling methods for the study and characterization of a small molecule probe, such as a covalent protein inhibitor.

The respective synthetic routes can be found in the Supporting Information of the publication.

5.2.3 Published Article

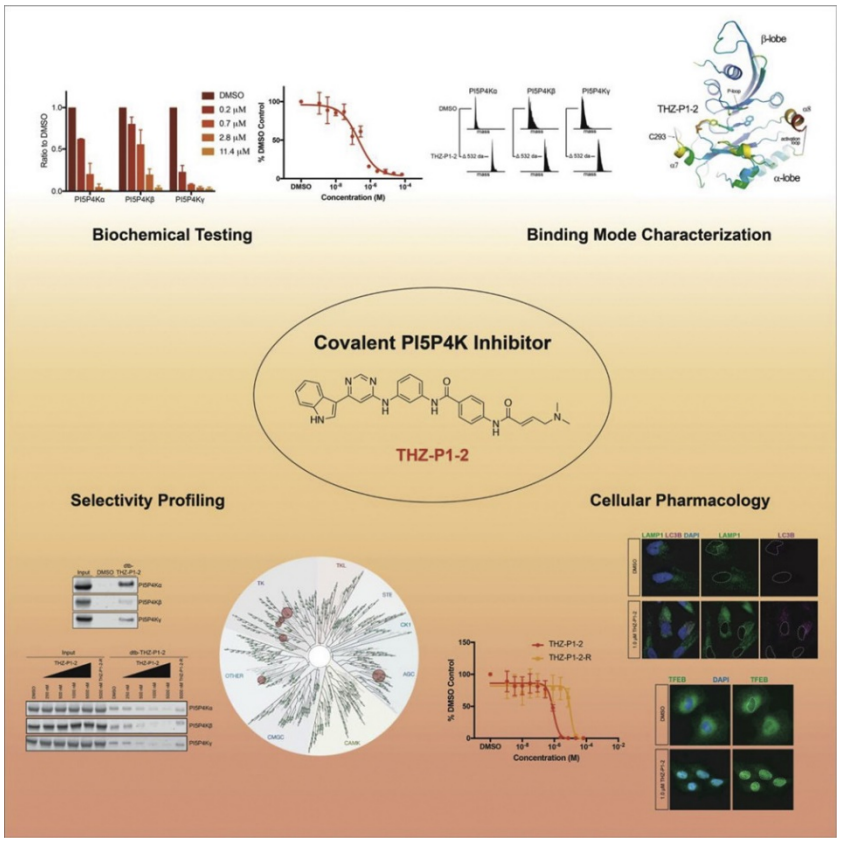
The following peer-reviewed publication was originally published as:

Targeting the PI5P4K Lipid Kinase Family in Cancer Using Covalent Inhibitors

Sindhu Carmen Sivakumaren, Hyeseok Shim, Tinghu, Zhang, Fleur M. Ferguson, Mark R. Lundquist, Christopher M. Browne, Hyuk-Soo Seo, Marcia N. Paddock, Theresa D. Manz, Baishan Jiang, Ming-Feng Hao, Pranav Krishnan, Diana G. Wang, T. Jonathan Yang, Nicholas P. Kwiatkowski, Scott B. Ficarro, James M. Cunningham, Jarrod A. Marto, Sirano Dhe-Paganon, Lewis C. Cantley, and Nathanael S. Gray.

Cell Chem. Biol. **2020**, *27* (5), 525-537.

DOI: https://doi.org/10.1016/j.chembiol.2020.02.003



This article, including Supporting Information, was reproduced with kind permission from the journal *Cell Chemical Biology* © 2020 Elsevier Ltd.

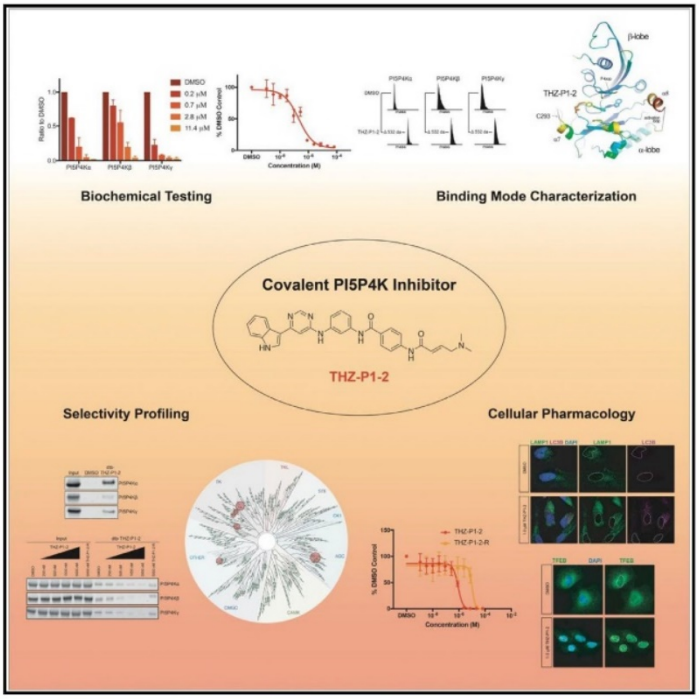
Note:
The publication was initially available as a preprint on the bioRxiv server.
DOI: https://doi.org/10.1101/819961 (posted: 10/25/2019)

Article

Cell Chemical Biology

Targeting the PI5P4K Lipid Kinase Family in Cancer Using Covalent Inhibitors

Graphical Abstract



Authors

Sindhu Carmen Sivakumaren, Hyeseok Shim, Tinghu Zhang, ..., Sirano Dhe-Paganon, Lewis C. Cantley, Nathanael S. Gray

Correspondence

lcantley@med.cornell.edu (L.C.C.), nathanael_gray@dfci.harvard.edu (N.S.G.)

In Brief

PI5P4K, an understudied kinase family, is essential in various disease contexts. Sivakumaren et al. develop and characterize PI5P4K inhibitor THZ-P1-2, which targets unique cysteines, exhibits effects in biochemical and cellular assays, displays anticancer activity in leukemia cell lines, and causes defects in autophagy similar to PI5P4K gene knockdown or deletion.

Highlights

- Inhibitor THZ-P1-2 shows PI5P4K enzyme inhibition and target engagement in cells
- THZ-P1-2 covalently targets unannotated cysteines outside the PI5P4K active site
- AML/ALL cell lines are broadly sensitive to THZ-P1-2's covalent effects
- PI5P4K inhibition causes autophagy disruption and upregulates TFEB signaling

Sivakumaren et al., 2020, Cell Chemical Biology 27, 525–537
May 21, 2020 © 2020 Elsevier Ltd.
https://doi.org/10.1016/j.chembiol.2020.02.003

CellPress

Cell Chemical Biology

CellPress

Article

Targeting the PI5P4K Lipid Kinase Family in Cancer Using Covalent Inhibitors

Sindhu Carmen Sivakumaren,[1,2,11] Hyeseok Shim,[3,11] Tinghu Zhang,[1,2,11] Fleur M. Ferguson,[1,2] Mark R. Lundquist,[3] Christopher M. Browne,[1,2,10] Hyuk-Soo Seo,[1,2] Marcia N. Paddock,[3] Theresa D. Manz,[1,2,4] Baishan Jiang,[1,2] Ming-Feng Hao,[1,2] Pranav Krishnan,[5] Diana G. Wang,[3] T. Jonathan Yang,[3] Nicholas P. Kwiatkowski,[1,2,6] Scott B. Ficarro,[1,7,8] James M. Cunningham,[5] Jarrod A. Marto,[1,7,8,9] Sirano Dhe-Paganon,[1,2] Lewis C. Cantley,[3,*] and Nathanael S. Gray[1,2,12,*]

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[3]Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, NY 10065, USA
[4]Department of Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbruecken, Germany
[5]Department of Medicine, Division of Hematology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[6]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[7]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[8]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[9]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[10]Present address: Discovery Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Boston, MA, USA
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: lcantley@med.cornell.edu (L.C.C.), nathanael_gray@dfci.harvard.edu (N.S.G.)
https://doi.org/10.1016/j.chembiol.2020.02.003

SUMMARY

The PI5P4Ks have been demonstrated to be important for cancer cell proliferation and other diseases. However, the therapeutic potential of targeting these kinases is understudied due to a lack of potent, specific small molecules available. Here, we present the discovery and characterization of a pan-PI5P4K inhibitor, THZ-P1-2, that covalently targets cysteines on a disordered loop in PI5P4Kα/β/γ. THZ-P1-2 demonstrates cellular on-target engagement with limited off-targets across the kinome. AML/ALL cell lines were sensitive to THZ-P1-2, consistent with PI5P4K's reported role in leukemogenesis. THZ-P1-2 causes autophagosome clearance defects and upregulation in TFEB nuclear localization and target genes, disrupting autophagy in a covalent-dependent manner and phenocopying the effects of PI5P4K genetic deletion. Our studies demonstrate that PI5P4Ks are tractable targets, with THZ-P1-2 as a useful tool to further interrogate the therapeutic potential of PI5P4K inhibition and inform drug discovery campaigns for these lipid kinases in cancer metabolism and other autophagy-dependent disorders.

INTRODUCTION

Phosphatidylinositol 5-phosphate (PI-5-P) is one of the seven phosphosinositides that regulate a wide range of cellular functions (Balla et al., 2009; Bulley et al., 2015). Known to localize in the nucleus, Golgi, endoplasmic reticulum, and at the plasma membrane, PI-5-P is an oxidative stress-induced regulator of AKT activation and is also regulated by proline isomerase Pin1 through Pin1's interactions with the phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) in times of cellular stress (Keune et al., 2012; Keune et al., 2013). The phosphorylation of the low abundance phosphoinositide PI-5-P at the 4-position, yielding the product phosphatidylinositol-4,5-bisphosphate (PI-4,5-P_2), is catalyzed by PI5P4Kα, β, and γ (encoded by genes *PIP4K2A*, *PIP4K2B*, and *PIP4K2C*) (Rameh et al., 1997; Rameh and Cantley, 1999). PI-4,5-P_2 is an important precursor for second messengers inositol-1,4,5-triphosphate, diacylgycerol, and phosphatidylinositol-3,4,5-trisphosphate (Martelli et al., 1992; Divecha et al., 1993; Fiume et al., 2012, 2015).

Although most of PI-4,5-P_2 is generated by phosphorylation of phosphatidylinositol 4-phosphate on the 5-position by the type I PI4P5K kinases, a PI5P4K-driven alternate route was discovered in 1997, hence the designation type II (Rameh et al., 1997). The PI5P4Ks were traditionally thought to mainly be crucial direct regulators of PI-5-P levels (Bulley et al., 2015; Stijf-Bultsma et al., 2015; Hasegawa et al., 2017). However, PI5P4Kα was found to synthesize a pool of PI-4,5-P_2 that is specifically important in mTORC2 regulation (Bulley et al., 2016) and to play a critical role in intracellular cholesterol transport by modulating PI-4,5-P_2 homeostasis on peroxisome membranes (Hu et al., 2018). The low-activity isoform PI5P4Kγ was demonstrated to positively regulate Notch1 signaling by facilitating receptor recycling, suggesting that endosome-localized production of PI(4,5)P_2 is involved Notch transport (Zheng and Conner, 2018). PI5P4Kα/β were also shown to be required for autophagosome-lysosome fusion during times of metabolic stress, suggesting that they were evolved by multicellular organisms to produce sufficient PI-4,5-P_2 in nutrient-deficient conditions (Lundquist et al.,

2018). These findings have dispelled the notion of PI5P4K as simply being functionally redundant in PI-4,5-P_2 production.

PI5P4K has been suggested to be important in several diseases. *PIP4K2B*-null mice were found to have increased insulin sensitivity and reduced adiposity (Lamia et al., 2004). *PIP4K2A* was found to be a dependency in acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) (Jude et al., 2014; Rosales-Rodríguez, et al., 2016; Urayama et al., 2018) and *PIP4K2A*$^{-/-}$, *PIP4K2B*$^{+-}$, *TP53*$^{-/-}$ mice had a dramatic tumor-free life extension compared with *TP53*$^{-/-}$ mice, uncovering a potential synthetic lethality of PI5P4K with p53 (Emerling et al., 2013). Knockdown of *PIP4K2A/B* in human retinal pigment epithelial cells and rabbit models abrogated the pathogenesis of proliferative vitreoretinopathy (Ma et al., 2016). Deletion of *PIP4K2C* in mice resulted in an increase in proinflammatory cytokines and T helper cells, as well as a decrease in regulatory T cells via hyperactivation of mTORC1 signaling (Shim et al., 2016). Pharmacological inhibition or knockdown of PI5P4Kγ reduced mutant huntingtin protein in human patient fibroblasts and aggregates in neurons, and relieved neuronal degeneration in *Drosophila* models of Huntington disease (Al-ramahi et al., 2017). The critical role of the PI5P4Ks in mediating autophagy may explain their induced essentiality in various disease pathologies (Emerling et al., 2013; Vicinanza et al., 2015; Al-ramahi et al., 2017; Lundquist et al., 2018). Collectively, these studies suggest that the PI5P4Ks represent a lipid kinase family whose underlying biology is important to numerous cellular processes and warrants further investigation of their therapeutic potential across a range of disease states.

The relevance of PI5P4K in a wide range of diseases has motivated efforts to develop PI5P4K inhibitors. Reported pan-PI5P4K inhibitors (Kitagawa et al., 2017) and isoform-specific inhibitors of PI5P4Kα (Davis et al., 2013), PI5P4Kβ (Voss et al., 2014), and PI5P4Kγ (Clarke et al., 2015; Al-ramahi et al., 2017) have laid the foundation for evidence of PI5P4K druggability and motivated a need for inhibitors with further improved pharmacological properties. Here, we present the identification of a potent and selective covalent PI5P4K inhibitor, THZ-P1-2, and characterize its cellular pharmacology in the contexts of autophagy and cancer. Using a multipronged approach combining biochemical and cellular assays, mass spectrometry, and crystallography, we discovered that THZ-P1-2 inhibits the PI5P4K family at sub-micromolar concentrations by reacting covalently with cysteine residues in a flexible loop outside the kinase domain of all three kinase isoforms. We show that THZ-P1-2 exhibits a reasonable selectivity profile across the kinome, with an S score ($S_{(10)}$) of 0.02 (Karaman et al., 2008; Davis et al., 2011) and inhibits cell proliferation at micromolar concentrations in a panel of leukemia cell lines in a manner dependent on covalent targeting. Finally, further investigation of the mechanism of action of THZ-P1-2 led us to the observation that THZ-P1-2 causes defects in the clearance of autophagosomes and lysosomes, halting autophagy and phenocopying the effects of genetic deletion of PI5P4Kα/β/γ (Al-ramahi et al., 2017; Lundquist et al., 2018). These studies provide evidence that irreversible inhibition of PI5P4K by THZ-P1-2 compromises autophagy, an essential alternative energy source during periods of metabolic stress which cancer cells depend on to maintain cellular homeostasis and prolonged cell viability, further suggesting PI5P4K as a potentially important target in cancer metabolism and other autophagy-dependent diseases.

RESULTS

Chemoproteomic Profiling and Synthetic Chemistry Approaches Reveal Lead PI5P4K Inhibitor THZ-P1-2

A previously developed acrylamide-based covalent JNK inhibitor, JNK-IN-7, was serendipitously discovered to also have activity on PI5P4Kγ through KiNativ cellular selectivity profiling (Figures 1A and S1A) (Zhang et al., 2012). JNK-IN-7 was found to inhibit the kinase activity of PI5P4Kα/β/γ in a radiometric thin-layer chromatography (TLC) assay (Figures S1B and S1C), and mass spectrometry revealed labeling of cysteine residues located on a disordered loop that is not observed in the available PI5P4K crystal structures (Figures S1D and S1E). We pursued a focused medicinal chemistry campaign guided by biochemical kinase assays and cellular pulldowns to optimize the potency and selectivity of phenylamino-pyrimidine, which resulted in the development of THZ-P1-2 (Figure 1A) (full structure-activity relationship study described in Manz et al., 2019). THZ-P1-2 demonstrated inhibition of PI5P4Kα kinase activity (Figure 1B), with a half maximal inhibitory concentration (IC_{50}) of 190 nM using an ADP-Glo (Promega) bioluminescence assay. THZ-P1-2 exhibited approximately 75% inhibition of PI-4,5-P_2 formation by PI5P4Kα and PI5P4K γ and 50% inhibition by PI5P4Kβ at a concentration of 0.7 μM monitored using a TLC assay (Figures 1C and 1D).

THZ-P1-2 Binds in the Active Site and Covalently Modifies Cysteine Residues of the PI5P4Ks

Based on the initial evidence of ATP-competitive inhibition and the covalent binding of JNK-IN-7 to PI5P4Kγ, we sought to confirm covalent binding of THZ-P1-2 to each PI5P4K isoform. Upon incubating THZ-P1-2 with recombinant PI5P4Kα/β/γ protein and analyzing each sample by electrospray mass spectrometry, we discovered that the compound covalently labeled each isoform, as demonstrated by the mass shift corresponding to the labeled protein, more efficiently than observed with JNK-IN-7 (Figure 2A). Notably, THZ-P1-2 achieved 100% labeling of PI5P4Kα/β within 2 h, and PI5P4Kγ in 30 min, corroborating earlier observations in the TLC assay. This observed labeling depends on the compound possessing the cysteine-reactive acrylamide moiety, and we confirmed this using a non-cysteine-reactive analog of THZ-P1-2 containing a 4-(dimethylamino)butanamide warhead (Figure 1A). The reversible counterpart, denoted by THZ-P1-2-R, did not label the recombinant PI5P4K protein (Figure S2B).

To ascertain the specific residues covalently modified by THZ-P1-2, we performed protease digestion and tandem mass spectrometry. Here, we identified peptides containing previously unannotated cysteine residues on a conserved disordered loop residing outside of the PI5P4K isoform kinase ATP-binding sites—Cys293 on PI5P4Kα, Cys307 and Cys318 on PI5P4Kβ, and Cys313 on PI5P4Kγ—that were covalently modified by THZ-P1-2 (Figures 2B and S2A). We hypothesize that this interaction is possible as the disordered loop region is brought into close proximity of the ATP-binding pocket through tertiary structure. These observations add PI5P4K to the list of kinases that

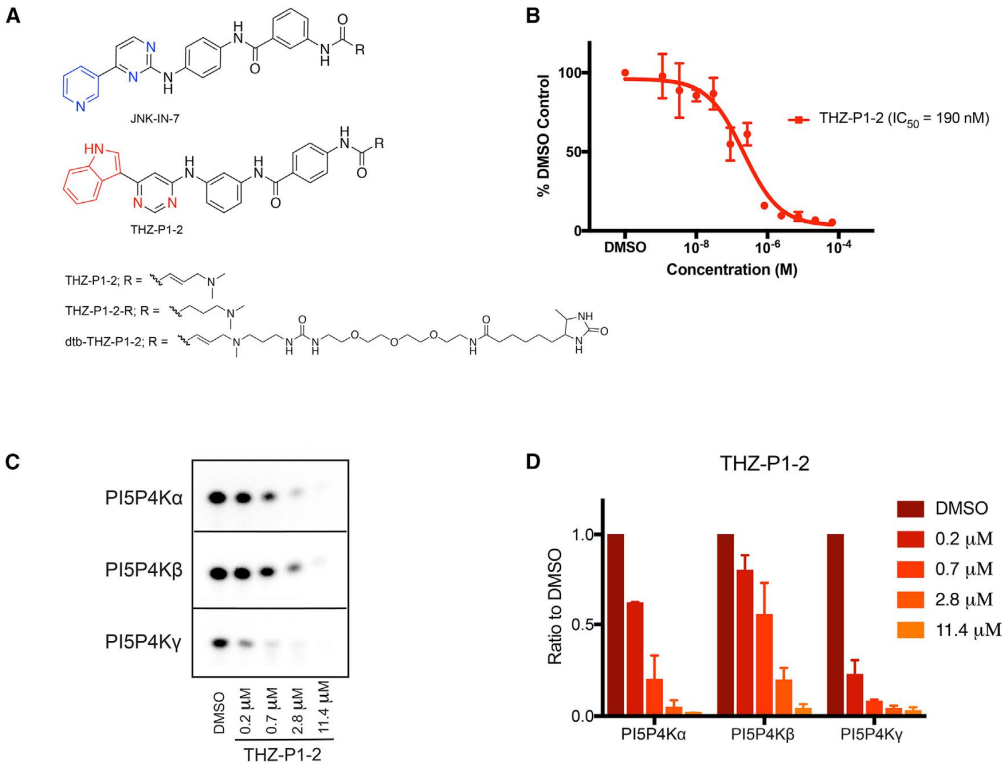


Figure 1. Chemoproteomic Profiling and Synthetic Chemistry Approaches Reveal PI5P4K Inhibitor THZ-P1-2

(A) Chemical structures of compounds JNK-IN-7 as reported previously in Zhang et al. (2012) and THZ-P1-2 with its reversible and desthiobiotinylated counterparts. See also Figure S1.

(B) THZ-P1-2 potently inhibits PI5P4Kα kinase activity in the ADP-Glo luminescence assay. Data shown are the mean of two biological replicates ± SEM.

(C) THZ-P1-2 shows potent inhibition of kinase activity of all three PI5P4K isoforms in a radiometric TLC assay measuring radiolabeled PI-4,5-P_2. TLC image shown is representative of two independent experiments.

(D) Quantification of (C). Radiolabeled PI-4,5-P_2 spots were imaged by autoradiography and quantified by densitometry.

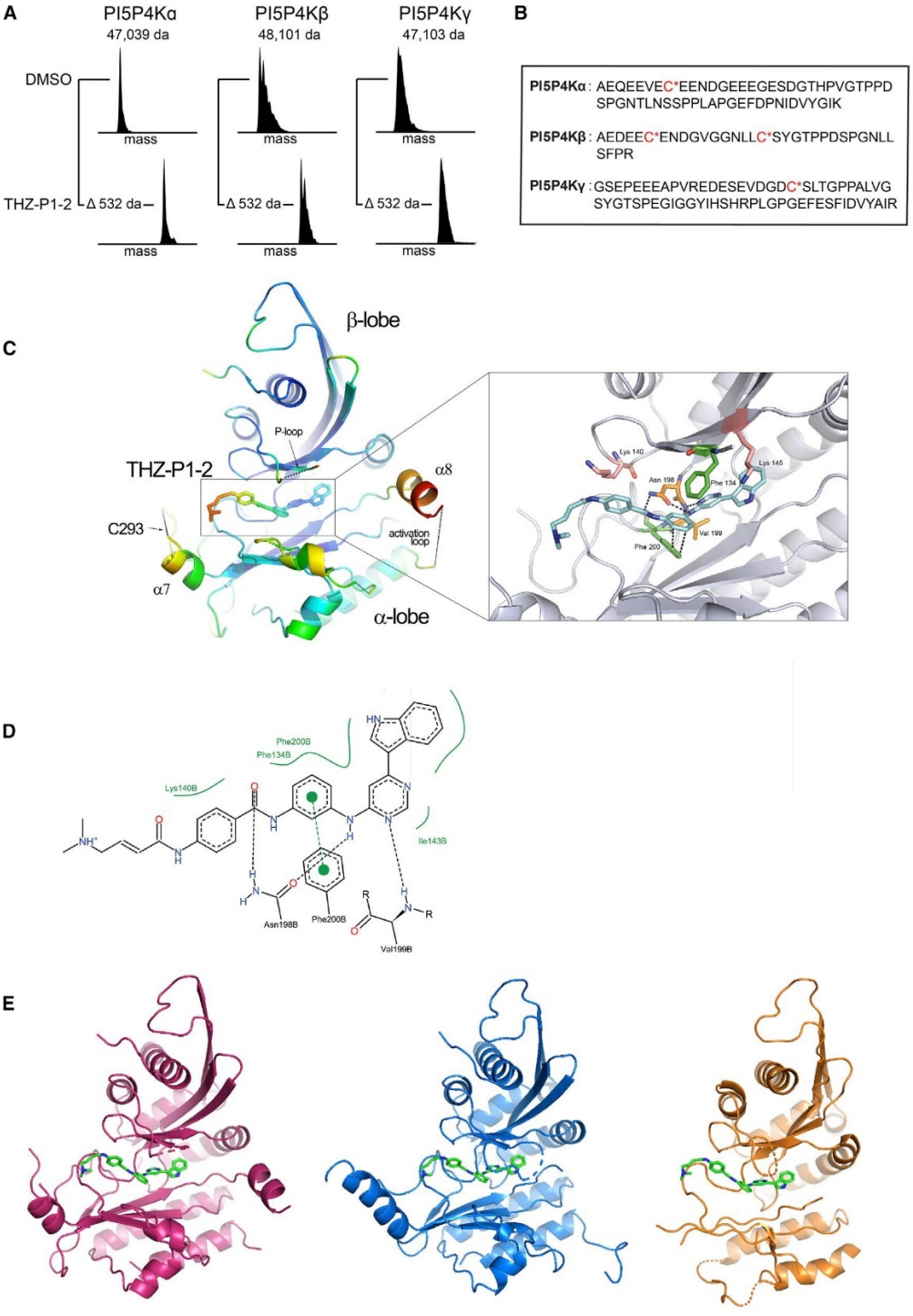
possess targetable cysteines distant from the ATP site in sequence yet accessible proximal to the ATP site due to the overall kinase conformation (Zhang et al., 2016; Kwiatkowski et al., 2014; Browne et al., 2018).

Elucidation of the Crystal Structure of THZ-P1-2 Bound to PI5P4Kα

To elucidate the detailed molecular interactions involved with the covalent and noncovalent interactions of THZ-P1-2 with the PI5P4Ks, we determined a crystal structure of PI5P4Kα in complex with THZ-P1-2 to 2.21 Å resolution (Figures 2C and S2C). The crystal structure showed crucial noncovalent interactions within the ATP-binding pocket. Asn198 forms hydrogen-bonding interactions between the carbonyl of the benzamide and the amine linker connecting the phenylenediamine with the pyrimidine moiety of THZ-P1-2. Furthermore, aromatic π-stacking interactions between the phenylenediamine and Phe200, as well as hydrophobic interactions between the indole head group and linker regions of THZ-P1-2 and several residues in the active site were detected (Figures 2C and 2D). Unfortunately, as has been observed in previously published crystal structures, there was insufficient electron density to resolve the disordered loop containing the reactive Cys293 (Figure S2D). However, it is clear from the co-structure that the cysteine-containing loop could easily traverse the region adjacent to where the acrylamide warhead of THZ-P1-2 is situated. Alignment of crystal structures of PI5P4Kβ (PDB: 1BO1) and PI5P4Kγ (PDB: 2GK9) deposited in the PDB to our resolved co-crystal structure in PyMOL allowed for a more direct comparison between the three isoforms, showing that THZ-P1-2 likely binds to the other two isoforms in a similar manner given the high degree of homology in the active sites and cysteine-containing loops (Figure 2E).

Cellular Target Engagement and Kinome Selectivity of THZ-P1-2

To confirm that the covalent binding observed *in vitro* is also observed in a cellular context, we synthesized a desthiobiotinylated derivative, dtb-THZ-P1-2, which maintained the biochemical potency of the parental compound (Figure S3A). A streptavidin pulldown assay in HEK293T cell lysate confirmed that dtb-THZ-P1-2 could capture PI5P4Kα/β/γ at a



(legend on next page)

concentration of 1 μM. We further demonstrated cell permeability and cellular on-target engagement of THZ-P1-2 by performing the streptavidin pulldown assay in a competitive fashion, first pre-treating HEK293T cells with the parent compound, followed by pulldown in cell lysate with the desthiobiotinylated derivative. THZ-P1-2 demonstrated the ability to compete and block pulldown of PI5P4K with dtb-THZ-P1-2 in a dose-dependent manner (Figure 3B). We observed that engagement measured by the streptavidin pulldown assay also determined covalent binding, as pre-treatment with THZ-P1-2-R was not able to block pulldown by dtb-THZ-P1-2 (Figure 3B) even though the reversible inactive control compound was still able to inhibit kinase activity to a lesser extent than the covalent compound in the ADP-Glo biochemical assay (Figure S3A). We validated THZ-P1-2's continuous on-target engagement typical of covalent inhibitors by performing an additional pulldown with a washout at 2 and 4 h, observing engagement as early as 2 h with 1 μM compound treatment (Figure S3B).

To broadly evaluate the kinome-wide selectivity of THZ-P1-2 we used the commercially available DiscoverX KINOMEscan profiling platform which revealed appreciable selectivity across the kinome with a selectivity score at a 10% DMSO control cutoff, $S_{(10)}$, of 0.02 (indicating THZ-P1-2 bound to 2% of queried kinases at the specified cutoff) (Figure 3C) (Karaman et al., 2008; Davis et al., 2011). We found the related lipid kinase PIKfyve to be an off-target (6.4% DMSO control). Because PIKfyve and PI4P5Kα/β/γ could reasonably be off-targets due to overall similar lipid kinase structure, we first tested THZ-P1-2 inhibitory activity on these kinases. We observed low micromolar IC_{50} values against PI4P5Kα/β/γ and an IC_{50} of 40 nM on PIKfyve in a fixed time point ADP-Glo assay (Carna Biosciences) (Figure S3C). To further assess if THZ-P1-2 biochemical inhibition of these closely related lipid kinases was maintained in cells, we evaluated dtb-THZ-P1-2 binding of these targets by streptavidin pulldown. We observed no engagement of PI4P5Kα/β/γ and PIKfyve, establishing the cellular selectivity of THZ-P1-2 against these related type 1 and type 3 lipid kinases (Figure 3D). Because of THZ-P1-2's high potency on PIKfyve in the biochemical context, we took a step further to investigate its cellular activity on PIKfyve in a vacuolar enlargement assay. THZ-P1-2 produced the established PIKfyve-inhibitory phenotype in Vero cells only at a high concentration of 10 μM, 1,000× that of positive control apilimod (Cai et al., 2013; Choy et al., 2018; Sharma et al., 2019), which causes substantial vacuolar enlargement at 10 nM (Figure S3D), but failed to show inhibition of PIKfyve at 1 μM. THZ-P1-2 was also unable to impair the ability of PIKfyve inhibitor apilimod to induce vacuolar enlargement when cells were either pre-incubated with THZ-P1-2 followed by apilimod, or co-incubated with both compounds (Figure S3E).

The additional potential off-targets identified through KINOMEscan were further evaluated using the Z'LYTE kinase assay (Invitrogen). All top off-targets except for BRK and ABL1 were found to have IC_{50} values in the micromolar range (Figure S4A). Engagement of both BRK and ABL1, both of which lack cysteines in or near the vicinity of the ATP-site, were assessed in the cellular pulldown assay with dtb-THZ-P1-2 and found to have little to no pulldown, suggesting that the affinity observed in the Z'LYTE assay may have been due to tight noncovalent binding which was not perpetuated in the cellular context (Figure S4B). THZ-P1-2 was markedly less potent than the ABL inhibitors, imatinib, nilotinib, and dasatinib, at killing BCR-ABL translocation containing cell lines (K562 and KU812F), indicating that THZ-P1-2 does not effectively inhibit BCR-ABL in cells (Figure S4C). Interestingly, even with retention of the indole head group of THZ1 and THZ531, the switch in 2,4- to 4,6-pyrimidine from JNK-IN-7 to THZ-P1-2 was sufficient to alleviate engagement of kinases targeted by related phenylaminopyrimidine acrylamides, including JNK, IRAK1, PKN3, CDK7, and CDK12, when probed in a streptavidin pulldown with their respective biotinylated or desthiobiotinylated compounds (Zhang et al., 2012; Kwiatkowski et al., 2014; Browne et al., 2018) (Figure S4D).

THZ-P1-2 as a Probe for Potential PI5P4K Dependencies in Cancer

PIP4K2A was found to be essential for cell survival in AML (Jude et al., 2014), and several variant SNPs in the *PIP4K2A* locus have been associated with ALL susceptibility and leukemogenesis (Rosales-Rodríguez, et al., 2016; Urayama et al., 2018). We evaluated the sensitivity of a small panel of leukemia cell lines to THZ-P1-2, using a 72-h CellTiter-Glo assay. We used THZ-P1-2-R as a negative compound to control for effects of THZ-P1-2 that are mediated solely by reversible binding. THZ-P1-2 demonstrated modest anti-proliferative activity in all six AML/ALL cell lines with IC_{50} values in the low micromolar range. We observed approximately 10-fold IC_{50} shifts between THZ-P1-2 and THZ-P1-2-R (Figure 4), indicating that the presence of the PI5P4K cysteine-targeting acrylamide partially contributes to the anti-proliferative activity. ABL1 fusion genes are also heavily involved in hematological malignancies (De Braekeleer et al., 2011), but the weak engagement of ABL1 in the cellular context (Figures S4B and S4C), together with the covalent dependency, gave confidence that the anti-proliferative activity observed is not solely due to ABL1 inhibition.

Figure 2. THZ-P1-2 Binds Covalently to All Isoforms of the PI5P4K Family on Unique Cysteine Residues Located on a Disordered Loop Outside the Kinase Domain

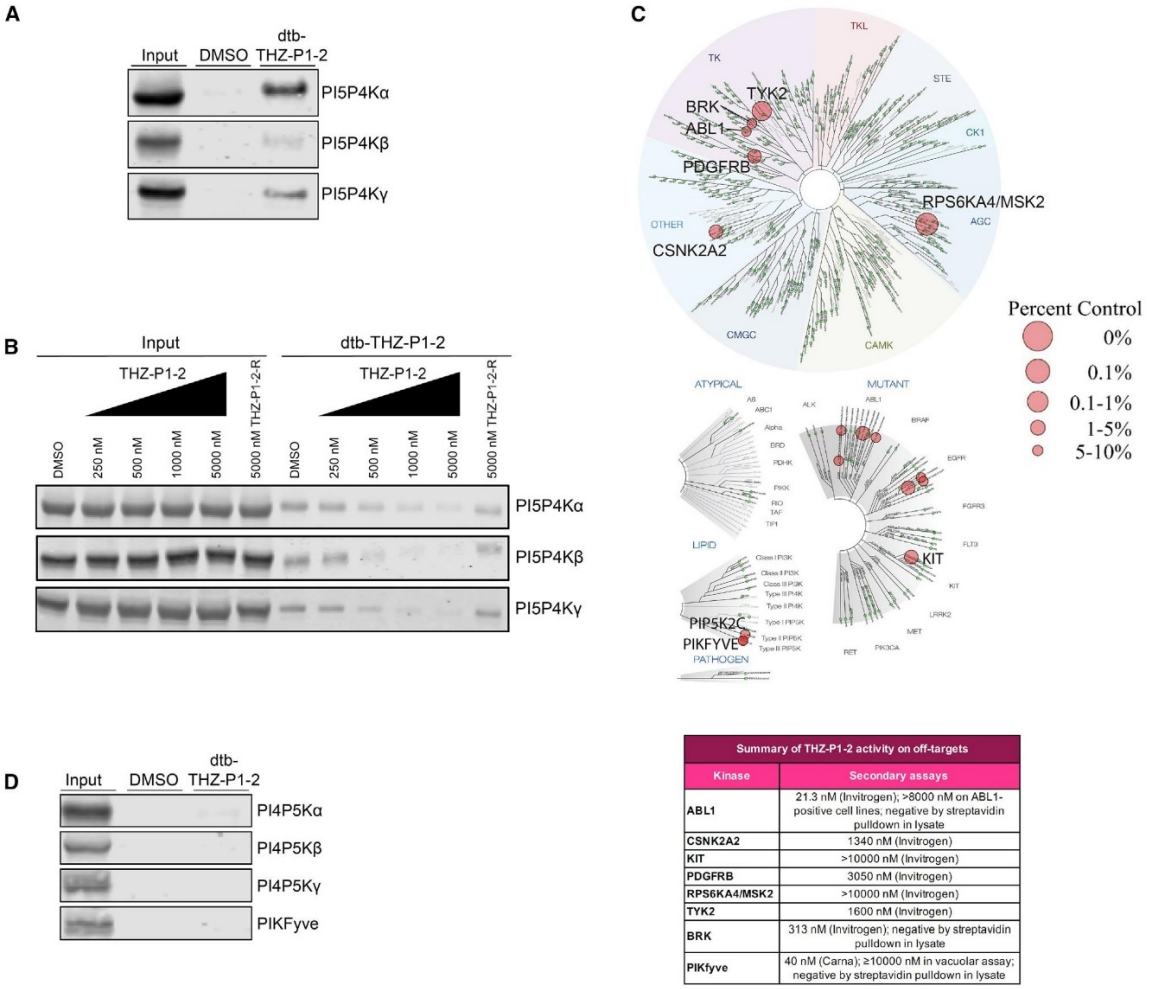
(A) Electrospray mass spectrometry of recombinant PI5P4Kα/β/γ incubated with THZ-P1-2 demonstrates covalent labeling of PI5P4K isoforms. See also Figure S2B.

(B) Subsequent protease digestion and tandem mass spectrometry confirms that THZ-P1-2 covalently labels cysteine residues. See also Figure S2A.

(C) Crystal structure of PI5P4Kα in complex with THZ-P1-2, colored according to B factor and shown with covalent warhead extended out toward the covalently targeted cysteine, C293 (labeled; not resolved in crystal structure). See also Table S1.

(D) Ligand interaction map of THZ-P1-2 with residues in the ATP-binding pocket of PI5P4Kα. Related to Figures S2C and S2D.

(E) Modeled THZ-P1-2 binding showed for all three isoforms based on alignment of published PI5P4Kβ (PDB: 1BO1; Rao et al., 1998) and γ (PDB: 2GK9; Uppenberg et al., 2006) structures with obtained PI5P4Kα structure. PI5P4Kα (magenta), PI5P4Kβ (blue), and PI5P4Kγ (orange).



Summary of THZ-P1-2 activity on off-targets	
Kinase	**Secondary assays**
ABL1	21.3 nM (Invitrogen); >8000 nM on ABL1-positive cell lines; negative by streptavidin pulldown in lysate
CSNK2A2	1340 nM (Invitrogen)
KIT	>10000 nM (Invitrogen)
PDGFRB	3050 nM (Invitrogen)
RPS6KA4/MSK2	>10000 nM (Invitrogen)
TYK2	1600 nM (Invitrogen)
BRK	313 nM (Invitrogen); negative by streptavidin pulldown in lysate
PIKfyve	40 nM (Carna); ≥10000 nM in vacuolar assay; negative by streptavidin pulldown in lysate

Figure 3. Cellular On-Target Engagement and Selectivity Profile of THZ-P1-2

(A) A desthiobiotinylated analog of THZ-P1-2 irreversibly engages all PI5P4K isoforms in a streptavidin pulldown in HEK293T lysate.

(B) THZ-P1-2, but not its reversible analog, exhibits dose-dependent on-target engagement of PI5P4K isoforms in a competitive streptavidin pulldown with 1 μM desthiobiotinylated-THZ-P1-2. See also Figure S3B.

(C) Kinome selectivity profile of THZ-P1-2. Complete dataset for off-targets included as Table S2. See also Figures S4A–S4C.

(D) THZ-P1-2 does not engage the PI4P5Ks or PIKfyve in a streptavidin pulldown. See also Figures S3C.

Inhibition of PI5P4K by THZ-P1-2 Results in Autophagy Defects

Recently, PI5P4K was demonstrated to be necessary for the proper completion of autophagy in mouse models (Lundquist et al., 2018). Loss of *PIP4K2A* and *PIP4K2B* in mouse embryonic fibroblasts and mouse liver leads to accumulation of lysosomes and autophagosomes. Lysosomes are enlarged, fused to multiple autophagosomes, and clustered at the nuclear membrane, indicating a defect in the autophagy process. The nuclear localization of the master autophagy transcription factor TFEB was also increased and downstream transcriptional targets were upregulated, indicating an activation of the autophagy gene program (Settembre et al., 2011). However, it has yet to be demonstrated whether these defects were caused by the loss of the full-length protein or the enzymatic activity.

To validate that PI5P4K enzymatic activity is necessary for completion of autophagy, we sought to demonstrate that THZ-P1-2 phenocopies the genetic loss of PI5P4Kα and PI5P4Kβ. We first treated HeLa cells with either DMSO or THZ-P1-2 for 18 h and stained for the lysosomal and autophagosomal markers LAMP1 and LC3B, respectively. Similar to the genetic knockouts of PI5P4K, treatment with THZ-P1-2 at concentrations as low as 250 nM results in the formation of numerous, enlarged LAMP1 puncta with fusion defects to LC3B-stained autophagosomes (Figure 5A). This upregulation in autophagosomal/autolysosomal puncta was accompanied by little to no increase in LAMP1

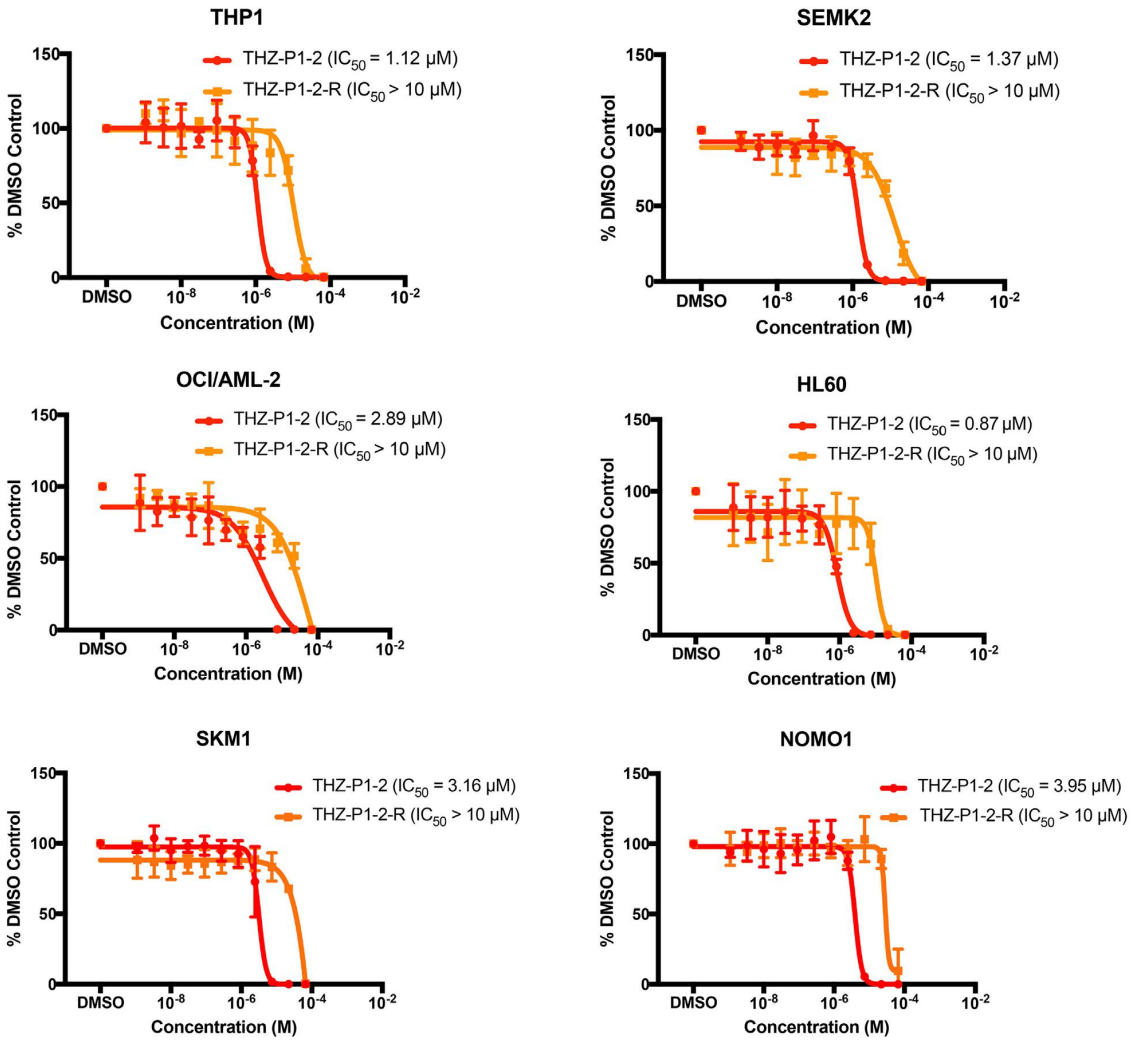


Figure 4. Preliminary Cancer Cell Line Profiling with THZ-P1-2 to Identify Potential PI5P4K Dependencies

A panel of AML and ALL cell lines were treated with THZ-P1-2 and THZ-P1-2-R for 72 h, and cell proliferation was measured with a CellTiter-Glo luminescence assay. Data shown are the mean of three biological replicates ± SEM.

protein levels with THZ-P1-2 treatment for 24 h, and an increase in LC3B-II and p62 protein levels in serum starvation conditions (HeLa cells cultured in medium supplemented with 0.3% fetal bovine serum) when compared with known autophagy inhibitors bafilomycin A1 and chloroquine (Figure S5A). Interestingly, we did not observe cumulative effects on autophagy markers by western blot when THZ-P1-2 treated in combination with these commercially available inhibitors. However, using an RFP-GFP-LC3B tandem-tagged sensor, we were able to further confirm that THZ-P1-2 treatment leads to a marked increase in GFP signal and a moderate increase in RFP signal, suggesting a disruption in autophagy flux, more specifically the induction and upregulation of autophagy without the autolysosome fusion that leads to GFP fluorescence signal loss (Figure S5B). We also observed an increase in nuclear TFEB accumulation (Figures 5B and 5C) and the upregulation of downstream TFEB targets with THZ-P1-2 treatment (Figure 5F).

We next investigated whether the effects of THZ-P1-2 are specific to PI5P4K by rescuing the autophagy defects with the expression of PI5P4Kα/β Cys-Ser mutant isoforms, which are unable to bind the inhibitor covalently. We infected HeLa cell lines to stably express a GFP-only vector, GFP-tagged PI5P4Kα/β, and GFP-tagged PI5P4K Cys-Ser mutants (α: C293S; β: C307,318S). We treated the cells for 18 h with either DMSO or THZ-P1-2. As observed with uninfected cells, cell lines expressing GFP alone saw an increase in lysosome/autophagosome dysfunction with THZ-P1-2 treatment (Figures 5D and 5E). These effects were partially ameliorated by expression of the

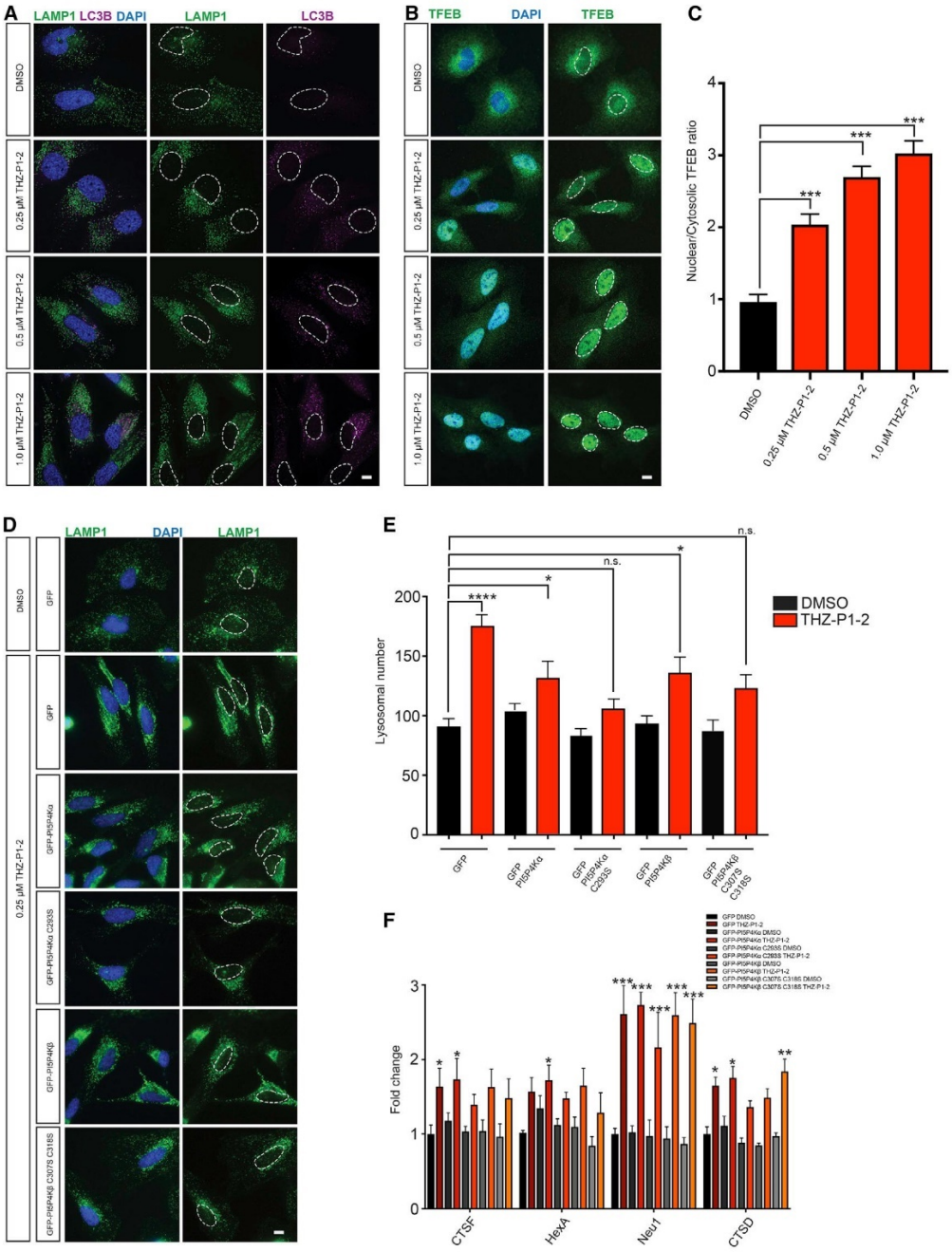


Figure 5. Inhibition of PI5P4K with THZ-P1-2 Leads to Lysosomal-Autophagosomal Defects and Increased TFEB Activation

(A) Similar to genetic loss of *PIP4K2A/B*, inhibition of PI5P4K activity with THZ-P1-2 increases LAMP1-positive lysosomal size, number, and contact with LC3B-positive autophagosomes. HeLa cells were cultured overnight with DMSO, 0.25, 0.5, or 1.0 μM THZ-P1-2 and stained for LC3B (magenta) or LAMP1 (green) with nuclei in blue. Scale bars, 10 μm. See also Figure S5.

(B) Inhibition of PI5P4K activity with THZ-P1-2 increases TFEB nuclear localization. HeLa cells were cultured overnight with DMSO, 0.25, 0.5, or 1.0 μM THZ-P1-2 and stained for TFEB (green) with nuclei in blue. Scale bars, 10 μm.

(legend continued on next page)

PI5P4K constructs and even more so by the cysteine to serine mutants (Figures 5D and 5E). The PI5P4K Cys-Ser mutants also partially rescued the nuclear localization of TFEB and the expression of downstream TFEB target genes (Figures 5D–5F).

DISCUSSION

The PI5P4Ks comprise a family of lipid kinases that regulate levels of their substrate and product PI-5-P and PI-4,5-P_2, maintaining compartment-specific concentrations of these phosphoinositides at a subcellular level. Initially thought to be mainly a PI-5-P modulator within a noncanonical PI-4,5-P_2 generation pathway, the PI5P4Ks have garnered a lot of interest in recent years because of their purported importance in specific disease contexts, spurring efforts to develop PI5P4K inhibitors. In this study, we designed and characterized THZ-P1-2, a first-in-class covalent inhibitor of the PI5P4K family. THZ-P1-2, a small molecule containing an electrophilic acrylamide moiety capable of undergoing a Michael addition, bound in the active site and irreversibly inhibited the enzymatic activity of these kinases. A covalent targeting strategy with THZ-P1-2 resulted in prolonged on-target engagement in cells and modest anti-proliferative activity in a panel of AML and ALL cell lines, while retaining a reasonable selectivity across the kinome. Finally, we found that THZ-P1-2 treatment phenocopied the effects of genetic deletion of PI5P4K $\alpha/\beta/\gamma$ in mice by causing autophagy defects in HeLa cells.

Our initial point of compound development was JNK-IN-7, a previously reported covalent JNK inhibitor that was found to be capable of binding to PI5P4Kγ by KiNativ profiling. Originally inspired by the imatinib core structure, this ''privileged'' kinase scaffold served as a promising starting point for developing a selection of analogs with chemical modifications with which to explore structure-activity relationships. Among this panel of molecules (Manz et al., 2019), we found a critical feature resulting in superior potency to be a switch from a 2,4-pyrimidine to a 4,6-pyrimidine hinge-interacting motif. We established THZ-P1-2's inhibitory activity with an apparent IC_{50} of 190 nM on PI5P4Kα in a fixed time point assay, the isoform with the highest kinase activity and thus easiest to analyze using an ATPase assay (ADP-Glo). We confirmed that THZ-P1-2 specifically prevents phosphate transfer to the PI-5-P substrate at nanomolar concentrations for all isoforms with an orthogonal TLC assay measuring radiolabeled PI-4,5-P_2 formation.

THZ-P1-2 covalently labels all three PI5P4K isoforms, with the reactive cysteine residues located on analogous disordered loops of unknown function outside the kinase domain. There has been a resurgence of interest in covalent inhibitors, especially with recent clinical successes of cysteine-directed covalent BTK inhibitor ibrutinib (Pan et al., 2007), epidermal growth factor receptor inhibitors neratinib (Rabindran et al., 2004), afatinib (Li et al., 2008), and osimertinib (Cross et al., 2014). Covalent irreversible inhibitors characteristically exhibit an enhanced potency and prolonged pharmacodynamics due to permanent disabling of kinase enzymatic activity, and high selectivity as covalent inhibitors will only bind irreversibly to a kinase with an appropriately placed cysteine within the trajectory of the electrophilic warhead (Liu et al., 2013). Our approach in developing probes THZ1 against CDK7 (Kwiatkowski et al., 2014), TH531 against CDK12/13 (Zhang et al., 2016), and most recently JZ128 against PKN3 (Browne et al., 2018) showcased to great effect the modification of a reactive cysteine residue outside the active site, widening the scope of ''accessible'' cysteines. We sought to apply these principles and strategies toward PI5P4K. Our study validates the PI5P4Ks as additional candidate kinases with cysteines that are distant from the active site in sequence but brought close to the ATP-binding site due to kinase tertiary structure, rendering them amenable to covalent targeting. Moreover, our approach in targeting these unique cysteines provides a proof-of-concept in using a combination of inhibitor covalent and noncovalent affinities to achieve selectivity, due to a lack of equivalent off-target cysteines across the kinome.

Structural insights into the binding mode of THZ-P1-2 with PI5P4Kα were obtained through a co-crystal structure. As underscored in the results for Figures 2C–2E, we were unable to resolve the covalent bond between the acrylamide warhead and Cys293 likely due to the degree of disorder in the cysteine-containing loop, as demonstrated by the lack of electron density (Figure S2D). However, we believe the combination of techniques in our study, particularly the mass spectrometry (MS) results (Figures 2A and 2B) and Cys-Ser rescue in autophagy dysfunction (Figures 5D–5F), led to observations that taken together support the covalent binding of THZ-P1-2 to the specific PI5P4K cysteine residues, once again emphasizing the value of our multipronged approach. Besides the noncovalent interactions highlighted in the results, hydrogen-bonding between the backbone of Val199 and the 4,6-pyrimidine may account for the increased potency compared with analogs with a 2,4-pyrimidine, especially if they contain substituents on the heterocycle linker. Notably, the Phe205 residue that engages in aromatic π-stacking interactions with the phenylenediamine moiety of THZ-P1-2 is conserved among the PI5P4Ks and has

(C) Quantification of results in (B). The intensity of TFEB immunofluorescence was quantified in the nucleus and the cytoplasm, and used to calculate the ratio. Statistical significance determined by ANOVA (***$p < 0.0005$) with Dunnett multiple comparison post-test. Each group was compared with control HeLa cells treated with DMSO ($n \geq 30$).

(D) Expression of PI5P4K cysteine to serine mutants alleviates lysosomal dysfunction induced by THZ-P1-2 treatment. HeLa cells were infected with viruses expressing GFP, GFP-PI5P4Kα, GFP- PI5P4Kα C293S, GFP-PI5P4Kβ, or GFP-PI5P4Kβ C307S C318S and treated with either DMSO or 250 nM THZ-P1-2 overnight. Expression of both the PI5P4Kα and PI5P4Kβ cysteine to serine mutants alleviated dysfunctional lysosomes (green) with nuclei in blue. Scale bars, 10 μm.

(E) Quantification of results in (D). The number of lysosomes was quantified per cell. Statistical significance determined by ANOVA (***$p < 0.0005$) with Dunnett multiple comparison post-test. Each group was compared with control HeLa cells expressing GFP and treated with DMSO ($n \geq 30$).

(F) HeLa cells expressing GFP, GFP-PI5P4Kα, GFP- PI5P4Kα C293S, GFP-PI5P4Kβ, or GFP-PI5P4Kβ C307S C318S were treated with either DMSO or 250 nM THZ-P1-2 overnight for 16 h and subsequently harvested for qPCR of TFEB targets. Fold change is calculated by comparison with HeLa cells expressing GFP treated with DMSO. *$p < 0.05$, Student's t test ($n \geq 8$).

been identified as the residue responsible for the GTP-sensing ability of PI5P4Kβ (Sumita et al., 2016). This co-crystal structure also serves as an important resource for the scientific community, enabling structure-guided design of PI5P4K inhibitors with improved properties and continuing to bring to attention the challenges of structurally capturing the targeting of regions of disorder. Cell permeability and dose-dependent on-target engagement of THZ-P1-2 was validated in HEK293T cells, chosen for their robust expression of all three PI5P4K isoforms, and the necessity for covalency verified through observing effects after washout and in anti-proliferation assays alongside THZ-P1-2-R.

Our first bona fide look at covalent-dependent on-target engagement was using a streptavidin pulldown assay with a desthiobiotinylated analog of THZ-P1-2 (Figure 3). In the direct pulldown, THZ-P1-2 engages with PI5P4K preferentially over PI4P5K and PIKfyve. Because we saw a very mild decrease in activity in the ADP-Glo biochemical assay from covalent to reversible compound (190–319 nM, Figure S3A), but observed that THZ-P1-2-R was not able to compete and prevent pulldown by dtb-THZ-P1-2, this indicates that the covalent contribution of the compound's binding mechanism truly comes into effect and is relevant in the cellular context. Furthermore, this also suggests that our conditions for the ADP-Glo assay may really only be accurate for determining the noncovalent binding potency for this family of kinases, with the covalent aspect needing to be verified in cellular assays.

Our biochemical selectivity profiling of THZ-P1-2 indicates that, although the compound exhibits satisfactory selectivity, there are some off-targets observed. The most potent off-targets, such as ABL1, PIKfyve, and BRK, were not engaged by dtb-THZ-P1-2 in a cellular pulldown assay. PIKfyve was inhibited by THZ-P1-2 only at concentrations of 10 μM or greater, higher than concentrations relevant for observable PI5P4K targeting. These results indicate that THZ-P1-2 does have some off-target activity on PIKfyve, but maintains a preference for the PI5P4Ks, again possibly due to a combination of covalent and noncovalent binding modes.

A principal part of our study was to further confirm recently discovered functions of the PI5P4Ks beyond their asserted primary role of noncanonical PI-4,5-P_2 generation. Treatment with THZ-P1-2 phenocopied genetic deletion of PI5P4K in causing autophagosome and lysosome dysfunction, TFEB nuclear localization, and TFEB target gene upregulation, indicating that PI5P4K kinase activity is required for the role of these kinases in autophagy. Cancer cell line profiling demonstrates AML/ALL cancer cell lines to be sensitive to THZ-P1-2 covalent targeting. Consistent with findings that PI5P4K mediates autophagy in times of nutrient stress (Lundquist et al., 2018) our studies using THZ-P1-2 as a tool suggest that the PI5P4Ks display an induced essentiality by facilitating autophagy as an energy source in a metabolically stressed environment, such as cancer (Kimmelman and White, 2017). Additional studies are required to fully dissect the connection between PI5P4K inhibition, autophagy flux disruption, and subsequent effects on cell viability, as well as investigate the sensitivities of various cancer types to specific PI5P4K inhibition. THZ-P1-2 may be a useful probe of the pharmacological consequences of covalent inhibition of PI5P4K that can complement genetic approaches (such as CRISPR- or RNAi-based screening) to identify unique vulnerabilities in cancer.

In conclusion, we have characterized a potent and selective covalent PI5P4K inhibitor that exhibits durable cellular pharmacodynamics, causes autophagy dysfunction, and demonstrates modest anti-proliferative activity against leukemia-derived cell lines. THZ-P1-2, THZ-P1-R, and the inhibitor-resistant PI5P4K Cys to Ser mutants described in this article provide a chemical biology toolbox serving as a valuable resource with which to investigate the pharmacological consequences of covalent inhibition of PI5P4Ks. Our discovery of THZ-P1-2 may be used as a model for targeting other understudied gene families, especially those bearing targetable cysteines in unique sites. Finally, our small-molecule strategy for disrupting autophagy through covalent PI5P4K inhibition exposes a new Achilles heel and indicates that exploiting this metabolic vulnerability may be a viable therapeutic approach in cancer and potentially other autophagy-addicted disorders.

SIGNIFICANCE

Our discovery and characterization of a pan-PI5P4K inhibitor, THZ-P1-2, that covalently targets cysteines on a disordered loop in PI5P4Kα/β/γ presents a new way to target kinases, especially noncanonical lipid kinases, by taking advantage of unique unannotated cysteines outside the kinase domain. THZ-P1-2 demonstrates cellular on-target engagement with limited off-targets across the kinome, displaying requisite potency and selectivity to be a useful tool compound. The sensitivity of leukemia cancer cells to THZ-P1-2 covalent inhibition is evidence of the potential of THZ-P1-2 to be used more universally in larger-scale cancer cell line screens to identify PI5P4K dependencies in other cancer types, or other diseases beyond cancer. This can be done in parallel with typical genetic approaches (such as CRISPR- or RNAi-based screening) to identify unique vulnerabilities in a cheaper, faster, less labor-intensive manner. One of the major advantages of THZ-P1-2, as with covalent inhibitors in general, is the ability to verify on-target activity through inhibitor-resistant cysteine to serine mutations. In the case of the PI5P4K family of enzymes which has three isoforms, the utility of having both covalent and reversible compounds is extremely valuable, where both compounds can be used in combination to broadly assess PI5P4K dependencies in disease, with follow-up mutant rescue studies to verify the PI5P4K contribution once the scope is narrowed down. Finally, THZ-P1-2-induced autophagy dysfunction phenocopying the effects of PI5P4K genetic deletion lends confidence to the on-target activity of the compound and presents an alternative strategy for targeting autophagy-lysosome homeostasis in disease. Our studies demonstrate that PI5P4Ks are tractable targets, with THZ-P1-2, its reversible and desthiobiotinylated analogs, and wild-type and mutant cell lines as a useful toolbox to further interrogate the therapeutic potential of PI5P4K inhibition, providing a new starting point for drug discovery campaigns for these lipid kinases in cancer metabolism and other autophagy-dependent disorders.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Chemistry
  - ADP-Glo Kinase Assays
  - Radiometric Kinase Assays
  - Immunoblotting
  - Pulldown Assays
  - Mass Spectrometry
  - Protein Expression and Purification
  - Crystallization
  - Data Collection and Structure Determination
  - Proliferation Assays
  - qRT-PCR
  - Fluorescence Microscopy
  - Autophagy Flux Assays
  - Generation and Transduction of GFP *PIP4K2* Lentiviral Constructs
  - Virus Production and Infection
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chembiol.2020.02.003.

ACKNOWLEDGMENTS

The authors acknowledge generous support from NIH grants 1U19AI109740 (to J.M.C.), R21 CA178860 (to J.A.M.), R21CA188881 (to J.A.M.), R35 CA197588 (to L.C.C.), R01 GM041890 (to L.C.C.), U54 U54CA210184 (to L.C.C.), R01 CA197329 (to N.S.G. and S.D.-P.), the Dana–Farber Strategic Research Initiative (to J.A.M.), and the Breast Cancer Research Foundation (to L.C.C.). Synchrotron data collection was based on research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines (NIGMS P41 GM103403). The authors acknowledge the Nikon Imaging Center at Harvard Medical School for valuable training and support with imaging experiments. The authors thank the James D. Griffin and James E. Bradner labs (Dana-Farber Cancer Institute) for generously donating cell lines, and Milka Kostic for advice and guidance with the manuscript submission and revision.

AUTHOR CONTRIBUTIONS

L.C.C., H.S., T.Z., and N.S.G. conceived the original project. S.C.S. led the project and designed and performed the majority of experiments. T.Z., B.J., M.-F.H., and T.D.M. designed and synthesized compounds. C.M.B. and S.B.F. conducted MS experiments. M.R.L. performed autophagy experiments (except autophagy flux experiments performed by S.C.S.). M.N.P. performed TLC experiments. H.-S.S. performed protein purification and crystallography experiments. H.-S.S. and S.D.-P. performed crystallography data collection and analysis. P.K., D.G.W., and T.J.Y. performed additional experiments. N.P.K. provided expertise and feedback. J.M.C., J.A.M., S.D.-P., L.C.C., and N.S.G. supervised the research. L.C.C. and N.S.G. secured funding. S.C.S. wrote the manuscript, with guidance from F.M.F., T.Z., and N.P.K. S.C.S., H.S., F.M.F., T.D.M., and N.S.G. revised the manuscript. All authors read and provided feedback on the manuscript.

DECLARATION OF INTERESTS

J.A.M. is a member of the scientific advisory board (SAB) of 908 Devices. L.C.C. is a founder and member of the Board of Directors of Agios Pharmaceuticals and is a founder and receives research support from Petra Pharmaceuticals. These companies are developing novel therapies for cancer. N.S.G. is a founder, SAB member, and equity holder in Gatekeeper, Syros, Petra, C4, B2S, and Soltego. The Gray lab receives or has received research funding from Novartis, Takeda, Astellas, Taiho, Janssen, Kinogen, Voronoi, Her2llc, Deerfield, and Sanofi. N.S.G., T.Z., and N.P.K. are inventors on a patent application covering chemical matter in this publication owned by the Dana-Farber Cancer Institute.

Received: May 13, 2019
Revised: November 14, 2019
Accepted: February 13, 2020
Published: March 3, 2020

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Li-Wei, H., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Struct. Biol. *66*, 213–221.

Al-ramahi, I., Giridharan, S.S.P., Chen, Y.C., Patnaik, S., Safren, N., Hasegawa, J., de Haro, M., Wagner Gee, A.K., Titus, S.A., Jeong, H., and Clarke, J. (2017). Inhibition of PIP4Kγ ameliorates the pathological effects of mutant huntingtin protein. Elife *6*, https://doi.org/10.7554/eLife.29123.

Balla, T., Szentpetery, Z., and Kim, Y.J. (2009). Phosphoinositide signaling: new tools and insights. Physiology *24*, 231–244.

De Braekeleer, E., Douet-Guilbert, N., Rowe, D., Bown, N., Morel, F., Berthou, C., Férec, C., and De Braekeleer, M. (2011). ABL1 fusion genes in hematological malignancies: a review. Eur. J. Haematol. *86*, 361–371.

Browne, C.M., Jiang, B., Ficarro, S.B., Doctor, Z.M., Johnson, J.L., Card, J.D., Sivakumaren, S.C., Alexander, W.M., Yaron, T.M., Murphy, C.J., et al. (2018). A chemoproteomic strategy for direct and proteome-wide covalent inhibitor target-site identification. J. Am. Chem. Soc. *141*, 191–203.

Bulley, S.J., Clarke, J.H., Droubi, A., Giudici, M.L., and Irvine, R.F. (2015). Exploring phosphatidylinositol 5-phosphate 4-kinase function. Adv. Biol. Regul. *57*, 193–202.

Bulley, S.J., Droubi, A., Clarke, J.H., Anderson, K.E., Stephens, L.R., Hawkins, P.T., and Irvine, R.F. (2016). B cells, phosphatidylinositol 5-phosphate 4-kinase-α synthesizes PI(4,5)P2 to impact mTORC2 and Akt signaling. Proc. Natl. Acad. Sci. U S A *113*, 10571–10576.

Cai, X., Xu, Y., Cheung, A.K., Tomlinson, R.C., Alcázar-Román, A., Murphy, L., Billich, A., Zhang, B., Feng, Y., Klumpp, M., and Rondeau, J.M. (2013). PIKfyve, a class III PI kinase, is the target of the small molecular IL-12/IL-23 inhibitor apilimod and a player in Toll-like receptor signaling. Chem. Biol. *20*, 912–921.

Choy, C.H., Saffi, G., Gray, M.A., Wallace, C., Dayam, R.M., Ou, Z.A., Lenk, G., Puertollano, R., Watkins, S.C., and Botelho, R.J. (2018). Lysosome enlargement during inhibition of the lipid kinase PIKfyve proceeds through lysosome coalescence. J. Cell Sci. *131*, https://doi.org/10.1242/jcs.213587.

Clarke, J.H., Giudici, M., Burke, J.E., Williams, R.L., Maloney, D.J., Marugan, J., and Irvine, R.F. (2015). The function of phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ) explored using a specific inhibitor that targets the PI5P-binding site. Biochem. J. *367*, 359–367.

Cross, D.A., Ashton, S.E., Ghiorghiu, S., Eberlein, C., Nebhan, C.A., Spitzler, P.J., Orme, J.P., Finlay, M.R., Ward, R.A., Mellor, M.J., and Hughes, G. (2014). AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. Cancer Discov. *4*, 1046–1061.

Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., and Zarrinkar, P.P. (2011). Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol. *29*, 1046–1051.

Davis, M.I., Sasaki, A.T., Shen, M., Emerling, B.M., Thorne, N., Michael, S., Pragani, R., Boxer, M., Sumita, K., Takeuchi, K., et al. (2013). A homogeneous, high-throughput assay for phosphatidylinositol 5-phosphate 4-kinase with a novel, rapid substrate preparation. PLoS One *8*, e54127.

Divecha, N., Rhee, S.G., Letcher, A.J., and Irvine, R.F. (1993). Phosphoinositide signalling enzymes in rat liver nuclei: phosphoinositidase C isoform beta 1 is specifically, but not predominantly, located in the nucleus. Biochem. J. *289* (Pt 3), 617–620.

Emerling, B.M., Hurov, J.B., Poulogiannis, G., Tsukazawa, K.S., Choo-Wing, R., Wulf, G.M., Bell, E.L., Shim, H.S., Lamia, K.A., Rameh, L.E., et al. (2013). XDepletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-Null tumors. Cell *155*, 844–857.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fiume, R., Keune, W.J., Faenza, I., Bultsma, Y., Ramazzotti, G., Jones, D.R., Martelli, A.M., Somner, L., Follo, M.Y., Divecha, N., and Cocco, L. (2012). Nuclear phosphoinositides: location, regulation and function. Subcell. Biochem. *59*, 335–361.

Fiume, R., Stijf-Bultsma, Y., Shah, Z.H., Keune, W.J., Jones, D.R., Jude, J.G., and Divecha, N. (2015). PIP4K and the role of nuclear phosphoinositides in tumour suppression. Biochim. Biophys. Acta *1851*, 898–910.

Hasegawa, J., Strunk, B.S., and Weisman, L.S. (2017). PI5P and PI(3,5)P2: minor, but essential phosphoinositides. Cell Struct. Funct. *42*, 49–60.

Hu, A., Zhao, X., Tu, H., Xiao, T., Fu, T., Wang, Y., Liu, Y., Shi, X.J., Luo, J., and Song, B. (2018). PIP4K2A regulates intracellular cholesterol transport through modulating PI(4,5)P2homeostasis. J. Lipid Res. *59*, 507–514.

Jude, J.G., Spencer, G.J., Huang, X., Somerville, T.D.D., Jones, D.R., Divecha, N., and Somervaille, T.C.P. (2014). A targeted knockdown screen of genes coding for phosphoinositide modulators identifies PIP4K2A as required for acute myeloid leukemia cell proliferation and survival. Oncogene *34*, 1–10.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr. D Biol. Crystallogr. *66*, 133–144.

Karaman, M.W., Herrgard, S., Treiber, D.K., Gallant, P., Atteridge, C.E., Campbell, B.T., Chan, K.W., Ciceri, P., Davis, M.I., Edeen, P.T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. Nat. Biotechnol. *26*, 127–132.

Keune, W.J., Jones, D.R., Bultsma, Y., Sommer, L., Zhou, X.Z., Lu, K.P., and Divecha, N. (2012). Regulation of phosphatidylinositol-5-phosphate signaling by Pin1 determines sensitivity to oxidative stress. Sci. Signal. *5*, 1–13.

Keune, W.J., Jones, D.R., and Divecha, N. (2013). PtdIns5P and Pin1 in oxidative stress signaling. Adv. Biol. Regul. *53*, 179–189.

Kimmelman, A.C., and White, E. (2017). Autophagy and tumor metabolism. Cell Metab. *25*, 1037–1043.

Kitagawa, M., Liao, P., Lee, K.H., Wong, J., Shang, S.C., Minami, N., Sampetrean, O., Saya, H., Lingyun, D., Prabhu, N., et al. (2017). Mitotic pathways leads to cancer-selective lethality. Nat. Commun. *8*, 2200.

Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature *511*, 616–620.

Lamia, K.A., Peroni, O.D., Kim, Y., Rameh, L.E., Kahn, B.B., and Cantley, L.C. (2004). Increased insulin sensitivity and reduced adiposity in phosphatidylinositol 5-phosphate 4-kinase –/– mice. Mol. Cell Biol. *24*, 5080–5087.

Li, D., Ambrogio, L., Shimamura, T., Kubo, S., Takahashi, M., Chirieac, L.R., Padera, R.F., Shapiro, G.I., Baum, A., Himmelsbach, F., et al. (2008). BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. Oncogene *27*, 4702–4711.

Liu, Q., Sabnis, Y., Zhao, Z., Zhang, T., Buhrlage, S.J., Jones, L.H., and Gray, N.S. (2013). Developing irreversible inhibitors of the protein kinase cysteinome. Chem. Biol. *20*, 146–159.

Lundquist, M.R., Goncalves, M.D., Loughran, R.M., Possik, E., Vijayaraghavan, T., Yang, A., Pauli, C., Ravi, A., Verma, A., Yang, Z., et al. (2018). Phosphatidylinositol-5-phosphate 4-kinases regulate cellular lipid metabolism by facilitating autophagy. Mol. Cell *70*, 531–544.

Ma, G., Duan, Y., Huang, X., Qian, C.X., Chee, Y., Mukai, S., Cui, J., Samad, A., Matsubara, J.A., Kazlauskas, A., et al. (2016). Prevention of proliferative vitreoretinopathy by suppression of phosphatidylinositol 5-phosphate 4-kinases. Invest. Opthalmol. Vis. Sci. *57*, 3935.

Manz, T.D., Sivakumaren, S.C., Yasgar, A., Hall, M.D., Davis, M.I., Seo, H.-S., Card, J.D., Ficarro, S.B., Shim, H., Marto, J.A., et al. (2019). Structure-activity relationship study of covalent pan-phosphatidylinositol 5-phosphate 4-kinase inhibitors. ACS Med. Chem. Lett. https://doi.org/10.1021/acsmedchemlett.9b00402.

Martelli, A.M., Gilmour, R.S., Bertagnolo, V., Neri, L.M., Manzoli, L., and Cocco, L. (1992). Nuclear localization and signalling activity of phosphoinositidase C beta in Swiss 3T3 cells. Nature *358*, 242–245.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Naviaux, R.K., Costanzi, E., Haas, M., and Verma, I.M. (1996). The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. J. Virol. *70*, 5701–5705.

Pan, Z., Scheerens, H., Li, S.J., Schultz, B.E., Sprengeler, P.A., Burrill, L.C., Mendonca, R.V., Sweeney, M.D., Scott, K.C., Grothaus, P.G., et al. (2007). Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. ChemMedChem *2*, 58–61.

Perera, R.M., Stoykova, S., Nicolay, B.N., Ross, K.N., Fitamant, J., Boukhali, M., Lengrand, J., Deshpande, V., Selig, M.K., Ferrone, C.R., et al. (2015). Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. Nature *524*, 361–365.

Rabindran, S.K., Discafani, C.M., Rosfjord, E.C., Baxter, M., Floyd, M.B., Golas, J., Hallett, W.A., Johnson, B.D., Nilakantan, R., Overbeek, E., et al. (2004). Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. Cancer Res. *64*, 3958–3965.

Rameh, L.E., and Cantley, L.C. (1999). The role of phosphoinositide 3-kinase lipid products in cell function. J. Biol. Chem. *274*, 8347–8350.

Rameh, L.E., Tolias, K.F., Duckworth, B.C., and Cantley, L.C. (1997). A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate. Nature *390*, 192–196.

Rao, V.D., Misra, S., Boronenkov, I.V., Anderson, R.A., and Hurley, J.H. (1998). Structure of type IIβ phosphatidylinositol phosphate kinase. Cell *94*, 829–839.

Rosales-Rodríguez, B., Fernández-Ramírez, F., Núñez-Enríquez, J.C., Velázquez-Wong, A.C., Medina-Sansón, A., Jiménez-Hernández, E., Flores-Lujano, J., Peñaloza-González, J.G., Espinosa-Elizondo, R.M., Pérez-Saldívar, M.L., et al. (2016). Copy number alterations associated with acute lymphoblastic leukemia in Mexican children. A report from the Mexican inter-institutional group for the identification of the causes of childhood leukemia. Arch. Med. Res. *47*, 706–711.

Settembre, C., Di Malta, C., Polito, V.A., Garcia Arencibia, M., Vetrini, F., Erdin, S., Erdin, S.U., Huynh, T., Medina, D., Colella, P., et al. (2011). TFEB links autophagy to lysosomal biogenesis. Science *332*, 1429–1433.

Sharma, G., Guardia, C.M., Roy, A., Vassilev, A., Saric, A., Griner, L.N., Marugan, J., Ferrer, M., Bonifacino, J.S., and DePamphilis, M.L. (2019). A family of PIKFYVE inhibitors with therapeutic potential against autophagy-dependent cancer cells disrupt multiple events in lysosome homeostasis. Autophagy *15*, 1694–1718.

Shim, H., Wu, C., Ramsamooj, S., Bosch, K.N., Chen, Z., and Emerling, B.M. (2016). Deletion of the gene Pip4k2c, a novel phosphatidylinositol kinase, results in hyperactivation of the immune system. Proc. Natl. Acad. Sci. U S A *113*, 7596–7601.

Stijf-Bultsma, Y., Sommer, L., Tauber, M., Baalbaki, M., Giardoglou, P., Jones, D.R., Gelato, K.A., van Pelt, J., Shah, Z., Rahnamoun, H., et al. (2015). The basal transcription complex component TAF3 transduces changes in nuclear phosphoinositides into transcriptional output. Mol. Cell *58*, 453–467.

Sumita, K., Lo, Y.H., Takeuchi, K., Senda, M., Kofuji, S., Ikeda, Y., Terakawa, J., Sasaki, M., Yoshino, H., Majd, N., and Zheng, Y. (2016). The lipid kinase PI5P4Kβ is an intracellular GTP sensor for metabolism and tumorigenesis. Mol. Cell *61*, 187–198.

Uppenberg, J., Hogbom, M., Ogg, D., Arrowsmith, C., Berglund, H., Collins, R., Ehn, M., Flodin, S., Flores, A., Graslund, S., et al. (2006). Human phosphatidylinositol-4-phosphate 5-kinase, type II, gamma. Struct. Genomics Consortium. https://doi.org/10.2210/pdb2gk9/pdb.

Urayama, K.Y., Takagi, M., Kawaguchi, T., Matsuo, K., Tanaka, Y., Ayukawa, Y., Arakawa, Y., Hasegawa, D., Yuza, Y., Kaneko, T., et al. (2018). Regional evaluation of childhood acute lymphoblastic leukemia genetic susceptibility loci among Japanese. Sci. Rep. *8*, 789.

Vicinanza, M., Korolchuk, V.I., Ashkenazi, A., Puri, C., Menzies, F.M., Clarke, J.H., and Rubinsztein, D.C. (2015). PI(5)P regulates autophagosome biogenesis. Mol. Cell *57*, 219–234.

Voss, M.D., Czechtizky, W., Li, Z., Rudolph, C., Petry, S., Brummerhop, H., Langer, T., Schiffer, A., and Schaefer, H.L. (2014). Discovery and pharmacological characterization of a novel small molecule inhibitor of phosphatidylinositol-5-phosphate 4-kinase, type II, beta. Biochem. Biophysical Res. Commun. *449*, 327–331.

Zhang, T., Inesta-Vaquera, F., Niepel, M., Zhang, J., Ficarro, S.B., Machleidt, T., Xie, T., Marto, J.A., Kim, N., Sim, T., et al. (2012). Discovery of potent and selective covalent inhibitors of JNK. Chem. Biol. *19*, 140–154.

Zhang, T., Kwiatkowski, N., Olson, C.M., Dixon-Clarke, S.E., Abraham, B.J., Greifenberg, A.K., Ficarro, S.B., Elkins, J.M., Liang, Y., Hannett, N.M., et al. (2016). Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. Nat. Chem. Biol. *12*, 876–884.

Zheng, L., and Conner, S.D. (2018). PI5P4Kγ functions in DTX1-mediated Notch signaling. Proc. Natl. Acad. Sci. U S A *115*, E1983–E1990.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti PI5P4Kα	Cell Signaling	Cat# 5527; RRID: AB_2722636
Rabbit monoclonal anti PI5P4Kβ	Cell Signaling Technologies	Cat# 9694; RRID: AB_2164572
Rabbit monoclonal anti PI5P4Kγ	Sigma-Aldrich	Cat# HPA058551; RRID: AB_2683752
Rabbit monoclonal anti PI4P5Kα	Cell Signaling Technologies	Cat# 9693; RRID: AB_2164698
Rabbit monoclonal anti PI4P5Kβ	Proteintech	Cat# 12541-1-AP; RRID: AB_10734598
Rabbit monoclonal anti PI4P5Kγ	Cell Signaling Technologies	Cat# 3296; RRID: AB_2164719
Mouse monoclonal anti PIKfyve	EMD Millipore	Cat# MABS522; RRID: AB_1904178
Mouse monoclonal anti c-Abl	Santa Cruz	Cat# sc-56887; RRID: AB_781732
Mouse monoclonal anti Brk	Santa Cruz	Cat# sc-137045; RRID: AB_2174228
Rabbit monoclonal anti IRAK1	Cell Signaling Technologies	Cat# 4504; RRID: AB_1904032
Mouse polyclonal anti PKN3	Novus Biologicals	Cat# H00029941-B01P; RRID: AB_2237066
Rabbit monoclonal anti SAP/JNK	Cell Signaling Technologies	Cat# 9252; RRID: AB_2250373
Mouse monoclonal anti CDK7	Cell Signaling Technologies	Cat# 2916; RRID: AB_2077142
Rabbit polyclonal anti CDK12	Cell Signaling Technologies	Cat# 11973; RRID: AB_2715688
Rabbit monoclonal anti LAMP1	Cell Signaling Technologies	Cat# 9091; RRID: AB_2687579
Rabbit monoclonal anti LC3B	Cell Signaling Technologies	Cat# 2775; RRID: AB_915950
Rabbit monoclonal anti LC3B	Cell Signaling Technologies	Cat# 3868; RRID: AB_2137707
Rabbit polyclonal anti TFEB	Sigma-Aldrich	Cat# SAB4503154; RRID: AB_10761398
Rat monoclonal anti LAMP1	Abcam	Cat# ab25245; RRID: AB_449893
Mouse monoclonal anti Tubulin	Cell Signaling Technologies	Cat# 3873; RRID: AB_1904178
Rabbit polyclonal anti Vinculin	Cell Signaling Technologies	Cat# 4650; RRID: AB_10559207
IRDye® 800CW Goat anti-Rabbit IgG (H + L) antibody	LI-COR Biosciences	Cat# 925-32211; RRID: AB_2651127
IRDye 680LT Goat anti-Mouse IgG (H + L) antibody	LI-COR Biosciences	Cat# P/N 925-68020; RRID: AB_2687826
Donkey anti-Rabbit IgG Alexa Fluor 647	Thermo Fisher Scientific	Cat# A-31573; RRID: AB_2536183
Donkey anti-Mouse IgG Alexa Fluor 647	Thermo Fisher Scientific	Cat# A-31571; RRID: AB_162542
Donkey anti-Mouse IgG Alexa Fluor 488	Thermo Fisher Scientific	Cat# A-21202; RRID: AB_141607
Donkey anti-Rabbit IgG Alexa Fluor 488	Thermo Fisher Scientific	Cat# A-21206; RRID: AB_2535792
Bacterial and Virus Strains		
NEB 10-beta	NEB	Cat# C3019I
NEB Stable	NEB	Cat# C3040I
Chemicals, Peptides, and Recombinant Proteins		
PI5P diC16	Echelon Biosciences	Cat# P-5016
DPPS (16:0/16:0 PS)	Echelon Biosciences	Cat# L-3116
Chloroquine	Cell Signaling Technologies	Cat# 14774
Bafilomycin A1	Cell Signaling Technologies	Cat# 54645
Critical Commercial Assays		
Z'LYTE (TYK2 | Km app)	Invitrogen	Assay ID 864
Z'LYTE (RPS6KA4 (MSK2) | Km app)	Invitrogen	Assay ID 801
Z'LYTE (CSNK2A2 (CK2 alpha 2) | Km app)	Invitrogen	Assay ID 363
Z'LYTE (ABL1 | Km app)	Invitrogen	Assay ID 220
Z'LYTE (KIT | Km app)	Invitrogen	Assay ID 542
Z'LYTE (PDGFRB (PDGFR beta) | Km app)	Invitrogen	Assay ID 687

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Z'LYTE (PTK6 (Brk) | Km app)	Invitrogen	Assay ID 771
Adapta (PIP5K1A | 10)	Invitrogen	Assay ID 2340
Adapta (PIP5K1B | 10)	Invitrogen	Assay ID 2341
Adapta (PIP5K1C | 10)	Invitrogen	Assay ID 2342
ADP-Glo (PIKfyve)	Carna Biosciences	N/A
ADP-Glo kinase assay	Promega	Cat# V6930
CellTiter Glo	Promega	Cat# G7571
Premo Autophagy Tandem Sensor RFP-GFP-LC3B Kit	Thermo Fisher Scientific	Cat# P36239
CalPhos Mammalian Transfection Kit	Clontech	Cat# 631312
RNeasy Mini Kit	Qiagen	Cat# 74104
SuperScript VILO cDNA Synthesis Kit	Thermo Fisher Scientific	Cat# 11754250
Quikchange XL Site-Directed Mutagenesis Kit	Agilent	Cat# 200517
Eppendorf MasterCycler RealPlex with SYBR Green	Thermo Fisher Scientific	Cat# K0251
Deposited Data		
PI5P4Kα/THZ-P1-2 co-crystal structure	This paper	PDB: 6OSP
PI5P4Kβ crystal structure	Rao et al., 1998	PDB: 1BO1
PI5P4Kγ crystal structure	Uppenberg et al., 2006	PDB: 2GK9
Experimental Models: Cell Lines		
HeLa (gender: female)	ATCC	Cat# CCL-2
HEK293T (gender: undetermined)	ATCC	Cat# CRL-3216
THP1 (gender: male)	ATCC	Cat# TIB-202
SEMK2 (gender: female)	DSMZ	RRID:CVCL_S906
HL-60 (gender: female)	ATCC	Cat# CCL-240
SKM1 (gender: male)	DSMZ	Cat# ACC 547
NOMO1 (gender: female)	DSMZ	Cat# ACC 542
K562 (gender: female)	ATCC	Cat# CCL-243
KU812F (gender: male)	ATCC	Cat# CRL-2101
OCI/AML-2 (gender: male)	DSMZ	Cat# ACC 99
HeLa-GFP/GFP-PI5P4K WT and Cys-Ser mutants (α: C293S; β: C307,318S;) (gender: female)	This paper	N/A
Oligonucleotides		
qPCR primers	Lundquist et al., 2018	N/A
hPI5P4Kα C293S QuikChange forward primer: catcgttctcctcactctccacttcctcctg	This paper	N/A
hPI5P4Kα C293S QuikChange reverse primer: caggaggaagtggagagtgaggagaacgatg	This paper	N/A
hPI5P4Kβ C307S QuikChange forward primer: gcagaggacgaggagagtgagaatgatgggg	This paper	N/A
hPI5P4Kβ C307S QuikChange reverse primer: ccccatcattctcactctcctcgtcctctgc	This paper	N/A
hPI5P4Kβ C318S QuikChange forward primer: gtgtgccataggagctgagtaggttgccacc	This paper	N/A
hPI5P4Kβ C318S QuikChange reverse primer: ggtggcaacctactcagctcctatggcacac	This paper	N/A
Recombinant DNA		
pCL-Eco	Naviaux et al., 1996	Addgene plasmid Cat# 12371
pLP1	Lundquist et al., 2018	N/A
pLP2	Lundquist et al., 2018	N/A
pVSV-G	Lundquist et al., 2018	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pGEX-2T-PIP4K2A	This paper	N/A
pNIC28-Bsa4-PIP4K2A	This paper	N/A
pGEX-2T-PIP4K2B	This paper	N/A
pNIC28-Bsa4-PIP4K2C	This paper	N/A
Software and Algorithms		
Prism	GraphPad	https://www.graphpad.com/scientific-software/prism/
ImageJ	NIH	https://imagej.nih.gov/ij/download.html
ZEN lite	Carl Zeiss	https://www.zeiss.com/microscopy/us/products/microscope-software/zen-lite.html
Fiji	NIH	https://imagej.net/Fiji
Phenix	Adams et al., 2010	https://www.phenix-online.org/
XDS	Kabsch, 2010	http://xds.mpimf-heidelberg.mpg.de/
Phaser	McCoy et al., 2007	http://www.ccp4.ac.uk/html/phaser.html
Coot	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Other		
Streptavidin-agarose beads	Thermo Fisher	Cat# 20349
ProLong Gold DAPI Antifade Mountant	Thermo Fisher	Cat# P36931
Polybrene	Millipore Sigma	Cat# TR-1003-G
Puromycin	Invivogen	Cat# ant-pr-1
4% Affymetrix/Alfa Aesar	Affymetrix/Alfa Aesar	Cat# J19943K2

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding authors Lewis C. Cantley (LCantley@med.cornell.edu) and Nathanael S. Gray (nathanael_gray@dfci.harvard.edu). All unique/stable reagents generated in this study are available from the corresponding authors with a completed Materials Transfer Agreement.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Cell Lines

Sources of Cell Lines

All cell lines were sourced from ATCC or DSMZ unless otherwise noted. K562 and KU812F cell lines were obtained from James D. Griffin (Dana-Farber Cancer Institute), and all other leukemia cell lines were obtained from James E. Bradner (formerly Dana-Farber Cancer Institute). HeLa-GFP cell lines were generated using lentiviral or retroviral transduction (see method details). The genders of all cell lines can be found in the Key Resources Table with additional information available on ATCC or DSMZ.

Culture and Maintenance of Cell Lines

All cell culture was performed using standard techniques. HEK293T and HeLa cells were cultured in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% FBS and 1% penicillin/streptomycin (Gibco). Leukemia cells were cultured according to ATCC or DSMZ recommendations. All cells were cultured at 37°C and 5% CO_2.

METHOD DETAILS

Chemistry

All solvents and reagents were used as obtained. ^{1}H-NMR spectra were recorded with a Varian Inova 600 NMR spectrometer and referenced to DMSO. Chemical shifts are expressed in ppm. Mass spectra were measured with Waters Micromass ZQ using an ESI source coupled to a Waters 2525 HPLC system operating in reverse mode with a Waters Sunfire C18 5 μm, 4.6 x 50 mm column. Purification of compounds was performed with either a Teledyne ISCO CombiFlash Rf system or a Waters Micromass ZQ preparative system. The purity was analyzed on a Waters LC-MS Symmetry (C18 column, 4.6 x 50 mm, 5 μM) using a gradient of 5%–95% methanol in water containing 0.05% trifluoroacetic acid (TFA). Detailed synthetic schemes and characterization data below (see Scheme S1).

ADP-Glo Kinase Assays

ADP-Glo assay protocol was modified from Davis et al. (2013). DPPS and PI5P (Echelon Biosciences) were dissolved in DMSO (333 μL/mg) and mixed by sonication and vortexing in a ratio of 2:1 DPPS:P15P. First, 63 μL of DMSO was added to 1255 μL of buffer 1 (30 mM Hepes pH 7.4, 1 mM EGTA, 0.1% CHAPS) and 2868 μL of buffer 2 (46 mM Hepes pH 7.4, 0.1% CHAPS). Volumes were multiplied according to number of assays to be run. PI5P4Kα enzyme was added to the buffer mixture at 32 nM concentration (predetermined on a batch-by-batch basis to give maximal signal to background). 10 μL was dispensed into white 384-well plates (Corning #3824). Then, 100 nL of compounds in DMSO were transferred by a pintool (JANUS, PerkinElmer). To initiate the reaction, 5 μL of ATP (Promega) at 15 μM (final assay concentration of 5 μM) and PI5P/DPPS (at concentrations of 0.06 μg/μL and 0.12 μg/μL respectively) in buffer 3 (20 mM Hepes pH 7.4, 60 mM $MgCl_2$, 0.015 mM ATP and 0.1% CHAPS) was added to each well. The final concentration of DMSO in the reaction was less than 5%. The resulting mixture was incubated at room temperature in the dark for one hour, at which time 5 μL of ADP-Glo reagent 1 were added to stop the reaction and remove any remaining ATP. After a 45-minute incubation, 10 μL of the ADP-Glo reagent 2 were added and allowed to incubate for 30 minutes. The luminescence was then read on an EnVision 2104 Multilabel Plate Reader (PerkinElmer). IC_{50}s were determined using the GraphPad Prism nonlinear regression curve fit.

Radiometric Kinase Assays

The PI5P4K assay was carried out as described in Rameh et al. (1997). Briefly, the kinase reaction was carried out in a total of 70 μL of kinase buffer (50 mM HEPES pH 7.4, 10 mM $MgCl_2$) with 0.1 μg of purified PI5P4K protein (preincubated with DMSO or compound at various concentrations), followed by addition of 20 μL of the resuspended lipids (sonicated and vortexed), 10 μM non-radiolabeled ATP, and 10 μCi [γ-^{32}P]-ATP for 10 minutes at room temperature. The reaction was terminated by adding 50 μL of 4 N HCl. Phosphoinositides were extracted by adding 100 μL of methanol/chloroform (approximately 1:1, vol:vol) mix and subjected to TLC (thin-layer chromatography) separation using heat-activated 2% oxaloacetate-coated silica gel 60 plates (20 cmx20 cm, EMD Millipore) and a 1-propanol/2 M acetic acid (65:35, vol:vol) solvent system. The radiolabeled product, $PI(4,5)P_2$, was quantified with a Phosphorimager (Molecular Dynamics, STORM840, GE Healthcare).

Immunoblotting

Cells were lysed with Pierce IP Lysis buffer supplemented with a cOmplete™ Mini Protease Inhibitor Cocktail tablet (Roche) on ice for 60 min. Lysates were clarified by centrifugation at 20,000 × *g* for 15 min at 4°C. Protein concentration
was determined by a BCA assay (Pierce), and all samples were run with equal
total protein content. Imaging was performed by detection of fluorescently labeled infrared secondary antibodies (IRDye) on the Odyssey CLx Imager (LI-COR).

Pulldown Assays

HEK293T cells were lysed with Pierce IP Lysis buffer supplemented with a cOmplete™ Mini Protease Inhibitor Cocktail tablet (Roche) on ice for 60 min. Lysates were clarified by centrifugation at 20,000 × *g* for 15 min at 4°C. Protein concentration
was determined by a BCA assay (Pierce), and all samples were equalized for protein content. DMSO or 1 μM dtb-THZ-P1-2 was added and rotated overnight at 4°C. Samples were rotated for 2h at room temperature to enhance covalent binding and then incubated with streptavidin resin for 2h at 4°C. Beads were washed with lysis buffer five times, resuspended in 1x LDS sample buffer, boiled at 95°C for 10 min and subjected to immunoblotting. For competitive pulldowns, HEK293T cells were grown to 90-95% confluence, pretreated for 6h with DMSO or varying concentrations of THZ-P1-2, washed twice with cold PBS, lysed, and streptavidin pulldown conducted as described above with normalized samples.

Mass Spectrometry

For intact MS analysis, recombinant PI5P4K protein was incubated with DMSO or 5 μM inhibitor (JNK-IN-7, THZ-P1-2 or THZ-P1-2-R) for 2 h at 37°C and analyzed by LC-ESI-MS essentially as described in Zhang et al. (2012). In each analysis, 5 μg protein was injected onto a self-packed reversed phase column (1/32" O.D. × 500 um I.D., 5 cm of POROS 10R2 resin). After desalting for four minutes, protein was eluted with an HPLC gradient (0–100% B in 4 min, A = 0. 2M acetic acid in water, B = 0.2 M acetic acid in acetonitrile, flow rate = 10 μL/min) into an LTQ ion trap mass spectrometer (ThermoFisher). Mass spectra were deconvoluted using MagTran1.03b2 software. To determine the site of modification, the samples used for intact analysis were reduced with TCEP (10 mM final concentration), alkylated with iodoacetamide (22.5 mM final concentration) and digested with Asp-N or Glu-C (37°C, overnight) and analyzed by nanoLC-MS. Digested peptides were injected onto the precolumn (4 cm POROS 10R2, Applied Biosystems) and eluted with an HPLC gradient (NanoAcquity UPLC system, Waters, Milford, MA; 10–70% B in 60 min; A = 0.1 M acetic acid in water, B = 0.1M acetic acid in acetonitrile). Peptides were resolved on a self-packed analytical column (50 cm Monitor C18, Column Engineering, Ontario, CA) and introduced to the mass spectrometer (LTQ Orbitrap XL) at a flow rate of ~30 nL/min (ESI spray voltage = 3.2 kV). The mass spectrometer was programmed to perform data-dependent MS/MS on the five most abundant precursors (35% collision energy) in each MS1 scan (image current detection, 30K resolution, *m/z* 300–2000). MS/MS spectra were matched to peptide sequences using Mascot (version 2.2.1) after conversion of raw data to .mgf using multiplierz scripts. Precursor and peptide ion mass tolerances were 10 ppm and 0.6 Da, respectively. Supplemental Intact MS Spectra for representative raw MS spectra used for charge deconvolution included separately as Data S1.

Protein Expression and Purification

Constructs of human PI5P4K in pNIC28-Bsa4/pGEX-2T vectors were overexpressed in E. coli BL21 (DE3) in TB medium in the presence of 50 μg/ml of kanamycin for pNIC28-Bsa4 or 100 μg/ml ampicillin for pGEX-2T. Cells were grown at 37°C to an OD of 0.7, induced overnight at 17°C with 400 μM isopropyl-1-thio-D-galactopyranoside, collected by centrifugation, and stored at -80°C. Cell pellets were resuspended in buffer A (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 10% glycerol, 20 mM Imidazole, and 14 mM BME), lysed by sonication, and the resulting lysate was centrifuged at 16,000 xg for 30 min. ~5mL Ni-NTA beads (Qiagen) were mixed with lysate supernatant for 45 min, washed with buffer A, and eluted with buffer B (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 10% glycerol, 300 mM Imidazole, and 14 mM BME). The eluted sample was gel-filtered through a Superdex-200 16/600 column in buffer C (20 mM HEPES, pH 7.5, 500 mM NaCl, 10% glycerol, 10mM DTT, and 1mM TCEP). Protein fractions were pooled, concentrated, and stored at -80°C. To purify GST-tag proteins, cell pellets were resuspended in buffer G (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 10% glycerol, 1 mM EDTA, and 14 mM BME), lysed, centrifuged, and purified as described above using ~5mL GSH beads (GE healthcare) and glutathione elution buffer (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 10% glycerol, 15 mM Glutathione, and 14 mM BME).

Crystallization

A sample of 400 μM protein and 500 μM THZ-P1-2 was co-crystallized in 20% PEG3350 and 200 mM NaMalate by sitting-drop vapor diffusion at 20°C. Crystals were transferred briefly into crystallization buffer containing 25% glycerol prior to flash-freezing in liquid nitrogen.

Data Collection and Structure Determination

Diffraction data from complex crystals were collected at beamline 24ID-E of the NE-CAT at the Advanced Photon Source at the Argonne National Laboratory. The following beamline settings were used: 12662.0 eV (0.9792 Å) beam energy, ADSC QUANTUM 315 detector, 0.2° oscillation, and 0.2s exposure. Multiple crystals were examined, the best of which diffracted to approximately 2.2 Å. XDS was used for integration and scaling (Kabsch, 2010). The structure was solved by molecular replacement using the program Phaser (McCoy et al., 2007) and the search model PDB entry 2YBX, which generated solutions having TFZ scores > 50. Refinement strategy included iterative model building, ligand fitting, and refinement using Phenix (Adams et al., 2010) and Coot (Emsley and Cowtan, 2004). Initial maps had well-defined density (Fo-Fc maps) for the core of the compound, but not for the linker nor electrophile. This was the case for multiple datasets, collected from crystals derived from different crystallization conditions, including primary precipitants Citrate at pH 8 and PEG50000 at pH 6. Despite repeated attempts at including the linker/electrophile segment, with various exit vectors, maps did not improve and did not suggest significant order or occupancy. Also, two residues upstream and 30 downstream of the cysteine in question, Cys293, were disordered. Final refinement led to a model with excellent statistics, shown in Table S1.

Proliferation Assays

Cells were plated in 96-well or 384-well format and treated with indicated compounds at varying concentrations for 72h. Anti-proliferative effects of compounds were assessed using the Cell Titer Glo assay kit (Promega) with luminescence measured on an EnVision 2104 Multilabel Plate Reader (PerkinElmer). IC_{50}s were determined using the GraphPad Prism nonlinear regression curve fit.

qRT-PCR

Total RNA was prepared using RNeasy (Qiagen). cDNA was synthesized using Superscript Vilo (Thermo) and qRT-PCR performed utilizing Fast SYBR green (Thermo) and the Realplex Mastercycler (Eppendorf). For a list of primers used see Lundquist et al. (2018). Isolation of mRNA and qPCR was performed as follows. 200,000 cells were plated in 6-well plastic dishes. 24 hours later, the RNA in the lysates was extracted using the RNeasy protocol. The RNA was resuspended in 50 μl H_2O at a concentration of 1 μg/μL. cDNA was transcribed using the SuperScript Vilo. The sequences of the oligonucleotides used as primers in the PCR reactions are given in Lundquist et al. (2018). The genes that were quantified here were previously shown to be regulated by TFEB (Perera et al., 2015).

Fluorescence Microscopy

HeLa were grown on glass coverslips pre-treated with poly-d-lysine. When indicated cells were treated with 250-1000 nM THZ-P1-2 for 16 hours. Adherent cell lines were rinsed with phosphate-buffered saline, pH 7.4 (PBS) and fixed with 4% paraformaldehyde (PFA) in PBS for 15 minutes at room temperature. After fixation, the cells were permeabilized for 10 minutes with PBS/0.1% Triton X-100, blocked for 30 minutes in blocking buffer (PBS with 3% BSA) and labeled with primary antibodies in blocking buffer for 1 hour at room temperature or overnight at 4°C. Alternatively, for staining, LC3B and LAMP1 (ab25245; Abcam) cells were fixed/permeabilized in -20 MeOH for 20 minutes. Coverslips were washed three times with blocking buffer and incubated with Alexa Fluor-conjugated goat secondary antibodies in blocking buffer for 1 hour at room temperature. After incubation with secondary antibodies, coverslips were washed three times with PBS, once with water, and then mounted on a glass microscope slide with Prolong Gold with DAPI (Thermo). The following primary antibodies were used: TFEB (SAB4503154; Sigma), LC3B (3868; Cell Signaling), LAMP1 (ab25245; Abcam). Alexa Fluor-conjugated secondary antibodies (Thermo) were used at 1:1000. Fluorescent and phase contrast images were acquired on a Nikon Eclipse Ti microscope equipped with an Andor Zyla sCMOS camera. Within each experiment,

exposure times were kept constant and in the linear range throughout. When using the 60x and 100x oil immersion objectives, stacks of images were taken and deconvoluted using AutoQuant (Media Cybernetics).

Autophagy Flux Assays

HeLa cells were plated in 24-well coverslip-bottomed dishes and transduced with 40 particles per cell of the Premo Autophagy Tandem Sensor kit (Thermo Fisher) and cultured for 16-24 hours. Cells were then incubated in either DMSO, 50 μM chloroquine or 1 μM THZ-P1-2 for an additional 16 hours. Cells were mounted in an Okolab Stage Top Incubator warmed to 37°C. DMEM without phenol red (Gibco) was used during image acquisition. All images were collected with a Yokagawa CSU-X1 spinning disk confocal with Spectral Applied Research Aurora Borealis modification on a Nikon Ti motorized inverted microscope equipped with Perfect Focus System. Images were acquired with a Hamamatsu ORCA-ER2 cooled-CCD camera controlled with MetaMorph image acquisition software. Within each experiment, exposure times were kept constant and in the linear range throughout. Images were analyzed and puncta quantified using Fiji software and values plotted using GraphPad Prism.

Generation and Transduction of GFP *PIP4K2* Lentiviral Constructs

Replication-deficient lentiviruses were prepared using a third-generation lentiviral system. GFP was inserted by itself or upstream of either human *PIP4K2A* or *PIP4K2B* in a lentiviral vector. Cysteine to Serine point mutations were made using the Quikchange XL II site-directed mutagenesis kit (Agilent). Virus was generated by cotransfecting three helper plasmids (pLP1, pLP2, pVSV-G) and the vector containing the gene of interest with *cis*-acting sequences for proper packaging were used to generate pseudovirions. A subconfluent culture of HEK293T cells was transfected using the CalPhos Mammalian Transfection Kit (Clontech). The titer of each virus was determined using HEK293T cells. Viral supernatant was then used to express PI5P4Ks in HeLa cells.

Virus Production and Infection

The 293T packaging cell line was used for retroviral amplification. Cells were transfected at 80-90% confluence with packaging plasmid and viral vectors. The packaging plasmid used for retroviral infection was pCL-Eco (Naviaux et al., 1996). Media was refreshed after 6 hours and the virus-containing media was collected 48 hours later, filtered, and subsequently used for infecting cells in the presence of 8μg/ml polybrene (Sigma). Media was once again refreshed 24 hours post-infection, and cells were subjected to puromycin selection or cell sorting for GFP 48 hours post-infection.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis, methods, and details for experiments can be found in the legend of each figure. Figures 1B, 1D, 4, 5C, 5E, 5F, S1C, S3A, S4C, and S5B were generated using GraphPad Prism 7.0.

DATA AND CODE AVAILABILITY

The PI5P4Kα/THZ-P1-2 co-crystal structure generated in this study has been deposited to the PDB under code 6OSP.

Cell Chemical Biology, Volume 27

Supplemental Information

Targeting the PI5P4K Lipid Kinase Family in Cancer Using Covalent Inhibitors

Sindhu Carmen Sivakumaren, Hyeseok Shim, Tinghu Zhang, Fleur M. Ferguson, Mark R. Lundquist, Christopher M. Browne, Hyuk-Soo Seo, Marcia N. Paddock, Theresa D. Manz, Baishan Jiang, Ming-Feng Hao, Pranav Krishnan, Diana G. Wang, T. Jonathan Yang, Nicholas P. Kwiatkowski, Scott B. Ficarro, James M. Cunningham, Jarrod A. Marto, Sirano Dhe-Paganon, Lewis C. Cantley, and Nathanael S. Gray

Supplemental Information

A

Kinase	Labeling site	% Inhibition (%)
JNK1, JNK2, JNK3	Lys2	99.7
p38a	Lys2	16.4
p38a	Other	14.7
GSK3A	Lys2	20.4
GSK3B	Lys2	12.3
PFTAIRE1	Lys1	75.3
CDK9	Lys2	16.6
ROCK1	Other	-11.9
ROCK1, ROCK2	Lys2	-21.5
PIP5K3	ATP	91.8
PIK3C3	ATP	77.5
PIK3C3	ATP	74.2
PIP4K2C	ATP	77.2
PIP4K2C	ATP	84.3

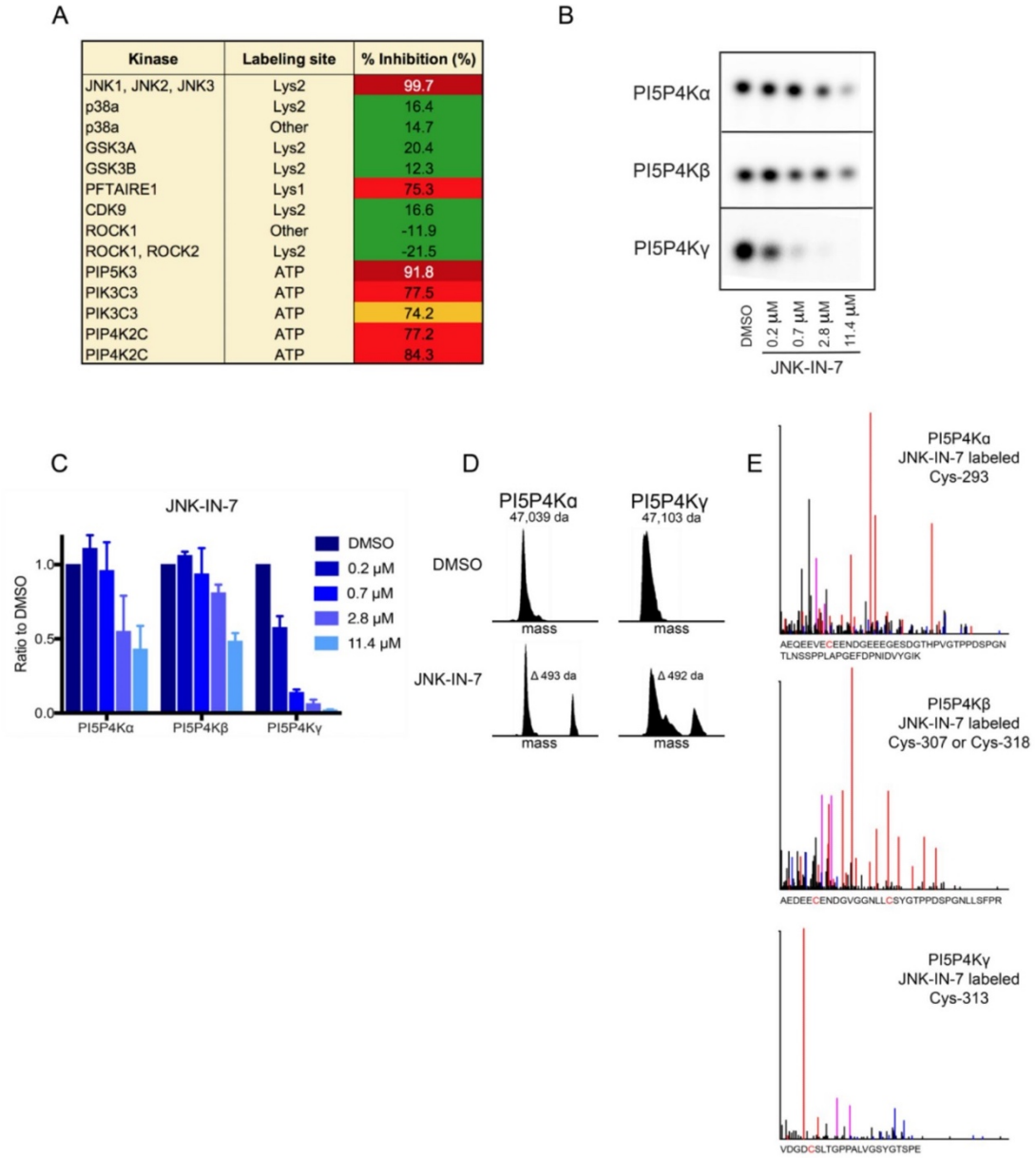


Figure S1, Related to Fig. 1A. (A) list of selected targets, including PI5P4Kγ (PIP4K2C), from KiNativ profiling of JNK-IN-7 at 1 μM for 3 hours in A375 cells. Full list is available in Zhang et al. (2012). (B) JNK-IN-7 inhibits kinase activity of all three PI5P4K isoforms in a radiometric TLC assay measuring radiolabeled PI-4,5-P_2. JNK-IN-7 was incubated with purified protein for 30min, followed by a 10min kinase reaction, quenching, lipid extraction and TLC development. (C) Quantification of (B). (D) Zero charge masses of recombinant PI5P4Kα and γ incubated with JNK-IN-7 for 2 hours at 37°C demonstrates covalent labeling of PI5P4K isoforms as determined by intact mass spectrometry. See Supplemental Intact MS Spectra for representative raw MS spectra used for charge deconvolution. (E) Subsequent protease digestion and tandem mass spectrometry confirms that THZ-P1-2 covalently labels cysteine residues on all three PI5P4K isoforms. The peptide for PI5P4Kβ was exclusively observed to be singly labeled at either Cys-307 or Cys-318. See Supplemental Intact MS Spectra for representative raw MS spectra used for charge deconvolution.

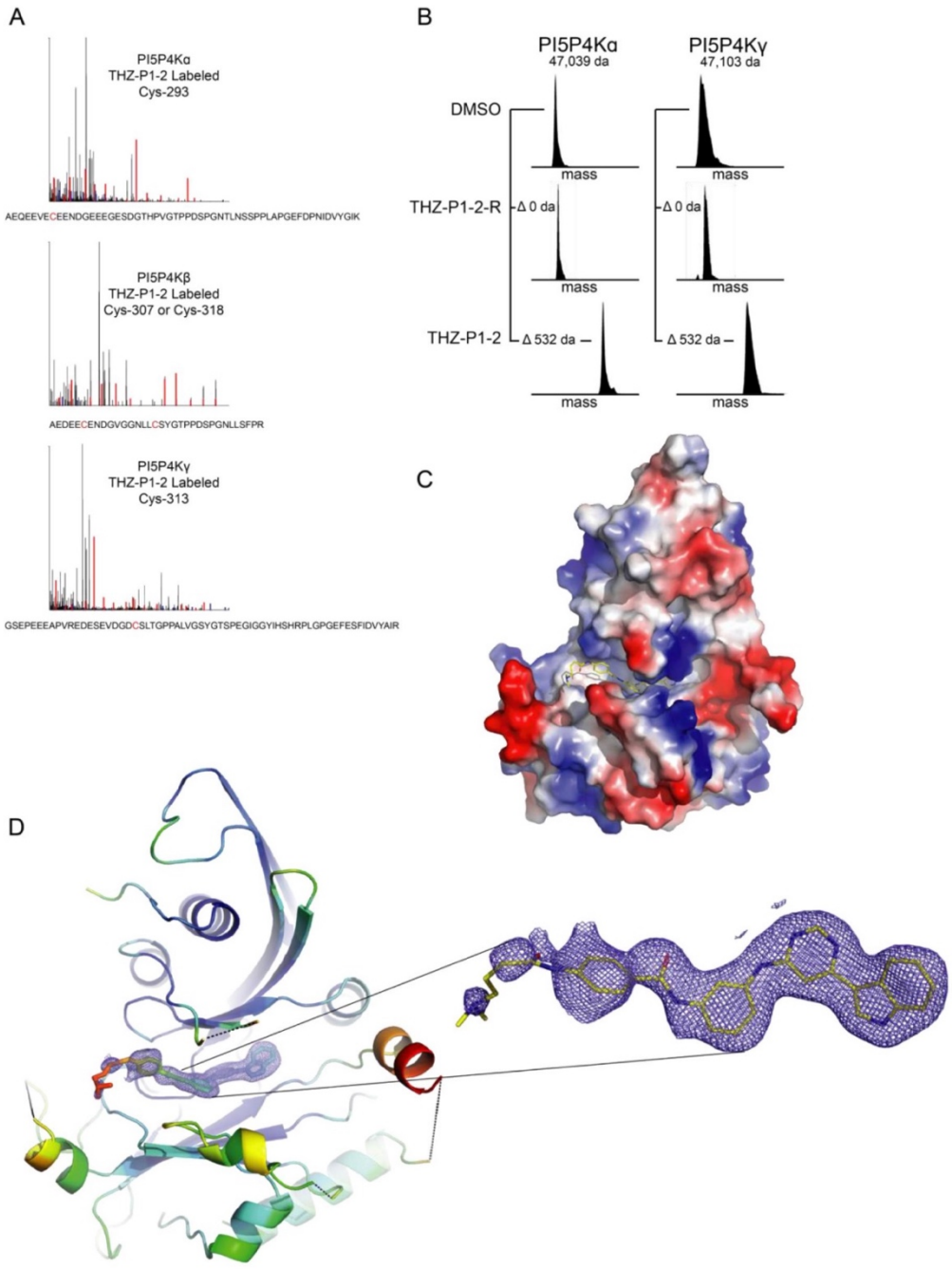


Figure S2, Related to Fig. 2. (A) Protease digestion and tandem mass spectrometry following intact mass labeling confirms that THZ-P1-2 covalently labels cysteine residues on all three PI5P4K isoforms. The peptide for PI5P4Kβ was exclusively observed to be singly labeled at either Cys-307 or Cys-318. Related to Fig. 2B. See Supplemental Intact MS Spectra for representative raw MS spectra used for charge deconvolution. (B) THZ-P1-2-R was found not to covalently label recombinant PI5P4Kα and γ when incubated with purified protein for 2 hours at 37°C. Mass labeling plot shown here is compared to mass shift observed with THZ-P1-2 from Fig. 2A. (C) Surface representation of co-crystal structure of PI5P4Kα in complex with THZ-P1-2. Related to Fig. 2C. (D) Zoom-in of ligand electron density within the active site of PI5P4Kα. See Table S1 for diffraction data collection and refinement statistics.

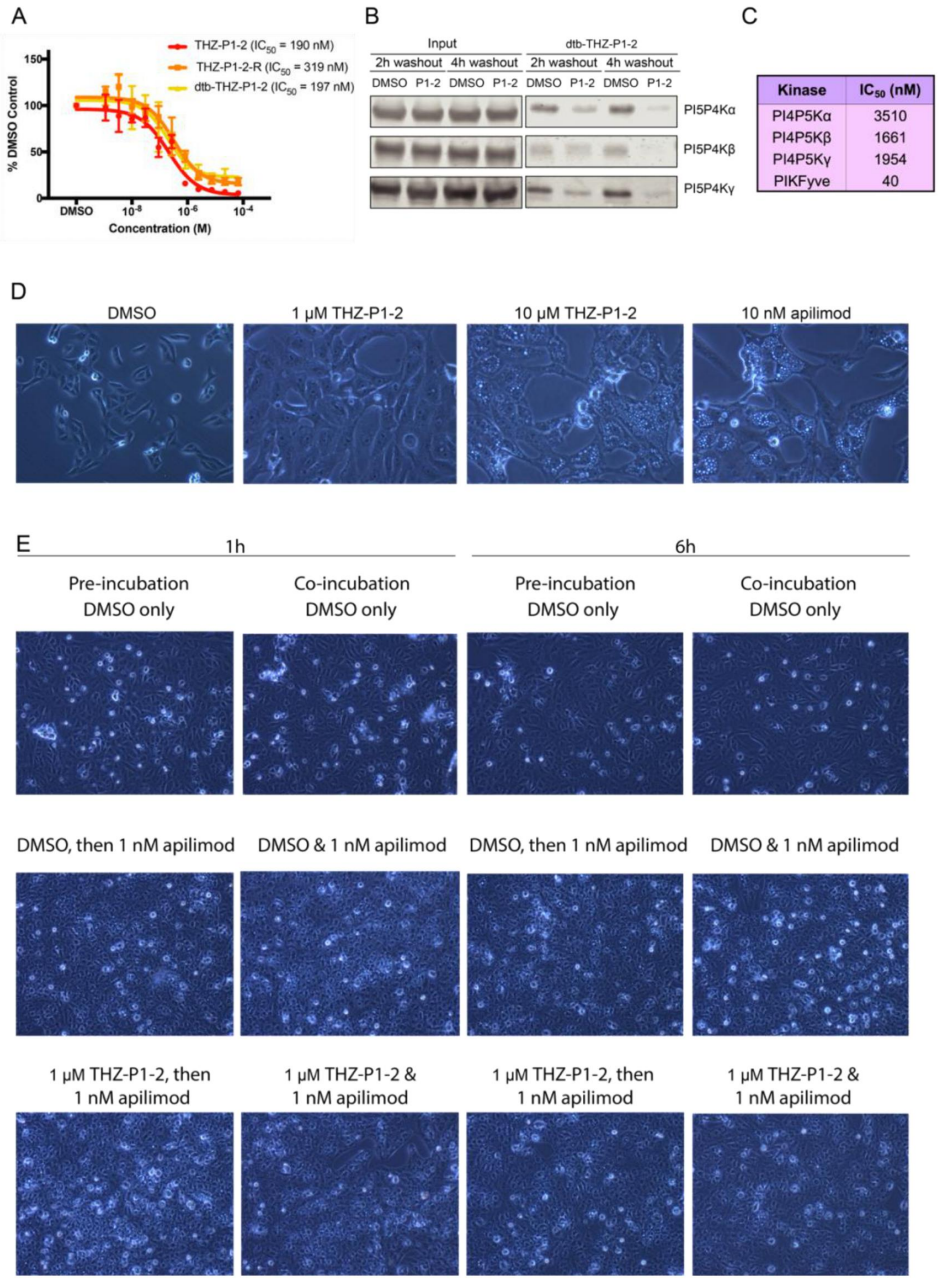


Figure S3, Related to Fig. 3. (A) THZ-P1-2-R and dtb-THZ-P1-2 were found to inhibit PI5P4Kα in an ADP-Glo assay, with a slight loss in potency with THZ-P1-2-R. Both compounds were incubated with PI5P4Kα for 15min before proceeding with a 1 hour reaction with ATP and development with ADP-Glo reagents. Data are represented as mean ± SEM. Plots are shown in comparison with THZ-P1-2 from Fig. 1B. (B) THZ-P1-2 engages PI5P4K isoforms at 2h and 4h timepoints, exhibiting prolonged engagement when a washout is performed. HEK293T cells were treated with DMSO or THZ-P1-2 at time, t=0 and cells were washed 2X with PBS at either 2 or 4h and harvested at the end of 6h. A streptavidin pulldown was then conducted in lysates after normalization of protein content with a BCA assay. (C) THZ-P1-2 inhibits the Type 1 PI4P5K kinases at a lower extent but shows potent inhibition on PIKfyve by ADP-Glo. *In vitro* kinase assays were performed by Invitrogen and Carna Biosciences. (D) THZ-P1-2 causes vacuolar enlargement, characteristic of PIKfyve inhibition, at a concentration of 10 μM but not 1 μM, compared to PIKfyve inhibitor apilimod at 10 nM treatment. Vero cells were treated with DMSO or compound for 6h and imaged. (E) Incubation of Vero cells with 1 μM THZ-P1-2 was unable to impair the PIKfyve inhibitory phenotype of apilimod. Preincubation: Vero cells were incubated with DMSO/THZ-P1-2 for 1 or 6h, washed with PBS to completely remove compound, treated with DMSO/apilimod and imaged after 6h. Coincubation: Vero cells were incubated with DMSO/THZ-P1-2 for 1 or 6h, and then co-treated with DMSO/apilimod and imaged after 6h.

A

Kinase	IC_{50} (nM)
ABL1	21.3
CSNK2A2	1340
KIT	>10000
PDGFRB	3050
RPS6KA4	>10000
TYK2	1600
BRK	313

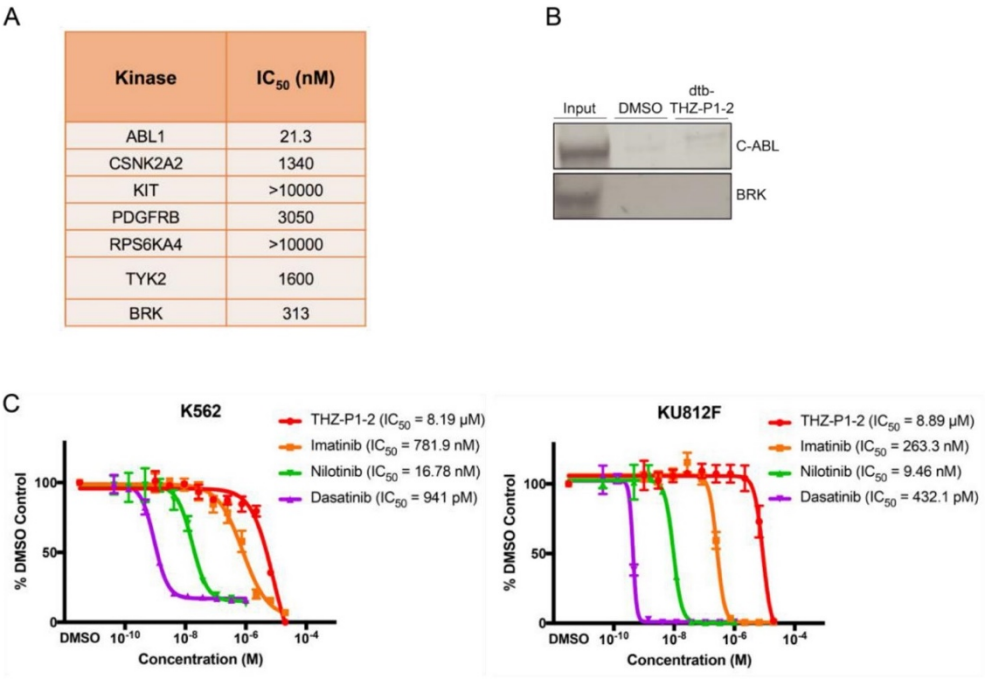


Figure S4, Related to Fig. 3 and 4. (A) THZ-P1-2 inhibits off-target kinases identified in the DiscoverX KINOMEScan panel to varying degrees. *In vitro* kinase assays were performed by Invitrogen using the Z'LYTE assay. (B) A streptavidin pulldown in HEK293T lysate shows little to no engagement of C-ABL and BRK by dtb-THZ-P1-2. (C) THZ-P1-2 shows mild antiproliferative activity on BCR-ABL addicted cell lines. Two cell lines containing BCR-ABL translocations, K562 and KU812F, were treated with THZ-P1-2 or known BCR-ABL inhibitors imatinib, nilotinib and dasatinib for 72h and assayed using Cell-Titer Glo. Data are represented as mean ± SEM. (D) THZ-P1-2 does not bind to and engage off-targets such as JNK, IRAK1, PKN3, CDK7 and CDK12 despite originating from the same core scaffold as inhibitors of these targets. Competitive streptavidin pulldowns in HEK293T cells were conducted with DMSO, 1 μM THZ-P1-2 and 1 μM of each target's corresponding inhibitor, followed by pulldown in lysate with 1 μM of the corresponding biotinylated or desthiobiotinylated inhibitor.

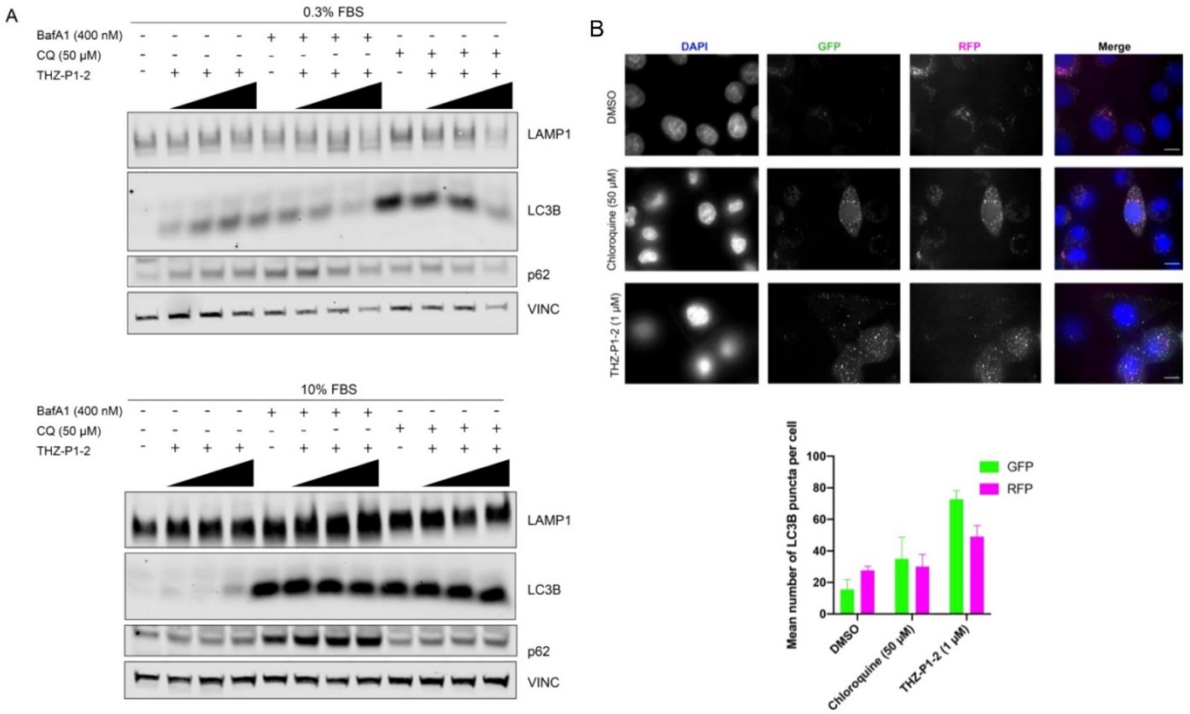
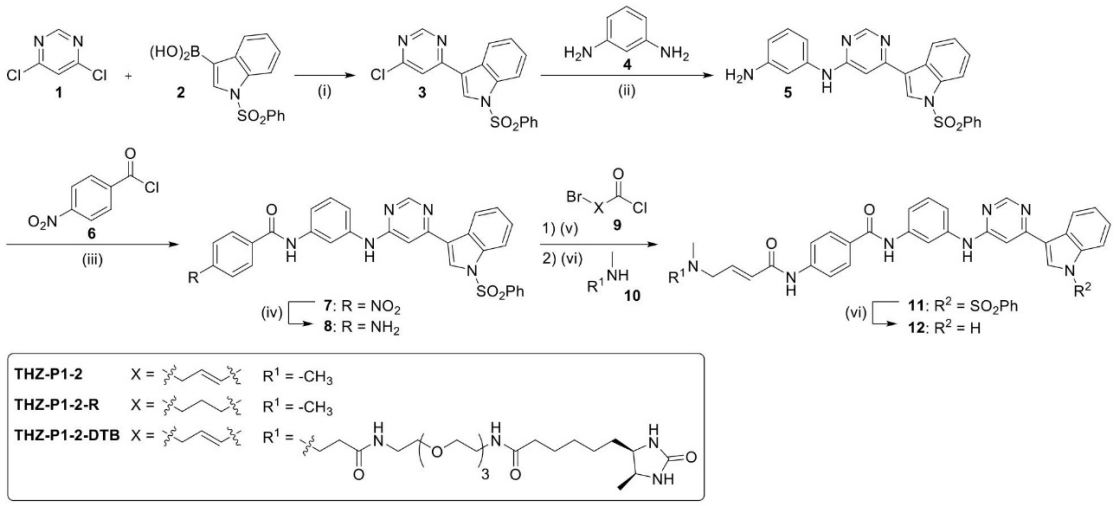


Figure S5, Related to Fig. 5. (A) THZ-P1-2 causes a slight increase in LAMP1, LC3B and p62 protein levels, as observed with positive control compounds bafilomycin A1 and chloroquine, and this effect is not cumulative in combination with these known autophagy inhibitors. HeLa cells were cultured for 24h in DMEM media supplemented with either 0.3% (serum-starved conditions) or 10% FBS. Cells were treated with DMSO, 250 nM, 500 nM, and 1000 nM concentrations of THZ-P1-2 and single doses of bafilomycin A1 and chloroquine, harvested after 24h, and analyzed by Western blot. Related to Fig. 5A. (B) THZ-P1-2 causes a dramatic increase in GFP signal and moderate increase in RFP signal in HeLa cells, as observed with positive control chloroquine, suggesting that autophagy was upregulated (leading to increase in LC3B fluorescence signal) but not completed with autolysosome fusion as GFP signal was not quenched. Cells were transfected with the Premo Autophagy Tandem Sensor for 16h and treated with DMSO or compounds for another 16h before analyzed using confocal microscopy. Fluorescence signal was quantified for at least n=10 per condition and data are represented as mean of two independent experiments ± SEM,

Scheme 1. Synthetic route of THZ-P1-2/THZ-1-2-R/dtb-THZ-P1-2[a], Related to Fig. 1A.



[a] Reagents and conditions: (i) $NaHCO_3$, $Pd(PPh_3)_2Cl_2$, ACN/H_2O (4:1, v/v), 90 °C, overnight, 61 % (ii) DIEA, NMP, 150 °C, overnight, 95 % (iii) pyridine, 80 °C, 3 h (iv) $SnCl_2 \cdot 2H_2O$, EtOAc/MeOH (5:2, v/v), 80 °C, overnight, 69 % over 2 steps (v) 1) DIEA, ACN, 0 °C, 5 min. 2) THF, rt, 2 h (vi) 1 m NaOH/1,4-dioxane (1:1, v/v), rt, 6 h, 7-73 % over 3 steps.

THZ-P1-2: MS m/z 532.24 $[M+H]^+$. ^1H NMR (400 MHz, DMSO-d_6) δ 11.73 (s, 1H), 10.50 (s, 1H), 10.19 (s, 1H), 9.55 (s, 1H), 8.63 (s, 1H), 8.27 (d, *J* = 7.2 Hz, 1H), 8.20 (s, 1H), 8.15 (d, *J* = 2.8 Hz, 1H), 8.00 (d, *J* = 8.7 Hz, 2H), 7.83 (d, *J* = 8.8 Hz, 2H), 7.59 – 7.44 (m, 2H), 7.37 (d, *J* = 8.3 Hz, 1H), 7.34 – 7.26 (m, 2H), 7.24 – 7.09 (m, 2H), 6.81 (dt, *J* = 15.4, 6.2 Hz, 1H), 6.40 (d, *J* = 15.4 Hz, 1H), 3.37 (d, *J* = 5.3 Hz, 2H), 2.39 (s, 6H).

THZ-P1-2-R: MS m/z 534.64 $[M+H]^+$. ^1H NMR (500 MHz, DMSO-d_6) δ 12.06 (s, 1H), 10.32 (s, 1H), 10.23 (s, 1H), 9.49 (s, 1H), 8.76 (s, 1H), 8.27 (d, *J* = 3.0 Hz, 1H), 8.24 (d, *J* = 2.1 Hz, 1H), 8.12 (d, *J* = 7.7 Hz, 1H), 7.97 (d, *J* = 8.6 Hz, 2H), 7.75 (d, *J* = 8.8 Hz, 2H), 7.56 (d, *J* = 7.7 Hz, 1H), 7.52 (d, *J* = 8.1 Hz, 1H), 7.43 (d, *J* = 8.2 Hz, 1H), 7.41 – 7.34 (m, 2H), 7.26 (p, *J* = 7.0 Hz, 2H), 3.16 – 3.04 (m, 2H), 2.81 (d, *J* = 4.6 Hz, 6H), 2.47 (d, *J* = 7.1 Hz, 2H), 2.02 – 1.90 (m, 2H).

dtb-THZ-P1-2 MS m/z 961.13 $[M+H]^+$. ^1H NMR (500 MHz, DMSO-d_6) δ 12.08 (s, 1H), 10.69 (s, 1H), 10.28 (s, 2H), 10.06 (s, 1H), 8.75 (s, 1H), 8.55 (s, 2H), 8.38 – 8.19 (m, 3H), 8.15 (d, *J* = 7.6 Hz, 1H), 8.00 (d, *J* = 8.8 Hz, 2H), 7.83 (d, *J* = 8.6 Hz, 2H), 7.54 (t, *J* = 8.6 Hz, 2H), 7.44 (d, *J* = 8.2 Hz, 1H), 7.40 – 7.32 (m, 2H), 7.32 – 7.15 (m, 2H), 6.83 (dt, *J* = 14.7, 7.2 Hz, 1H), 6.53 (d, *J* = 15.3 Hz, 1H), 3.69 – 3.56 (m, 4H), 3.52-3.49 (m, 8H), 3.43 (t, *J* = 6.0 Hz, 2H), 3.38 (t, *J* = 6.1 Hz, 2H), 3.25 (d, *J* = 7.0 Hz, 2H), 3.21 – 3.10 (m, 6H), 2.78 (s, 3H), 2.64 (t, *J* = 7.0 Hz, 2H), 2.05 (t, *J* = 7.5 Hz, 2H), 1.46 (q, *J* = 7.4 Hz, 2H), 1.28-1.22 (m, 4H), 0.95 (d, *J* = 6.4 Hz, 3H).

Table S1, Related to Figure 2C-D. Crystallization conditions, data collection and refinement statistics for co-crystal structure of complex THZ-P1-2 with PI5P4Kα.

Structure Name	**PI5P4Kα/THZ-P1-2**
Ligand	THZ-P1-2
RCSB accession code	6OSP
Data collection [a]	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	44.19 98.68 105.12
a, b, g (°)	90.00 93.66 90.00
Resolution (Å)	44.1-2.21 (2.29-2.21) [b]
R_{pim}	0.042 (0.525)
$I / \sigma I$	13.44 (1.53)
Completeness (%)	99.17 (99.44)
Redundancy	3.7 (3.7)
Structure solution	
PDB entries used for molecular replacement	2YBX
Refinement	
Resolution (Å)	44.1-2.21
No. reflections	44717
R_{work} / R_{free}	0.1779/0.2254
No. atoms	5450
Protein	5102
Ligand/ion	92
Water	256
B-factors	
Protein	53.08
Ligand/ion	70.65
Water	50.74
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.15
Ramachandran	
Preferred	98.21%
Allowed	1.63%
Not Allowed	0.16%

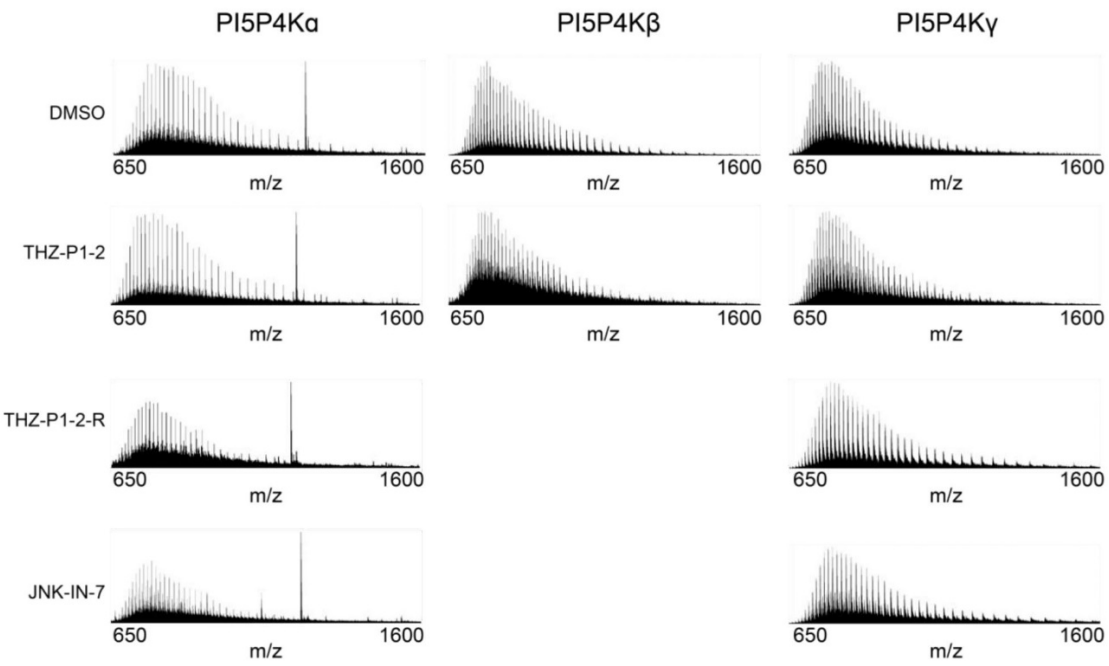
[a] A single crystal was used to collect data for each of the structures reported here.

[b] Values in parentheses are for highest-resolution shell.

Table S2, Related to Fig. 3C. Complete dataset for THZ-P1-2 kinome selectivity using the DiscoverX commercial profiling platform.

The kinome selectivity dataset for THZ-P1-2 is 11 pages and provided as a supplemental excel table.

Data S1. Supplemental Intact MS Spectra, Related to Fig. 2A, S1D, S2B.



Supplemental Excel Table (Kinome Selectivity Data Set):
The presented results were obtained with a 1 μM concentration of THZ-P1-2

Ambit Gene Symbol	Percent Control
RPS6KA4(Kin.Dom.1-N-terminal)	0.1
TYK2(JH2domain-pseudokinase)	0.9
CSNK2A2	1.3
EGFR(E746-A750del)	1.3
ABL1(Y253F)-phosphorylated	2.4
KIT(D816V)	4.2
PDGFRB	4.3
ABL1(H396P)-nonphosphorylated	5.8
PIKFYVE	6.4
ABL1(H396P)-phosphorylated	6.6
BRK	6.6
EGFR(L747-T751del,Sins)	6.8
ABL1-phosphorylated	8.1
EGFR(L747-E749del, A750P)	9
PIP5K2C	9
KIT(V559D)	11
ABL1(M351T)-phosphorylated	12
KIT(L576P)	12
ABL1-nonphosphorylated	13
EGFR(S752-I759del)	14
EGFR(L747-S752del, P753S)	16
EGFR(G719C)	18
EGFR(L858R)	18
ABL1(Q252H)-phosphorylated	20
PIP5K1C	20
ABL1(E255K)-phosphorylated	21
EGFR(L861Q)	21
MKNK2	21
EGFR	22
EGFR(G719S)	22
SRC	23
ABL1(Q252H)-nonphosphorylated	24
KIT	24
CSNK2A1	27
BLK	34
KIT(D816H)	35
ABL2	36
CDK7	37
FLT3(D835V)	37
STK36	38
SIK	39
ABL1(F317L)-phosphorylated	44
FLT3(D835Y)	45
HIPK4	47
RSK4(Kin.Dom.1-N-terminal)	47
SIK2	47
CDK4-cyclinD1	48
CIT	48
TNK2	48
FLT3(D835H)	50
FLT3(ITD,D835V)	51
GSK3A	52
IRAK3	56
KIT(A829P)	57
PDGFRA	58
KIT(V559D,V654A)	59
FLT3(N841I)	60
DRAK2	61
NEK11	61
FLT3(ITD)	62
CSNK1D	63
GRK3	63
LIMK2	63
LIMK1	64
LKB1	64
CSF1R	67
DDR1	67
RET	67
CDK4-cyclinD3	70
ERK4	70

Ambit Gene Symbol	Percent Control
ABL1(F317L)-nonphosphorylated	71
ALK(L1196M)	71
CDK4	71
EPHA2	71
MERTK	71
EPHB6	72
IKK-alpha	72
TRKA	72
EGFR(L858R,T790M)	73
IKK-beta	73
MLK3	73
RET(M918T)	73
ERK5	74
ERK8	74
JAK1(JH1domain-catalytic)	74
JAK1(JH2domain-pseudokinase)	74
DRAK1	75
EPHB1	75
GCN2(Kin.Dom.2,S808G)	75
PRKD2	75
MYLK4	76
PKN2	76
ADCK3	77
EPHA1	78
MAP3K1	78
PAK6	78
TAK1	78
LCK	79
MAP4K4	79
SRMS	79
SYK	80
TNNI3K	80
TRKC	80
BMX	81
CSNK1G2	81
SRPK1	81
TNK1	81
DMPK	82
EPHB2	82
FGR	82
FLT3(K663Q)	82
INSRR	82
MARK3	82
MST1	82
PFTK1	82
PIP5K2B	82
PRP4	82
RIOK3	82
ARK5	83
AURKC	83
CSF1R-autoinhibited	83
FLT3-autoinhibited	83
MLK1	83
CDK9	84
CSNK1E	84
MUSK	84
PIP5K1A	84
YANK3	84
CSNK1A1L	85
DYRK2	85
IRAK1	85
p38-delta	85
PKAC-beta	85
JNK1	86
KIT-autoinhibited	86
MLK2	86
NEK6	86
RIPK1	86
RSK1(Kin.Dom.1-N-terminal)	86
CDKL2	87
EPHA5	87
ERK2	87
ERK3	87
MKNK1	87
NEK5	87
PIK3CA(H1047Y)	87
RIOK1	87
TIE1	87
YANK1	87
CDK3	88
ERBB4	88
FGFR3	88
FLT4	88

Ambit Gene Symbol	Percent Control
GAK	88
MINK	88
STK16	88
TESK1	88
VEGFR2	88
BIKE	89
BMPR1A	89
CDK5	89
DYRK1B	89
EPHB4	89
FLT1	89
FLT3(ITD,F691L)	89
LYN	89
PAK3	89
SGK3	89
TEC	89
ACVRL1	90
FAK	90
FLT3	90
GRK2	90
JNK2	90
LRRK2(G2019S)	90
MET(M1250T)	90
PIM2	90
RSK2(Kin.Dom.2-C-terminal)	90
SRPK3	90
VPS34	90
VRK2	90
WEE1	90
ACVR2B	91
DAPK2	91
ERBB2	91
FGFR3(G697C)	91
LRRK2	91
MYLK2	91
PAK1	91
PAK2	91
PCTK2	91
PFPK5(P.falciparum)	91
PYK2	91
RIOK2	91
RIPK2	91
ROCK2	91
RSK3(Kin.Dom.1-N-terminal)	91
TRPM6	91
ALK	92
AXL	92
CASK	92
CDK11	92
CSNK1A1	92
DDR2	92
FGFR4	92
FYN	92
JNK3	92
KIT(V559D,T670I)	92
MAP3K4	92
MEK1	92
MEK6	92
PFTAIRE2	92
YES	92
ABL1(F317I)-nonphosphorylated	93
ABL1(T315I)-nonphosphorylated	93
ALK(C1156Y)	93
AMPK-alpha1	93
CAMK1	93
CAMK1D	93
CAMK2A	93
CDC2L2	93
CSK	93
DLK	93
DMPK2	93
HIPK3	93
IGF1R	93
LTK	93
MAPKAPK5	93
MEK5	93
PLK4	93
PRKCH	93
PRKCQ	93
TXK	93
ZAK	93
ACVR1	94
ACVR2A	94

Ambit Gene Symbol	Percent Control
CAMKK2	94
CDC2L1	94
CLK3	94
ERK1	94
LOK	94
MAP4K5	94
PRKD3	94
ROS1	94
RSK3(Kin.Dom.2-C-terminal)	94
TLK1	94
TTK	94
TYRO3	94
AKT1	95
AKT2	95
ASK2	95
BRSK2	95
BUB1	95
CLK4	95
DYRK1A	95
EGFR(T790M)	95
FER	95
FGFR1	95
MARK1	95
MARK4	95
PIK3C2B	95
PIK4CB	95
PIM1	95
QSK	95
RPS6KA4(Kin.Dom.2-C-terminal)	95
TBK1	95
TSSK1B	95
WNK4	95
YSK1	95
BRAF	96
BTK	96
CAMKK1	96
CDKL1	96
CDKL3	96
CLK1	96
CLK2	96
CSNK1G1	96
DCAMKL3	96
FRK	96
GRK4	96
PIM3	96
PRKCI	96
RET(V804M)	96
RPS6KA5(Kin.Dom.1-N-terminal)	96
RSK2(Kin.Dom.1-N-terminal)	96
SNRK	96
ADCK4	97
ANKK1	97
BMPR1B	97
CAMK4	97
FLT3(R834Q)	97
HASPIN	97
JAK3(JH1domain-catalytic)	97
MAST1	97
MST4	97
NDR2	97
NLK	97
PCTK3	97
PRKR	97
TGFBR2	97
CHEK1	98
EIF2AK1	98
FGFR2	98
GSK3B	98
HCK	98
MET	98
MLCK	98
NEK2	98
NEK7	98
NEK9	98
PAK4	98
PAK7	98
SNARK	98
TAOK1	98
TIE2	98
TNIK	98
AURKA	99
BRSK1	99

Ambit Gene Symbol	Percent Control	Ambit Gene Symbol	Percent Control
CSNK1G3	99	EPHA8	100
MAP4K2	99	EPHB3	100
MEK3	99	ERBB3	100
MST1R	99	ERN1	100
MYLK	99	FES	100
PIK3CA(Q546K)	99	GRK1	100
PKMYT1	99	GRK7	100
RET(V804L)	99	HIPK1	100
RIPK4	99	HIPK2	100
RPS6KA5(Kin.Dom.2-C-terminal)	99	HPK1	100
SLK	99	HUNK	100
TLK2	99	ICK	100
AAK1	100	IKK-epsilon	100
ABL1(F317I)-phosphorylated	100	INSR	100
ABL1(T315I)-phosphorylated	100	IRAK4	100
ACVR1B	100	ITK	100
AKT3	100	JAK2(JH1domain-catalytic)	100
AMPK-alpha2	100	LATS1	100
ASK1	100	LATS2	100
AURKB	100	LZK	100
BMPR2	100	MAK	100
BRAF(V600E)	100	MAP3K15	100
CAMK1B	100	MAP3K2	100
CAMK1G	100	MAP3K3	100
CAMK2B	100	MAP4K3	100
CAMK2D	100	MAPKAPK2	100
CAMK2G	100	MARK2	100
CDC2L5	100	MEK2	100
CDK2	100	MEK4	100
CDK8	100	MELK	100
CDKL5	100	MET(Y1235D)	100
CHEK2	100	MKK7	100
CTK	100	MRCKA	100
DAPK1	100	MRCKB	100
DAPK3	100	MST2	100
DCAMKL1	100	MST3	100
DCAMKL2	100	MTOR	100
EPHA3	100	MYO3A	100
EPHA4	100	MYO3B	100
EPHA6	100	NDR1	100
EPHA7	100	NEK1	100
		NEK10	100
		NEK3	100

Ambit Gene Symbol	Percent Control	Ambit Gene Symbol	Percent Control
NEK4	100	PRKG2	100
NIK	100	PRKX	100
NIM1	100	RAF1	100
OSR1	100	RIPK5	100
p38-alpha	100	ROCK1	100
p38-beta	100	RSK1(Kin.Dom.2-C-terminal)	100
p38-gamma	100	RSK4(Kin.Dom.2-C-terminal)	100
PCTK1	100	S6K1	100
PDPK1	100	SBK1	100
PFCDPK1(P.falciparum)	100	SGK	100
PHKG1	100	SgK110	100
PHKG2	100	SGK2	100
PIK3C2G	100	SRPK2	100
PIK3CA	100	STK33	100
PIK3CA(C420R)	100	STK35	100
PIK3CA(E542K)	100	STK39	100
PIK3CA(E545A)	100	TAOK2	100
PIK3CA(E545K)	100	TAOK3	100
PIK3CA(H1047L)	100	TGFBR1	100
PIK3CA(I800L)	100	TRKB	100
PIK3CA(M1043I)	100	TSSK3	100
PIK3CB	100	TYK2(JH1domain-catalytic)	100
PIK3CD	100	ULK1	100
PIK3CG	100	ULK2	100
PKAC-alpha	100	ULK3	100
PKN1	100	WEE2	100
PKNB(M.tuberculosis)	100	WNK1	100
PLK1	100	WNK2	100
PLK2	100	WNK3	100
PLK3	100	YANK2	100
PRKCD	100	YSK4	100
PRKCE	100	ZAP70	100
PRKD1	100		
PRKG1	100		

5.3 Publication B: *Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors*

5.3.1 Specific Aims

With the identification of the covalent pan-PI5P4K inhibitor THZ-P1-2 (see Chapter 5.2), the aim was to further explore its SAR, especially regarding biochemical and cellular target engagement as well as kinome-wide selectivity. For this, a small library of THZ-P1-2 analogs had to be designed, synthesized, and biochemically characterized.

Due to issues in the interpretation of the results from pulldown assays with the THZ-P1-2-DTB probe (see Chapter 5.2), the label-free CETSA method was to be established to investigate THZ-P1-2-analog efficacy *in cellulo*.

5.3.2 Main Outcomes and Context

The presented SAR study was able to confirm key interactions of the THZ-P1-2 scaffold with its identified binding site of PI5P4Ks, especially in regard to the interactions of the compound with the hinge of these lipid kinases. The findings were in accordance with the previously reported co-crystal structure (Chapter 5.2). Furthermore, changes in the indole and pyrimidine moieties of THZ-P1-2 led to the development of compound **30**, which exhibited improved kinome-wide selectivity, while retaining on-target activity in biochemical and cellular assays.

Using isothermal CETSA conditions circumvented read-out issues experienced with pulldown assays of the THZ-P1-2 analogs and enabled determination of cellular binding activity of the covalent PI5P4K inhibitor. This label-free method was further successfully applied to compare cellular PI5P4K binding of selected analogs of THZ-P1-2, leading to the identification of compound **30** as an interesting chemical probe to further study PI5P4K function in the cellular context. The study emphasizes the value of label-free methods in the characterization of a probe, as well as the importance of cellular characterization in addition to biochemical assays for the development of high-quality chemical probes.

Details regarding the synthesis of the THZ-P1-2-analog library as well as the isothermal CETSA conditions are found in the Supporting Information of the article.

5.3.3 Published Article

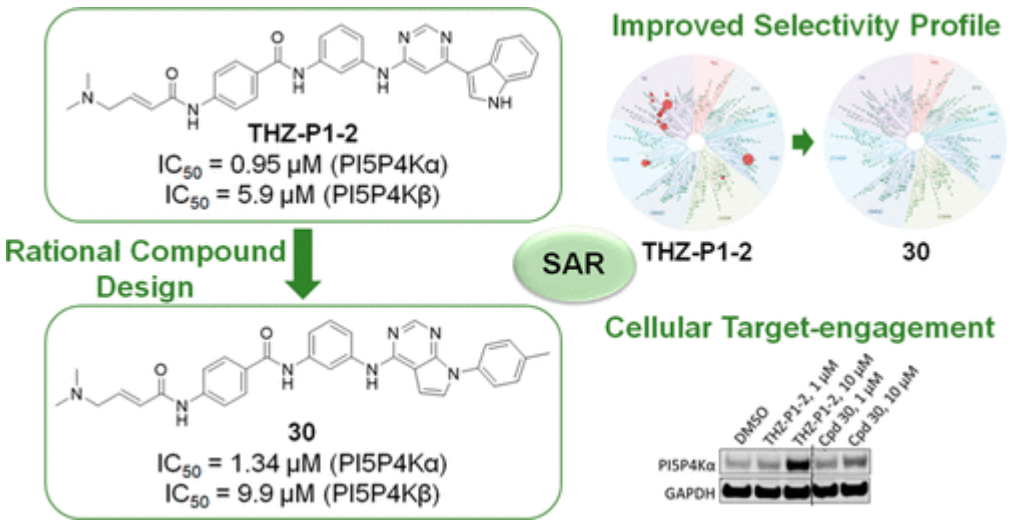
The following peer-reviewed publication was originally published as:

Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz, Sindhu C. Sivakumaren, Adam Yasgar, Matthew D. Hall, Mindy I. Davis, Hyuk-Soo Seo, Joseph D. Card, Scott B. Ficarro, Hyeseok Shim, Jarrod A. Marto, Sirano Dhe-Paganon, Atsuo T. Sasaki, Matthew B. Boxer, Anton Simeonov, Lewis C. Cantley, Min Shen, Tinghu Zhang, Fleur M. Ferguson, and Nathanael S. Gray.

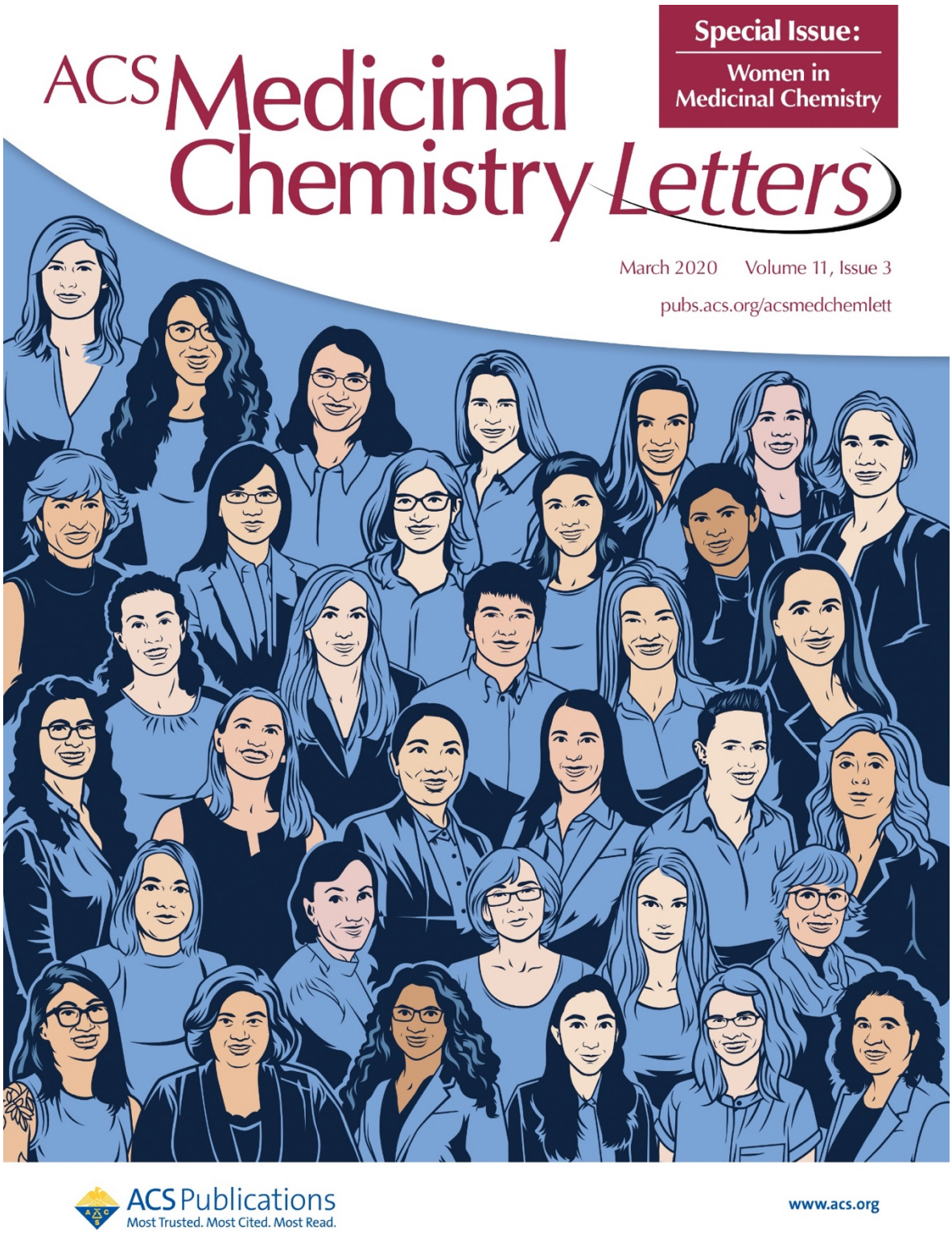
ACS Med. Chem. Lett. **2020**, *11* (3), 346-352.

DOI: https://doi.org/10.1021/acsmedchemlett.9b00402



This article, including Supporting Information, was reproduced with kind permission from *ACS Medicinal Chemistry Letters* © 2019 American Chemical Society.

Note:
This publication is part of the "Women in Medicinal Chemistry" special issue of the Journal *ACS Med. Chem. Lett.* © 2019 American Chemical Society.


Special Issue:
Women in Medicinal Chemistry
ACS Medicinal Chemistry Letters
March 2020 Volume 11, Issue 3
pubs.acs.org/acsmedchemlett
ACS Publications
Most Trusted. Most Cited. Most Read.
www.acs.org

About the Cover:

The cover shows some of the contributors to the Women in Medicinal Chemistry Special Issue (first and/or corresponding authors): Bottom Row (L-R): A.Cheng, M. Chang, N. Rajapaksa, H. Park, C.B. Cooley, M.-J. Blanco; Row 2 (L-R): K. Hiesinger, G. White. A. I. de Lucas. **T. Manz**. A. Newman; Row 3 (L-R): E.H. D'Agostino, C. Wells, X. Lu, K. Nelson, C. Garrison, M. L. Bolognesi; Row 4 (L-R): A. Axtman, W. Young, A.E. Garces, F. Ferguson, J. Farand; Row 5 (L-R): D.M. Huryn, Q. Wang, A. Hart, L. Ferrins, S. Varghese; Row 6 (L-R): I. Landel, L. Terrab, C. Haskell-Luevano, M. Bryan, N. Castilho, M. Grenier, M. Buskes. Image credit: Conception: Kyle Brimacombe, NCATS; Matthew Hall, NCATS & Donna M. Huryn (ACS MCL Assoc. Ed.); Illustration: Jeffrey Everett, NIH Medical Arts.

ACS Medicinal Chemistry Letters
Cite This: ACS Med. Chem. Lett. 2020, 11, 346−352
Letter
pubs.acs.org/acsmedchemlett

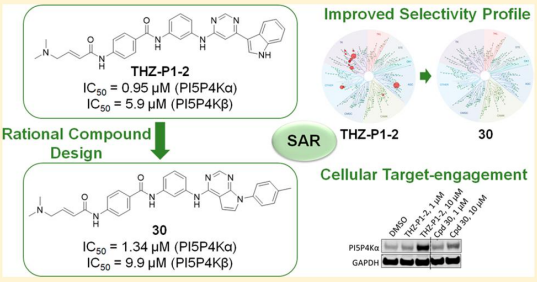
Structure−Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz,[†,‡,§] Sindhu C. Sivakumaren,[†,‡] Adam Yasgar,[∥] Matthew D. Hall,[∥] Mindy I. Davis,[∥] Hyuk-Soo Seo,[†,‡] Joseph D. Card,[#,∇,○] Scott B. Ficarro,[#,∇,○] Hyeseok Shim,[◆] Jarrod A. Marto,[#,∇,○] Sirano Dhe-Paganon,[†,‡] Atsuo T. Sasaki,[⊥] Matthew B. Boxer,[∥] Anton Simeonov,[∥] Lewis C. Cantley,[◆] Min Shen,[∥] Tinghu Zhang,[*,†,‡] Fleur M. Ferguson,[*,†,‡] and Nathanael S. Gray[*,†,‡]

[†]Department of Cancer Biology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, Massachusetts 02215, United States
[‡]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States
[§]Department of Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbruecken, Germany
[∥]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States
[⊥]Division of Hematology and Oncology, University of Cincinnati, 3125 Eden Avenue, Cincinnati, Ohio 45267-0508, United States
[#]Department of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States
[∇]Department of Oncologic Pathology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, Massachusetts 02215, United States
[○]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, United States
[◆]Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, New York 10065, United States

Supporting Information

ABSTRACT: Phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) are important molecular players in a variety of diseases, such as cancer. Currently available PI5P4K inhibitors are reversible small molecules, which may lack selectivity and sufficient cellular on-target activity. In this study, we present a new class of covalent pan-PI5P4K inhibitors with potent biochemical and cellular activity. Our designs are based on THZ-P1-2, a covalent PI5P4K inhibitor previously developed in our lab. Here, we report further structure-guided optimization and structure−activity relationship (SAR) study of this scaffold, resulting in compound **30**, which retained biochemical and cellular potency, while demonstrating a significantly improved selectivity profile. Furthermore, we confirm that the inhibitors show efficient binding affinity in the context of HEK 293T cells using isothermal CETSA methods. Taken together, compound **30** represents a highly selective pan-PI5P4K covalent lead molecule.



KEYWORDS: *PI5P4K, lipid kinase inhibitors, structure−activity relationship, covalent inhibitors, phosphoinositide, drug discovery*

The phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks), consisting of the three isoforms, PI5P4K*α*, *β*, and *γ*, are lipid kinases that catalyze phosphorylation of phosphatidylinositol 5-phosphate (PI5P) on its 4-position to form phosphatidylinositol-4,5-bisphosphate (PI-4,5-P_2).[1] In the cellular membrane, PI-4,5-P_2 is also produced by another signaling pathway in which phosphatidylinositol 4-phosphate (PI4P) is phosphorylated by phosphatidylinositol 4-phosphate 5-kinases (PI4P5Ks). Although the majority of PI-4,5-P_2 is produced via the PI4P5K pathway, the PI5P4Ks have been recognized as key regulators of many cell functions including metabolism, stress response, autophagy, and immunological processes.[2−8] Dysregulation of the PI5P4K signaling pathway has been further linked to diseases such as diabetes,[3] neurodegenerative disorders,[5] and cancers.[9−11] These findings indicate that the inhibition of PI5P4K kinase activity with a small molecule inhibitor might have a therapeutic potential across various diseases.

Special Issue: Women in Medicinal Chemistry

Received: August 30, 2019
Accepted: November 3, 2019
Published: November 3, 2019

ACS Publications © 2019 American Chemical Society

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

Downloaded via HARVARD UNIV on June 3, 2020 at 22:51:24 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

To date, several scaffolds have been developed as PI5P4K inhibitors.[5,12−15] For example, I-OMe Tyrphostin AG 538 was reported to inhibit PI5P4Kα with an IC_{50} of 1 μM,[12] while NIH-12848 was found to be a PI5P4Kγ selective compound[15] (Figure 1). However, though some of the reported compounds

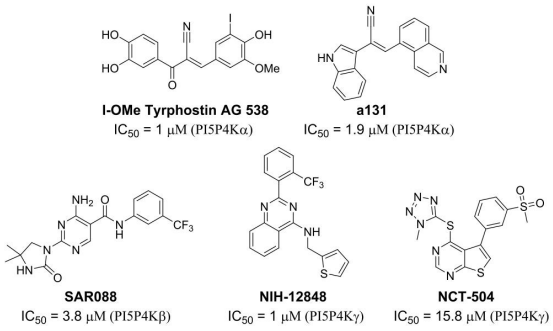


Figure 1. Reported PI5P4K inhibitors.

might seem able to covalently label kinases, such as a131, these inhibitors are reversible inhibitors with micromolar biochemical potency against PI5P4K, which limits their use in the investigation of PI5P4K function in the cellular context (Supporting Figure 6). In general, irreversible binding through a unique cysteine represents a viable strategy to improve potency and selectivity.[16] We recently reported the discovery of THZ-P1-2 as a pan-PI5P4K covalent small molecule inhibitor by targeting conserved cysteines across all three PI5P4K isoforms that are situated outside of the ATP-binding pocket. Here we report the structure−activity relationship (SAR) for this series of compounds.

Our recently reported irreversible inhibitor of PI5P4Ks, THZ-P1-2, displays submicromolar biochemical potency against PI5P4Kα kinase activity.[17] Although the THZ-P1-2/PI5P4Kα cocrystal structure (PDB-ID: 6OSP) did not resolve the loop that contains the covalently targeted cysteine (Cys293), the structure clearly revealed key noncovalent interactions, including two hydrogen bonds between the aminopyrimidine, as well as the benzamide carbonyl and Asn198, a hydrogen bond forming between the pyrimidine and Val199, and T-shaped π-stacking between the phenylenediamine group and Phe134 in the p-loop (Figure 2). In addition, Phe134 and Phe200 form a narrow passage encaging the phenylenediamine of THZ-P1-2 (Supporting Figure 7). Even though the electron-density for the warhead tail was insufficiently resolved, the acrylamide moiety seemed to be positioned toward the solvent exposed region, which allows proximity to the targetable cysteines for each isoform: Cys 293 (α), 307 or 318 (β), and 313 (γ) (Supporting Figure 8). Hydrophobic interactions between the indole and the hydrophobic pocket under the p-loop contribute to the overall affinity.

To understand the SAR of THZ-P1-2 analogs, we employed biochemical assays for PI5P4Kα (ADP Glo) and PI5P4Kβ (Transcreener), which have produced reliable IC_{50} values for THZ-P1-2. The THZ-P1-2 analogs, which contain either a variety of different indole ring systems or bulkier mono- or bicyclic aromatic six-membered rings, were first designed to investigate the tolerability of the hydrophobic pocket that is occupied by the indole group in THZ-P1-2. The synthetic

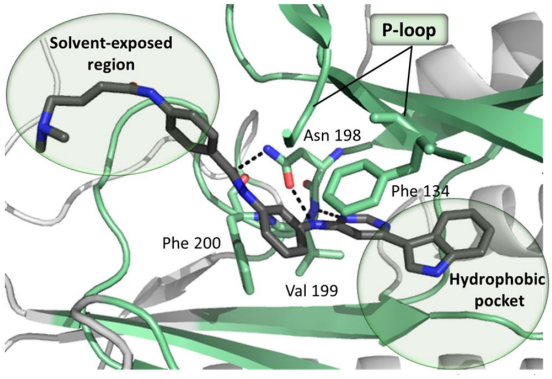
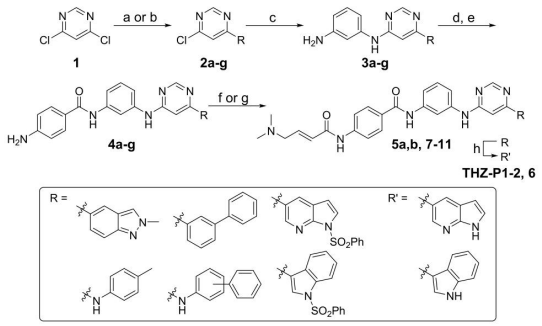


Figure 2. Cocrystal structure of PI5P4Kα and THZ-P1-2 (dark gray). The pocket bound by the inhibitor is highlighted in pale-green. H-bonds are shown as dashed bonds. For clarity, only Asn198, Val199, Phe134, and Phe200 are shown as capped sticks. PDB-ID: 6OSP.

strategy for these analogs revolved around the 4,6-dichloropyrimidine, which underwent either a Suzuki coupling reaction with aromatic boronic acids/esters or an S_N2 nucleophilic substitution reaction with aromatic amines. The second chloride group was then substituted with *m*-phenylenediamine, which was extended through an amide bond with 4-aminobenzoic acid, containing a free amino group for 4-dimethylaminocrotonamide transformation, as the final step (Scheme 1).

THZ-P1-2 showed submicromolar potency (IC_{50} of 950 nM) against PI5P4Kα, as well as an IC_{50} of 5.9 μM against PI5P4Kβ. Pyrrolo[2,3-*b*]pyridine installed in compound **6** led

Scheme 1. Synthesis of THZ-P1-2 and Compounds 6−11[a]

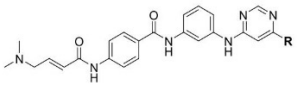


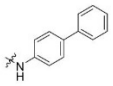
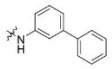
[a]Reagents and conditions: (a) $(HO)_2$B-R or respective pinacol boronic acid esters, $NaHCO_3$, $Pd(PPh_3)_2Cl_2$, ACN/H_2O, 90−100 °C, 2.5−19 h, 53−83% yield; (b) 4-methylaniline, TEA, BuOH, 140 °C, 1 h or 3/4-phenylaniline, TEA, EtOH, 120 °C, 3 h, 99%-quant. yield; (c) *m*-phenylenediamine, DIEA, NMP, 140−160 °C, 1.5 h to 2 d, 37−95% yield; (d) 4-nitrobenzoyl chloride, pyridine or TEA/DCM, rt, 5 h or 4-((*tert*-butoxycarbonyl)amino)benzoic acid, HATU, Hunig's Base, DMF, rt, 1 h, then TFA/DCM, rt, 2 h, 37% yield; (e) $SnCl_2{\cdot}2H_2O$, EtOAc/MeOH, 80 °C, 4.5 h to 2 d, or H_2, 10% Pd/C, EtOAc/MeOH or MeOH, rt, 16 h, 4−87% yield over 2 steps; (f) 4-bromocrotonyl chloride, DIEA, DCM or ACN, 0 °C, then dimethylamine, THF, rt, 1 h, 24−71% yield; (g) 4-dimethylaminocrotonic acid, DIEA, HATU, DCM, or oxalylic chloride, DMF rt, 1−12 h, 6−30% yield; (h) 1.0 M NaOH/1,4-dioxane, rt, 4−6 h, 8−60% yield.

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

to a slight decrease in potency on both isoforms, while a 2-methylindazole of compound **7** dampened the activity even further to 3.5 and 15.8 μM for PI5P4Kα and PI5P4Kβ, respectively (Table 1). Introducing an NH-linker, connecting

Table 1. IC_{50} Evaluation of THZ-P1-2 and Compounds 6–11



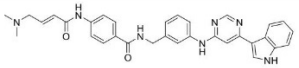
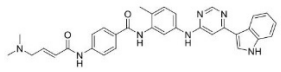
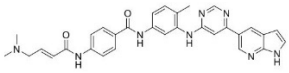
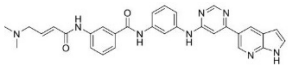
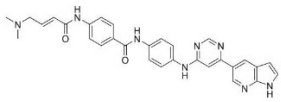
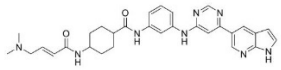
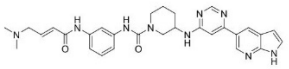
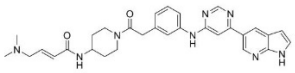
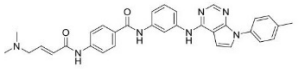
Cpd	R	IC_{50} [μM][a] PI5P4Kα	PI5P4Kβ
THZ-P1-2		0.95 ± 0.34	5.9 ± 5.4
6		2.1 ± 0.0	10.9 ± 4.0
7		3.5 ± 0.24	15.8 ± 7.5
8		1.3 ± 0.36	6.3 ± 2.6
9		1.3 ± 0.16	6.2 ± 1.3
10		19.3 ± 0.0	24.5 ± 6.1
11		17.4 ± 4.8	30.4 ± 32.8

[a]Average ± SD of three experiments, each using an 11-point titration.

the pyrimidine with a phenyl group carrying either a 4-methyl group or even a bulky 4-phenyl substituent (compounds **9** and **8**, respectively) were well tolerated and maintained IC_{50} values of around 1.3 μM and 6.2–6.3 μM for the α- and β-isoform of the PI5P4Ks, respectively. In contrast, a 3-phenylaniline or direct conjugation of pyrimidine with a 3-phenylbenzene group resulted in a more than 10-fold loss in PI5P4Kα activity (compounds **10** and **11**), underlining that these changes in the trajectory of binding seem to cause a steric clash with some residues in the hydrophobic pocket under the p-loop.

Next, to understand the SAR for the groups that bind the hinge region, as well as the solvent-exposed region, we designed a second batch of THZ-P1-2 analogs, shown in Table 2. To access these compounds, the second chloride moiety at the 6-position of the pyrimidine was substituted with different aromatic amines through S_N2 reactions, which were then further conjugated with either 3- or 4-nitro benzoic acid, 4-aminocyclohexane-1-carboxylic acid, or 4-aminopiperidine through amide bond formation to yield compounds **14**–**19** and **25** (Schemes 2 and 4). We also synthesized compound **22** with a distinct 3-aminopiperidine group installed at the 6-position of the pyrimidine, followed by a urea functionality, which was utilized to conjugate with 3-aminobenzeneamine (Scheme 3). All these compounds contain a free amino group, which was eventually converted to a 4-dimethylaminocrotonamide as a cysteine reactive warhead.

Table 2. IC_{50} Evaluation of Compounds 14–19, 22, 25, and 30

Cpd	Structure	IC_{50} [μM][a] PI5P4K α	β
14		5.2 ± 2.8	29.8 ± 12.1
15		5.3 ± 3.9	33.6 ± 8.2
16		6.5 ± 0.90	20.5 ± 11.9
17		2.1 ± 0.78	8.9 ± 2.9
18		1.99 ± 0.36	35.6 ± 2.4
19		2.0 ± 1.17	31.8 ± 4.1
22		inactive	inactive
25		2.6 ± 0.76	10.1 ± 10.2
30		1.34 ± 0.156	9.9 ± 2.8

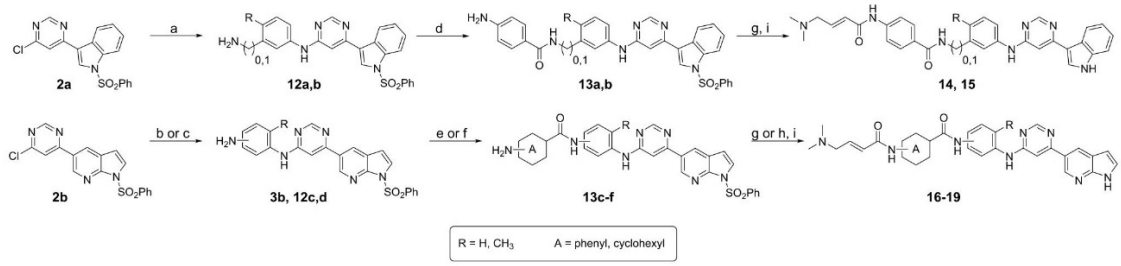
[a]Average ± SD of three experiments, each using an 11-point titration.

We first found that introducing a methyl group to either carbon-4 (compound **15**) or carbon-6 (compound **16**) led to up to 5-fold higher IC_{50}s compared to THZ-P1-2 and compound **6**. An explanation could be that the conformational changes in these compounds are potentially causing steric clashes with the relatively small passage (Supporting Figure 7) or are jeopardizing the π-stacking interaction with Phe134. Going even further, using a bulky, nonaromatic piperidine ring resulted in the overall inactive compound **22**.

A scan of different crotonamide warhead-containing linkers pointing toward the solvent-exposed area, including piperidine, cyclohexane, or phenyl rings with different substitution patterns, demonstrated high tolerability to both the composition and the conformation of the linker with respect to activity against PI5P4Kα. For example, both piperidine and cyclohexane containing compounds (compounds **25** and **19**, respectively) showed almost identical activity with IC_{50} values of 2.0–2.5 μM on PI5P4Kα. Interestingly, PI5P4Kβ seems to be more sensitive to changes in this motif since the IC_{50} of compound **19** drops by 5-fold, compared to THZ-P1-2. Similarly, the regiochemical variation of compound **6** to compound **17**, attaching the crotonamide residue in either the 3- or 4-position of the benzamide, or to compound **18**, carrying either a *m*- or *p*-phenylenediamine, displayed again only minor effects on the inhibitory activity on PI5P4Kα. While changing the warhead attachment-site on the benzamide did not have any effect on PI5P4Kβ activity either

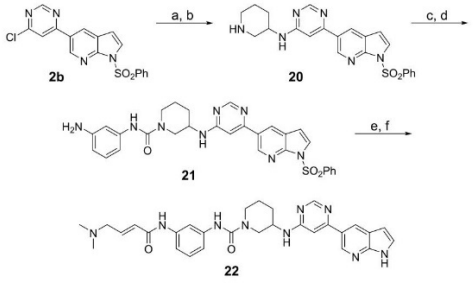
DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346–352

Scheme 2. Synthesis of Compounds 14−19[a]



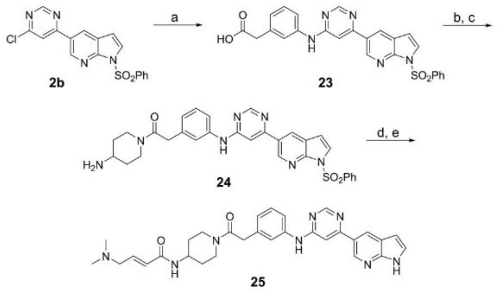
[a]Reagents and conditions: (a) anilines, DIEA, NMP, 150 °C, 6−8 h, then TFA/DCM, rt, 1 h, 67−84% yield; (b) anilines, DIEA, NMP, 150 °C, 8 h to overnight, 46−60% yield; (c) nitroaniline, Cs_2CO_3, X-Phos, $Pd_2(dba)_3$, 1,4-dioxane, 104 °C, 3 h, then $SnCl_2$, MeOH, 80 °C, 8 h, 61% yield; (d) carboxylic acid chlorides, pyridine, rt, overnight, then $SnCl_2{\cdot}2H_2O$, EtOAc/MeOH, 80 °C, 1.5−2 d, 23−39% yield over 2 steps; (e) carboxylic acid chlorides, DIEA or TEA, DCM, 0 °C, 5 min or rt, 1.5 h, then H_2, 10% Pd/C, MeOH or EtOAC/MeOH, rt, 12 h to overnight, 36−51% over 2 steps; (f) carboxylic acids, HATU, DIEA, DCM, rt, 0.5−3 h, then HCl, MeOH or EtOAc, rt, 0.5 h to overnight, 54−85% over 2 steps; (g) 4-bromocrotonyl chloride, DIEA, DCM or ACN, 0 °C, 5 min to 0.5 h, then dimethylamine (2.0 M in THF), rt, 0.5−2 h, 30−91% yield; (h) 4-dimethylaminocrotonic acid hydrochloride, DIEA, HATU, DCM, rt, 12 h, 30% yield; (i) 1.0 M NaOH/1,4-dioxane, 10 °C to rt, 3−6 h, 10−46% yield.

Scheme 3. Synthesis of Compound 22[a]



[a]Reagents and conditions: (a) *tert*-butyl 3-aminopiperidine-1-carboxylate, DIEA, NMP, 110 °C, overnight, 50% yield; (b) TFA/DCM, rt, overnight, quant. yield; (c) 3-nitroaniline, triphosgene, TEA, DCM, 0 °C, 2 h, then rt, overnight, 83% yield; (d) $SnCl_2$, EtOAc/MeOH, reflux, overnight, 74% yield; (e) 4-bromocrotonyl chloride, DIEA, DCM, 0 °C, 5 min, then dimethylamine (2.0 M in THF), rt, 1 h, 68% yield; (f) 1.0 M NaOH/1,4-dioxane, 10 °C to rt, 4 h, 30% yield.

Scheme 4. Synthesis of Compound 25[a]



[a]Reagents and conditions: (a) 2-(3-aminophenyl)acetic acid, DIEA, NMP, 140 °C, 12 h, 6% yield; (b) *tert*-butyl piperidin-4-ylcarbamate, HATU, DIEA, DMF, rt, 0.5 h, 51% yield; (c) HCl, EtOAc, rt, 0.5 h, 90% yield; (d) 4-dimethylaminocrotonic acid hydrochloride, DIEA, HATU, DCM, rt, 0.5 h, 50% yield; (e) 1.0 M NaOH/1,4-dioxane, rt, 4 h, 56% yield.

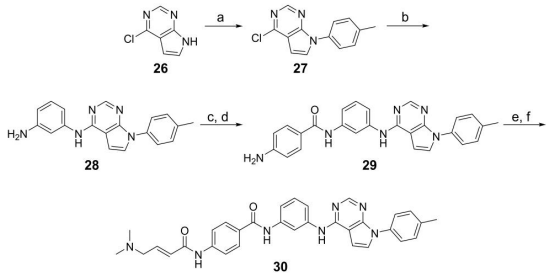
(compounds **6** and **17**), changes in the phenylenediamine moiety led to a 3.5-fold higher IC_{50} (compound **18**), underlining a higher sensitivity of the β-isoform to alterations in this region. However, an extended linker with an insertion of a methylene group, shifting the benzamide carbonyl further away from the hinge region, in compound **14** became more resistant to both, PI5P4Kα and PI5P4Kβ.

Inspired by the SAR information we obtained above, as well as the cocrystal structure of our inhibitor THZ-P1-2 with PI5P4Kα,[17] we cyclized the pyrimidine ring from carbon-4 and -5 to yield a bicyclic pyrrolo[2,3-*d*]pyrimidine and introduced a 4-methylphenyl, a favored residue (e.g., compound **9**), attached to the N atom of the pyrrolo ring system. We hypothesized that this could potentially improve the overall selectivity of the scaffold while maintaining on-target activity. The compound was synthesized through the coupling of 4-chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine with 4-iodotoluene in a copper-catalyzed nucleophilic substitution reaction. The free chloride was then coupled with *m*-phenylenediamine, extended with benzoic acid via an amide bond, and the free amino group was finally converted to a crotonamide as already described in Scheme 1 (Scheme 5), leading to active compound **30** with IC_{50} values of 1.3 and 9.9 μM for PI5P4Kα and PI5P4Kβ, respectively.

As previously reported, the selectivity profiling of THZ-P1-2[17] with KINOMEscan has shown that this compound exhibits cross activity with the type I lipid kinase PIKFYVE, as well as several protein kinases including BRK, TYK2, and Abl. We next questioned whether the bicyclic pyrrolo[2,3-*d*]pyrimidine scaffold on compound **30** could dial out these off-targets, making it a more selective PI5P4K inhibitor. Therefore, another KINOMEscan profiling assay for compound **30** was undertaken. At 1 μM, compound **30** lost almost all off-targets seen with THZ-P1-2, including the PI5P4K-related PIKFYVE kinase (Figure 3). An ADP-Glo biochemical assay (commercially available from Carna Biosciences, Inc.) done with PIKFYVE confirmed these findings as an IC_{50} value of 40.1 nM for THZ-P1-2, whereas compound **30** did not show any activity in this assay format, which was also true for other off-targets (Supporting Table 2). Since the two compounds solely differ in their headgroup moieties, facing the hydrophobic pocket in the binding site, potential reasons for the selectivity

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

Scheme 5. Synthesis of Compound 30[a]



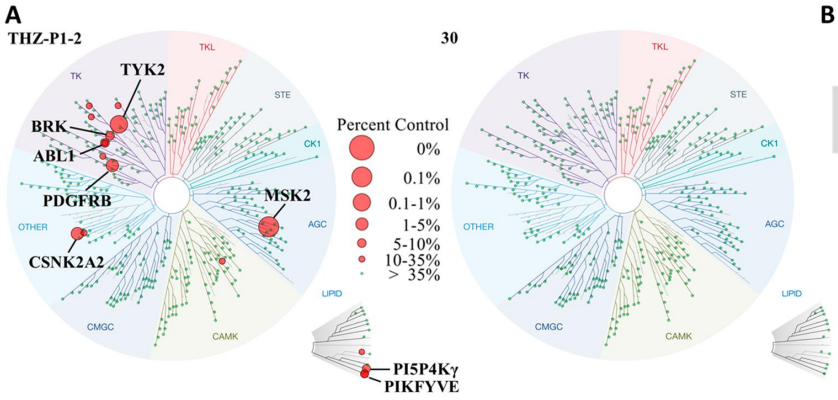
[a]Reagents and conditions: (a) 4-iodotoluidine, *trans*-1,2-cyclohexanediamine, K_3PO_4, CuI, toluene, 110 °C, 1 d, 48% yield; (b) *m*-phenylenediamine, DIEA, NMP, 150 °C, 4 d, quant. yield; (c) 4-nitrobenzoyl chloride, pyridine, rt, overnight; (d) $SnCl_2 \cdot 2H_2O$, EtOAc/MeOH, 80 °C, 1 d, 40% yield over 2 steps; (e) 4-bromocrotonyl chloride, DIEA, ACN, 0 °C; (f) dimethylamine (2.0 M in THF), rt, 1 h, 4% yield over 2 steps.

improvement could be the slightly bulkier headgroup in compound **30**, as well as the missing hydrogen bond offering NH group of the indole found in THZ-P1-2. It should be noted that PI5P4Kα is not assessed in the KINOMEscan panel. Although PI5P4Kβ and PI5P4Kγ were included in the kinases tested, we could not see significant binding of both compounds tested, despite their biochemical potency and cellular target engagement. Hence, the kinome-wide profiling data (Figure 3A) was used to investigate potential off-targets.

To assess the cellular activity of the THZ-P1-2 analogs, we used a cellular thermal shift assay (CETSA) in which we treated HEK 293T cells with our most potent pan-PI5P4K inhibitors. We first determined the optimal incubation temperature for the assay, by comparing melting curves of cells treated with either DMSO or 10 μM THZ-P1-2 at temperatures ranging from 40 to 67 °C (Figure 4A). While we observed loss of soluble PI5P4Kα and PI5P4Kβ between 49 and 52 °C with the DMSO control, THZ-P1-2 stabilized the enzymes to temperatures up to 58 to 61 °C, showing that THZ-P1-2 exhibits good pan-PI5P4K binding in the cellular context. We determined 52 °C to be the optimal assay temperature to investigate cellular target engagement of our compounds since incubation at this temperature resulted in still optimal stabilization of soluble PI5P4Kα/β with THZ-P1-2, while the DMSO-treated cells show already an immense loss of soluble protein. This temperature was used for further CETSA experiments. Next, an isothermal CETSA was performed to compare cellular target engagement of our top inhibitors, showing good target binding activity of all selected compounds in HEK 293T cells (Figure 4B,C). Dose−response curves were obtained using the same isothermal CETSA conditions. Surprisingly, compounds **8** and **30** generally showed slightly less cellular activity compared to THZ-P1-2, while compound **9** seems to be slightly more active in cells (Figure 4D,E and Supporting Figure 5A,B).

Ultimately, comparing our lead compounds, THZ-P1-2 and compound **30**, with their reversible analogs (THZ-P1-2-R and compound **30-R**, respectively) led to an overall decrease in biochemical activity of 3−4-fold, except for compound **30** and its PI5P4Kα activity, which did not change significantly (Supporting Figure 9A). The same trend was seen in an isothermal CETSA experiment, although compound **30** now seems to show similar potency changes on both, PI5P4Kα and -β (Supporting Figure 9B,C). However, the observed potency changes between covalent and noncovalent analogs were again more pronounced with THZ-P1-2. To make sure that compound **30** is still a covalent PI5P4K inhibitor, an intact mass spectrometry experiment was undertaken, which showed indeed covalent labeling of PI5P4Kα, -β, and -γ by compound **30** (Supporting Figure 10). These results indicate that compound **30** might be a slower covalent binder than THZ-P1-2, which could explain not only the biochemical results but also the weaker cellular activity of compound **30**.

In conclusion, the discovery of THZ-P1-2 and its analogs present a novel class of covalent inhibitors for PI5P4Ks. The SAR study reported here provides a rationale for further development, by systematically exploring the tolerability of each binding region for chemical modification. The selectivity profiling indicated a highly selective PI5P4K inhibitor is achievable through the modification of the core 4,6-pyrimidine scaffold, shown by compound **30**. Furthermore, investigating the cellular target engagement, using isothermal CETSA methods, we were able to confirm that the inhibitors show efficient binding affinity in the context of HEK 293T cells.



Kinase	THZ-P1-2 %contr.	30 %contr.
MSK2	0.1	100
TYK2b	0.9	100
CSNK2A2	1.3	90
PDGFRb	4.3	70
PIKFYVE	6.4	100

Figure 3. (A) KINOMEscan results for THZ-P1-2 and compound **30**. Selectivity profile was determined for 469 kinases as percent DMSO control at 1 μM inhibitor concentration (excluding mutant, atypical, and pathogen). (B) Table comparing percent DMSO control for top five off-targets of THZ-P1-2 to compound **30**.

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

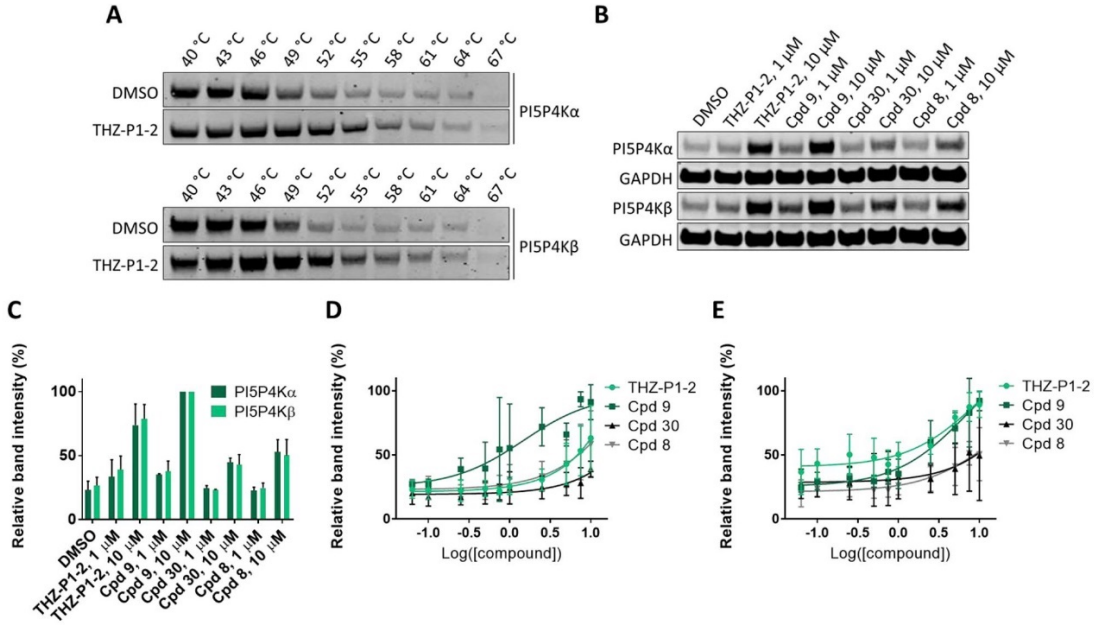


Figure 4. (A) Melting-curves (T_m) of PI5P4Kα and PI5P4Kβ. Treatment with THZ-P1-2 at 10 μM for 1 h, in comparison to DMSO. (B) Representative Western blot of isothermal CETSA at 52 °C (HEK 293T cells) with selected PI5P4Kα/β inhibitors at 1 or 10 μM. (C) Quantified relative band intensity (%) of isothermal CETSA experiment (Figure 4B). Mean of three independent experiments. (D) Dose−response curve of isothermal CETSA for PI5P4Kα at 52 °C (HEK 293T cells) with selected PI5P4Kα/β inhibitors. Mean of three independent experiments (representative Western blot, see Supporting Figure 5A). (E) Dose−response curve of isothermal CETSA for PI5P4Kβ at 52 °C (HEK 293T cells) with selected PI5P4Kα/β inhibitors. Mean of three independent experiments (representative Western blot, see Supporting Figure 5B).

These findings suggest that compound **30** is a selective, covalent pan-PI5P4K inhibitor, which may overcome limitations of cellular activity of reported PI5P4K inhibitors. Further efforts to optimize cellular potency of compound **30** and its pharmacokinetics for *in vivo* studies are ongoing.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmedchemlett.9b00402.

Experimental and characterization data for all new compounds; additional biological data and protocols (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(T.Z.) E-mail: Tinghu_Zhang@dfci.harvard.edu.
*(F.M.F.) E-mail: FleurM_Ferguson@dfci.harvard.edu.
*(N.S.G.) E-mail: Nathanael_gray@dfci.harvard.edu.

ORCID

Adam Yasgar: 0000-0001-7350-1402
Matthew D. Hall: 0000-0002-5073-442X
Hyuk-Soo Seo: 0000-0003-0646-2102
Fleur M. Ferguson: 0000-0003-4091-7617
Nathanael S. Gray: 0000-0001-5354-7403

Author Contributions

T.D.M., T.Z., F.M.F., and N.S.G. conceived the project and designed the research strategy. T.D.M., T.Z., and F.M.F. designed and synthesized the compounds. S.C.S. performed compound testing and helped with planning experimental strategies. M.B.B. synthesized additional compounds for the biochemical assay validation. A.Y., M.I.D., A.S., and M.D.H. developed and validated the biochemical assays. A.Y. performed the biochemical assays under supervision of M.D.H. and M.S. H.-S.S. and S.D.-P. generated the cocrystal structure. A.T.S. provided reagents and methods for the biochemical assays. S.C.S., J.D.C., S.B.F., and J.A.M. performed the intact mass spectrometry analysis. T.D.M. planned and performed the cellular assays. H.S. and L.C.C. were involved in the planning and provided expertise and feedback for the project. The manuscript was written by T.D.M. and contributions of all authors. All authors have given approval to the final version of the manuscript. T.Z., F.M.F., and N.S.G. supervised the research.

Funding

The authors would like to thank the National Institutes of Health for their generous financial support through grants R01 CA197329 (to N.S.G., subcontracted to S.D.P.), R01 NS089815 and R03 MH096575 (to A.T.S.), R01 CA222218 and R03 CA250020 (to J.A.M.), as well as R35 CA197588, R01 GM041890, and U54 CA210184 (to L.C.C.). This work was further supported by the Breast Cancer Research Foundation (to L.C.C.) and the NIH intramural research program (NCATS).

Notes

The authors declare the following competing financial interest(s): N.S.G. is a founder, SAB member, and equity holder in Gatekeeper, Syros, Petra, C4, B2S, and Soltego. The Gray lab receives or has received research funding from

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

Novartis, Takeda, Astellas, Taiho, Janssen, Kinogen, Voronoi, Her2llc, Deerfield, and Sanofi. N.S.G. and T.Z. are inventors on a patent application covering chemical matter in this publication, owned by Dana-Farber Cancer Institute. L.C.C. is a founder and member of the Board of Directors (BOD) of Agios Pharmaceuticals and is a founder and receives research support from Petra Pharmaceuticals. These companies are developing novel therapies for cancer. J.A.M. serves on the SAB of 908 Devices.

■ ACKNOWLEDGMENTS

The authors would like to thank M. Kostic for her help and expertise with writing the manuscript.

■ ABBREVIATIONS

ACN, acetonitrile; BRK, breast tumor-related kinase; CETSA, cellular thermal shift assay; Cpd, compound; DCM, dichloromethane; DIEA, *N*,*N*-diisopropylethylamine; DMF, dimethylformamide; DMSO, dimethyl sulfoxide; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; HATU, 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxide hexafluorophosphate; NMP, *N*-methyl-2-pyrrolidone; $Pd_2(dba)_3$, tris(dibenzylideneacetone)-dipalladium(0); PI-4,5-P_2, phosphatidylinositol-4,5-bisphosphate; PI4P, phosphatidylinositol 4-phosphate; PI4P5K, phosphatidylinositol 4-phosphate 5-kinase; PI5P, phosphatidylinositol 5-phosphate; PI5P4K, phosphatidylinositol 5-phosphate 4-kinase; PIKFYVE, phosphatidylinositol 3-phosphate 5-kinase; rt, room temperature; SAR, structure−activity relationship; SD, standard deviation; TEA, triethylamine; TFA, trifluoroacetic acid; THF, tetrahydrofuran; TYK2, tyrosine kinase 2; X-Phos, 2-dicyclohexylphosphino-2′,4′,6′-triisoropylbiphenyl

■ REFERENCES

(1) Rameh, L. E.; Tolias, K. F.; Duckworth, B. C.; Cantley, L. C. A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate. *Nature* **1997**, *390* (6656), 192−196.

(2) Hu, A.; Zhao, X.-T.; Xiao, T.; Fu, T.; Wang, Y.; Liu, Y.; Shi, X.-J.; Luo, J.; Song, B.-L. PIP4K2A regulates intracellular cholesterol transport through modulating $PI(4,5)_2$ homeostasis. *J. Lipid Res.* **2018**, *59*, 507−514.

(3) Lamia, K. A.; Peroni, O. D.; Kim, Y.-B.; Rameh, L. E.; Kahn, B. B.; Cantley, L. C. Increased Insulin Sensitivity and Reduced Adiposity in Phosphatidylinositol 5-Phosphate 4-Kinase $\beta^{-/-}$ Mice. *Mol. Cell. Biol.* **2004**, *24* (11), 5080−5087.

(4) Shim, H.; Wu, C.; Ramsamooj, S.; Bosch, K. N.; Chen, Z.; Emerling, B. M.; Yun, J.; Liu, H.; Choo-Wing, R.; Yang, Z.; Wulf, G. M.; Kuchroo, V. K.; Cantley, L. C. Deletion of the gene *Pip4k2c*, a novel phosphatidylinositol kinase, results in hyperactivation of the immune system. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (27), 7596−7601.

(5) Al-Ramahi, I.; Giridharan, S. S. P.; Chen, Y.-C.; Patnaik, S.; Safren, N.; Hasegawa, J.; de Haro, M.; Wagner Gee, A. K.; Titus, S. A.; Jeong, H.; Clarke, J.; Krainc, D.; Zheng, W.; Irvine, R. F.; Barmada, S.; Ferrer, M.; Southall, N.; Weisman, L. S.; Botas, J.; Marugan, J. J. Inhibition of PIP4Kγ ameliorates the pathological effects of mutant huntingtin protein. *eLife* **2017**, *6*, No. e29123.

(6) Lundquist, M. R.; Goncalves, M. D.; Loughran, R. M.; Possik, E.; Vijayaraghavan, T.; Yang, A.; Pauli, C.; Ravi, A.; Verma, A.; Yang, Z.; Johnson, J. L.; Wong, J. C. Y.; Ma, Y.; Hwang, K. S.-K.; Weinkove, D.; Divecha, N.; Asara, J. M.; Elemento, O.; Rubin, M. A.; Kimmelman, A. C.; Pause, A.; Cantley, L. C.; Emerling, B. M. Phosphatidylinositol-5-Phosphate 4-Kinases Regulate Cellular Lipid Metabolism by Facilitating Autophagy. *Mol. Cell* **2018**, *70* (3), 531−543.

(7) Bulley, S. J.; Droubi, A.; Clarke, J. H.; Anderson, K. E.; Stephens, L. R.; Hawkins, P. T.; Irvine, R. F. In B cells, phosphatidylinositol 5-phosphate 4-kinase-α synthesizes $PI(4,5)P_2$ to impact mTORC2 and Akt signaling. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (38), 10571−10576.

(8) Keune, W.-J.; Jones, D. R.; Divecha, N. PtdIns5*P* and Pin1 in oxidative stress signaling. *Advances in Biological Regulation* **2013**, *53* (2), 179−189.

(9) Jude, J. G.; Spencer, G. J.; Somerville, T. D. D.; Jones, D. R.; Divecha, N.; Somervaille, T. C. P. A targeted knockdown screen of genes coding for phosphoinositide modulators identifies *PIP4K2A* as required for acute myeloid leukemia cell proliferation and survival. *Oncogene* **2015**, *34* (10), 1253−1262.

(10) Luoh, S.-W.; Venkatesan, N.; Tripathi, R. Overexpression of the amplified *Pip4k2β* gene from 17q11−12 in breast cancer cells confers proliferation advantage. *Oncogene* **2004**, *23*, 1354−1363.

(11) Emerling, B. M.; Hurov, J. B.; Poulogiannis, G.; Tsukazawa, K. S.; Choo-Wing, R.; Wulf, G. M.; Bell, E. L.; Shim, H.-S.; Lamia, K. A.; Rameh, L. E.; Bellinger, G.; Sasaki, A. T.; Asara, J. M.; Yuan, X.; Bullock, A.; DeNicola, G. M.; Song, J.; Brown, V.; Signoretti, S.; Cantley, L. C. Depletion of a Putatively Druggable Class of Phosphatidylinositol Kinases Inhibits Growth of p53-Null Tumors. *Cell* **2013**, *155* (4), 844−857.

(12) Davis, M. I.; Sasaki, A. T.; Shen, M.; Emerling, B. M.; Thorne, N.; Michael, S.; Pragani, R.; Boxer, M.; Sumita, K.; Takeuchi, K.; Auld, D. S.; Li, Z.; Cantley, L. C.; Simeonov, A. A Homogeneous, High-Throughput Assay for Phosphatidylinositol 5-Phosphate 4-Kinase with a Novel, Rapid Substrate Preparation. *PLoS One* **2013**, *8* (1), No. e54127.

(13) Kitagawa, M.; Liao, P.-J.; Lee, K. H.; Wong, J.; Shang, S. C.; Minami, N.; Sampetrean, O.; Saya, H.; Lingyun, D.; Prabhu, N.; Diam, G. K.; Sobota, R.; Larsson, A.; Nordlung, P.; McCormick, F.; Ghosh, S.; Epstein, D. M.; Dymock, B. W.; Lee, S. H. Dual blockade of the lipid kinase PIP4Ks and mitotic pathways leads to cancer-selective lethality. *Nat. Commun.* **2017**, *8* (1), 2200.

(14) Voss, M. D.; Czechtizky, W.; Li, Z.; Rudolph, C.; Petry, S.; Brummerhop, H.; Langer, T.; Schiffer, A.; Schaefer, H.-L. Discovery and pharmacological characterization of a novel small molecule inhibitor of phosphatidylinositol-5-phosphate 4-kinase, type II, beta. *Biochem. Biophys. Res. Commun.* **2014**, *449* (3), 327−331.

(15) Clarke, J. H.; Giudici, M.-L.; Burke, J. E.; Williams, R. L.; Maloney, D. J.; Marugan, J.; Irvine, R. F. The function of phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ) explored using a specific inhibitor that targets the PI5P-binding site. *Biochem. J.* **2015**, *466* (2), 359−367.

(16) Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing Irreversible Inhibitors of the Protein Kinase Cysteinome. *Chem. Biol.* **2013**, *20* (2), 146−159.

(17) Sivakumaren, S. C.; Shim, H.; Zhang, T.; Ferguson, F. M.; Lundquist, M. R.; Browne, C. M.; Seo, H.-S.; Paddock, M. N.; Manz, T. D.; Jiang, B.; Hao, M.-F.; Krishnan, P.; Wang, D. G.; Yang, J.; Kwiatkowski, N. P.; Ficarro, S. B.; Cunningham, J. M.; Marto, J. A.; Dhe-Paganon, S.; Cantley, L. C.; Gray, N. S. Targeting the PI5P4K lipid kinase family in cancer using novel covalent inhibitors. *bioRxiv* **2019**; DOI: 10.1101/819961.

DOI: 10.1021/acsmedchemlett.9b00402
ACS Med. Chem. Lett. 2020, 11, 346−352

Supporting Information

Structure-Activity Relationship Study of Covalent Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz,[†,‡,⊥] Sindhu C. Sivakumaren,[†,‡] Adam Yasgar,[Δ] Matthew D. Hall,[Δ] Mindy I. Davis,[Δ] Hyuk-Soo Seo,[†,‡] Joseph D. Card,[Ψ,Θ,φ] Scott B. Ficarro,[Ψ,Θ,φ] Hyeseok Shim,[§] Jarrod A. Marto,[Ψ,Θ,φ] Sirano Dhe-Paganon,[†,‡] Atsuo T. Sasaki,[π] Matthew B. Boxer,[Δ] Anton Simeonov,[Δ] Lewis C. Cantley,[§] Min Shen,[Δ] Tinghu Zhang,[†,‡,*] Fleur M. Ferguson,[†,‡,*] and Nathanael S. Gray[†,‡,*]

†Department of Cancer Biology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, MA 02215, USA

‡Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA

⊥Department of Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbruecken, Germany

ΔNational Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA

πDivision of Hematology and Oncology, University of Cincinnati, 3125 Eden Ave., Cincinnati, OH 45267-0508, USA

ΨDepartment of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Ma 02215, USA

ΘDepartment of Oncologic Pathology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, MA 02215, USA

φDepartment of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA

§Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, NY 10065, USA

Table of Contents

1. Supporting Tables and Figures 3
   1.1 Supporting Tables 3
   1.2 Supporting Figures 14
2. Experimental Procedures 21
   2.1 Synthesis 21
      2.1.1 General Methods 21
      2.1.2 Synthesis of THZ-P1-2 and compounds 6-11 22
      2.1.3 Synthesis of compounds 14-19 33
      2.1.4 Synthesis of compound 22 41
      2.1.5 Synthesis of compound 25 42
      2.1.6 Synthesis of compound 30 44
   2.2 X-ray Crystallography 45
   2.3 Mass Spectrometry 45
   2.4 Biochemical Assay Protocols 45
      2.4.1 Materials and Methods 45
      2.4.2 ADP-Glo^{TM} Kinase Assay (PI5P4Kα) 46
      2.4.3 GDP Transcreener Assay (PI5P4Kβ) 47
      2.4.4 Assay Optimization and Miniaturization 47
   2.5 Kinome Profiling 49
   2.6 CETSA Protocol 49
      2.6.1 Cell culture 49
      2.6.2 Compounds and Reagents 49
      2.6.3 Cellular Thermal Shift Assay 49
      2.6.4 Western Blot Analysis 50
3. Abbreviations 51
4. References 51

1. Supporting Tables and Figures

1.1 Supporting Tables

Supporting Table 1. Full KINOMEscan datasets

Kinase ID	Compound conc. [μM]	Percent Control	
		THZ-P1-2	Cpd 30
AAK1	1	100	98
ABL1(E255K)-phosphorylated	1	21	78
ABL1(F317I)-nonphosphorylated	1	93	100
ABL1(F317I)-phosphorylated	1	100	100
ABL1(F317L)-nonphosphorylated	1	71	100
ABL1(F317L)-phosphorylated	1	44	100
ABL1(H396P)-nonphosphorylated	1	5.8	88
ABL1(H396P)-phosphorylated	1	6.6	89
ABL1(M351T)-phosphorylated	1	12	100
ABL1(Q252H)-nonphosphorylated	1	24	64
ABL1(Q252H)-phosphorylated	1	20	78
ABL1(T315I)-nonphosphorylated	1	93	100
ABL1(T315I)-phosphorylated	1	100	100
ABL1(Y253F)-phosphorylated	1	2.4	100
ABL1-nonphosphorylated	1	13	99
ABL1-phosphorylated	1	8.1	76
ABL2	1	36	92
ACVR1	1	94	77
ACVR1B	1	100	91
ACVR2A	1	94	100
ACVR2B	1	91	100
ACVRL1	1	90	70
ADCK3	1	77	89
ADCK4	1	97	93
AKT1	1	95	100
AKT2	1	95	88
AKT3	1	100	98
ALK	1	92	93
ALK(C1156Y)	1	93	100
ALK(L1196M)	1	71	100
AMPK-alpha1	1	93	100
AMPK-alpha2	1	100	100
ANKK1	1	97	100
ARK5	1	83	100
ASK1	1	100	99
ASK2	1	95	100
AURKA	1	99	93

AURKB	1	100	100
AURKC	1	83	100
AXL	1	92	100
BIKE	1	89	100
BLK	1	34	87
BMPR1A	1	89	91
BMPR1B	1	97	100
BMPR2	1	100	100
BMX	1	81	92
BRAF	1	96	100
BRAF(V600E)	1	100	100
BRK	1	6.6	100
BRSK1	1	99	86
BRSK2	1	95	100
BTK	1	96	100
BUB1	1	95	100
CAMK1	1	93	100
CAMK1B	1	100	100
CAMK1D	1	93	100
CAMK1G	1	100	100
CAMK2A	1	93	95
CAMK2B	1	100	89
CAMK2D	1	100	96
CAMK2G	1	100	97
CAMK4	1	97	78
CAMKK1	1	96	59
CAMKK2	1	94	96
CASK	1	92	100
CDC2L1	1	94	100
CDC2L2	1	93	100
CDC2L5	1	100	100
CDK11	1	92	96
CDK2	1	100	100
CDK3	1	88	100
CDK4	1	71	100
CDK4-cyclinD1	1	48	96
CDK4-cyclinD3	1	70	77
CDK5	1	89	100
CDK7	1	37	95
CDK8	1	100	100
CDK9	1	84	100
CDKL1	1	96	99
CDKL2	1	87	100
CDKL3	1	96	98
CDKL5	1	100	100
CHEK1	1	98	73

CHEK2	1	100	97
CIT	1	48	82
CLK1	1	96	100
CLK2	1	96	98
CLK3	1	94	100
CLK4	1	95	91
CSF1R	1	67	86
CSF1R-autoinhibited	1	83	98
CSK	1	93	100
CSNK1A1	1	92	79
CSNK1A1L	1	85	100
CSNK1D	1	63	100
CSNK1E	1	84	100
CSNK1G1	1	96	100
CSNK1G2	1	81	94
CSNK1G3	1	99	100
CSNK2A1	1	27	100
CSNK2A2	1	1.3	90
CTK	1	100	100
DAPK1	1	100	100
DAPK2	1	91	86
DAPK3	1	100	64
DCAMKL1	1	100	94
DCAMKL2	1	100	100
DCAMKL3	1	96	90
DDR1	1	67	67
DDR2	1	92	89
DLK	1	93	100
DMPK	1	82	96
DMPK2	1	93	81
DRAK1	1	75	100
DRAK2	1	61	100
DYRK1A	1	95	100
DYRK1B	1	89	96
DYRK2	1	85	100
EGFR	1	22	74
EGFR(E746-A750del)	1	1.3	80
EGFR(G719C)	1	18	75
EGFR(G719S)	1	22	88
EGFR(L747-E749del, A750P)	1	9	79
EGFR(L747-S752del, P753S)	1	16	71
EGFR(L747-T751del,Sins)	1	6.8	88
EGFR(L858R)	1	18	75
EGFR(L858R,T790M)	1	73	100
EGFR(L861Q)	1	21	72
EGFR(S752-I759del)	1	14	89

EGFR(T790M)	1	95	81
EIF2AK1	1	98	100
EPHA1	1	78	96
EPHA2	1	71	98
EPHA3	1	100	100
EPHA4	1	100	84
EPHA5	1	87	100
EPHA6	1	100	95
EPHA7	1	100	94
EPHA8	1	100	94
EPHB1	1	75	82
EPHB2	1	82	92
EPHB3	1	100	94
EPHB4	1	89	93
EPHB6	1	72	100
ERBB2	1	91	70
ERBB3	1	100	100
ERBB4	1	88	100
ERK1	1	94	90
ERK2	1	87	97
ERK3	1	87	70
ERK4	1	70	100
ERK5	1	74	100
ERK8	1	74	100
ERN1	1	100	100
FAK	1	90	99
FER	1	95	93
FES	1	100	100
FGFR1	1	95	100
FGFR2	1	98	100
FGFR3	1	88	91
FGFR3(G697C)	1	91	100
FGFR4	1	92	100
FGR	1	82	87
FLT1	1	89	100
FLT3	1	90	100
FLT3(D835H)	1	50	86
FLT3(D835V)	1	37	61
FLT3(D835Y)	1	45	86
FLT3(ITD)	1	62	100
FLT3(ITD,D835V)	1	51	13
FLT3(ITD,F691L)	1	89	70
FLT3(K663Q)	1	82	82
FLT3(N841I)	1	60	92
FLT3(R834Q)	1	97	100
FLT3-autoinhibited	1	83	100

FLT4	1	88	93
FRK	1	96	100
FYN	1	92	100
GAK	1	88	98
GCN2(Kin.Dom.2,S808G)	1	75	94
GRK1	1	100	84
GRK2	1	90	98
GRK3	1	63	100
GRK4	1	96	98
GRK7	1	100	100
GSK3A	1	52	99
GSK3B	1	98	90
HASPIN	1	97	100
HCK	1	98	89
HIPK1	1	100	87
HIPK2	1	100	95
HIPK3	1	93	100
HIPK4	1	47	96
HPK1	1	100	97
HUNK	1	100	93
ICK	1	100	100
IGF1R	1	93	80
IKK-alpha	1	72	100
IKK-beta	1	73	100
IKK-epsilon	1	100	100
INSR	1	100	91
INSRR	1	82	100
IRAK1	1	85	100
IRAK3	1	56	94
IRAK4	1	100	100
ITK	1	100	94
JAK1(JH1domain-catalytic)	1	74	100
JAK1(JH2domain-pseudokinase)	1	74	80
JAK2(JH1domain-catalytic)	1	100	90
JAK3(JH1domain-catalytic)	1	97	73
JNK1	1	86	90
JNK2	1	90	100
JNK3	1	92	100
KIT	1	24	93
KIT(A829P)	1	57	100
KIT(D816H)	1	35	88
KIT(D816V)	1	4.2	98
KIT(L576P)	1	12	100
KIT(V559D)	1	11	86
KIT(V559D,T670I)	1	92	91
KIT(V559D,V654A)	1	59	100

KIT-autoinhibited	1	86	100
LATS1	1	100	97
LATS2	1	100	100
LCK	1	79	92
LIMK1	1	64	78
LIMK2	1	63	91
LKB1	1	64	71
LOK	1	94	96
LRRK2	1	91	59
LRRK2(G2019S)	1	90	55
LTK	1	93	100
LYN	1	89	100
LZK	1	100	100
MAK	1	100	100
MAP3K1	1	78	100
MAP3K15	1	100	92
MAP3K2	1	100	99
MAP3K3	1	100	100
MAP3K4	1	92	100
MAP4K2	1	99	100
MAP4K3	1	100	94
MAP4K4	1	79	100
MAP4K5	1	94	84
MAPKAPK2	1	100	100
MAPKAPK5	1	93	100
MARK1	1	95	92
MARK2	1	100	100
MARK3	1	82	95
MARK4	1	95	98
MAST1	1	97	100
MEK1	1	92	100
MEK2	1	100	83
MEK3	1	99	98
MEK4	1	100	100
MEK5	1	93	100
MEK6	1	92	95
MELK	1	100	100
MERTK	1	71	83
MET	1	98	86
MET(M1250T)	1	90	100
MET(Y1235D)	1	100	100
MINK	1	88	91
MKK7	1	100	100
MKNK1	1	87	100
MKNK2	1	21	100
MLCK	1	98	98

MLK1	1	83	68
MLK2	1	86	70
MLK3	1	73	98
MRCKA	1	100	94
MRCKB	1	100	100
MST1	1	82	97
MST1R	1	99	86
MST2	1	100	94
MST3	1	100	100
MST4	1	97	98
MTOR	1	100	96
MUSK	1	84	98
MYLK	1	99	95
MYLK2	1	91	100
MYLK4	1	76	100
MYO3A	1	100	95
MYO3B	1	100	94
NDR1	1	100	96
NDR2	1	97	73
NEK1	1	100	92
NEK10	1	100	100
NEK11	1	61	100
NEK2	1	98	99
NEK3	1	100	86
NEK4	1	100	100
NEK5	1	87	79
NEK6	1	86	100
NEK7	1	98	100
NEK9	1	98	98
NIK	1	100	100
NIM1	1	100	100
NLK	1	97	100
OSR1	1	100	100
p38-alpha	1	100	99
p38-beta	1	100	89
p38-delta	1	85	86
p38-gamma	1	100	100
PAK1	1	91	100
PAK2	1	91	98
PAK3	1	89	91
PAK4	1	98	94
PAK6	1	78	96
PAK7	1	98	100
PCTK1	1	100	100
PCTK2	1	91	100
PCTK3	1	97	95

PDGFRA	1	58	100
PDGFRB	1	4.3	70
PDPK1	1	100	100
PFCDPK1(P.falciparum)	1	100	92
PFPK5(P.falciparum)	1	91	100
PFTAIRE2	1	92	100
PFTK1	1	82	100
PHKG1	1	100	100
PHKG2	1	100	86
PIK3C2B	1	95	100
PIK3C2G	1	100	81
PIK3CA	1	100	92
PIK3CA(C420R)	1	100	100
PIK3CA(E542K)	1	100	100
PIK3CA(E545A)	1	100	97
PIK3CA(E545K)	1	100	67
PIK3CA(H1047L)	1	100	100
PIK3CA(H1047Y)	1	87	85
PIK3CA(I800L)	1	100	66
PIK3CA(M1043I)	1	100	90
PIK3CA(Q546K)	1	99	100
PIK3CB	1	100	100
PIK3CD	1	100	95
PIK3CG	1	100	100
PIK4CB	1	95	100
PIKFYVE	1	6.4	100
PIM1	1	95	91
PIM2	1	90	100
PIM3	1	96	96
PIP5K1A	1	84	96
PIP5K1C	1	20	93
PIP5K2B	1	82	96
PIP5K2C	1	9	78
PKAC-alpha	1	100	94
PKAC-beta	1	85	100
PKMYT1	1	99	95
PKN1	1	100	100
PKN2	1	76	72
PKNB(M.tuberculosis)	1	100	85
PLK1	1	100	100
PLK2	1	100	100
PLK3	1	100	90
PLK4	1	93	86
PRKCD	1	100	96
PRKCE	1	100	57
PRKCH	1	93	96

PRKCI	1	96	81
PRKCQ	1	93	95
PRKD1	1	100	91
PRKD2	1	75	100
PRKD3	1	94	75
PRKG1	1	100	100
PRKG2	1	100	100
PRKR	1	97	68
PRKX	1	100	96
PRP4	1	82	100
PYK2	1	91	86
QSK	1	95	100
RAF1	1	100	97
RET	1	67	98
RET(M918T)	1	73	97
RET(V804L)	1	99	98
RET(V804M)	1	96	100
RIOK1	1	87	100
RIOK2	1	91	100
RIOK3	1	82	100
RIPK1	1	86	77
RIPK2	1	91	96
RIPK4	1	99	100
RIPK5	1	100	96
ROCK1	1	100	100
ROCK2	1	91	96
ROS1	1	94	94
RPS6KA4(Kin.Dom.1-N-terminal)	1	0.1	100
RPS6KA4(Kin.Dom.2-C-terminal)	1	95	100
RPS6KA5(Kin.Dom.1-N-terminal)	1	96	94
RPS6KA5(Kin.Dom.2-C-terminal)	1	99	60
RSK1(Kin.Dom.1-N-terminal)	1	86	97
RSK1(Kin.Dom.2-C-terminal)	1	100	100
RSK2(Kin.Dom.1-N-terminal)	1	96	94
RSK2(Kin.Dom.2-C-terminal)	1	90	100
RSK3(Kin.Dom.1-N-terminal)	1	91	100
RSK3(Kin.Dom.2-C-terminal)	1	94	72
RSK4(Kin.Dom.1-N-terminal)	1	47	96
RSK4(Kin.Dom.2-C-terminal)	1	100	98
S6K1	1	100	95
SBK1	1	100	100
SGK	1	100	91
SgK110	1	100	99
SGK2	1	100	100
SGK3	1	89	100
SIK	1	39	100

SIK2	1	47	93
SLK	1	99	89
SNARK	1	98	81
SNRK	1	96	100
SRC	1	23	98
SRMS	1	79	93
SRPK1	1	81	94
SRPK2	1	100	59
SRPK3	1	90	90
STK16	1	88	100
STK33	1	100	80
STK35	1	100	98
STK36	1	38	96
STK39	1	100	100
SYK	1	80	75
TAK1	1	78	99
TAOK1	1	98	100
TAOK2	1	100	94
TAOK3	1	100	97
TBK1	1	95	100
TEC	1	89	98
TESK1	1	88	76
TGFBR1	1	100	100
TGFBR2	1	97	83
TIE1	1	87	97
TIE2	1	98	100
TLK1	1	94	100
TLK2	1	99	93
TNIK	1	98	100
TNK1	1	81	48
TNK2	1	48	85
TNNI3K	1	80	100
TRKA	1	72	53
TRKB	1	100	86
TRKC	1	80	69
TRPM6	1	91	97
TSSK1B	1	95	98
TSSK3	1	100	100
TTK	1	94	69
TXK	1	93	85
TYK2(JH1domain-catalytic)	1	100	87
TYK2(JH2domain-pseudokinase)	1	0.9	100
TYRO3	1	94	100
ULK1	1	100	100
ULK2	1	100	100
ULK3	1	100	100

VEGFR2	1	88	100
VPS34	1	90	100
VRK2	1	90	100
WEE1	1	90	100
WEE2	1	100	96
WNK1	1	100	100
WNK2	1	100	93
WNK3	1	100	100
WNK4	1	95	100
YANK1	1	87	100
YANK2	1	100	100
YANK3	1	84	95
YES	1	92	79
YSK1	1	95	100
YSK4	1	100	83
ZAK	1	93	92
ZAP70	1	100	100

Supporting Table 2. Off-targets of THZ-P1-2 and compound 30

Kinase	THZ-P1-2		Cpd 30	
	%control	IC_{50} [μM][a]	%control	IC_{50} [μM][a]
PI5P4Kβ	82	5.9	96	9.9
PIP5K1C	20	0.0821	93	>10
MSK2	0.1	>10	100	>10
TYK2	0.9	2.2	100	>10
TNK1	81	7.33	48	>10
CSNK2A2	1.3	1.97	90	>10
PDGFRb	4.3	3.01	70	>10
PIKFYVE	6.4	0.0401	100	>10

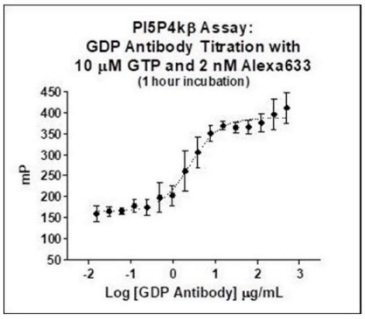
[a] IC_{50} data were obtained from our data set, Carna Biosciences, or Invitrogen (Thermo Fisher Scientific).

1.2 Supporting Figures

A

Time	60 min	
Conc. [µg/mL]	Mean	SD
500	412	36
250	396	37
125	376	22
62.5	366	16
31.25	365	12
15.6	369	10
7.8	351	20
3.91	306	36
1.95	261	48
0.98	203	24
0.488	198	36
0.244	174	18
0.122	178	14
0.061	166	8
0.0305	165	11
0.0153	159	19
HILLSLOP	1.228	
EC_{50}	2.553	
EC_{85}	**10.5**	

B

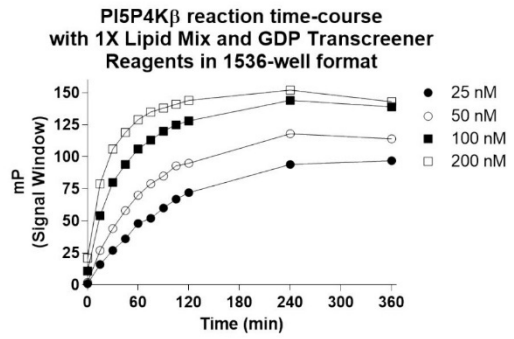


C

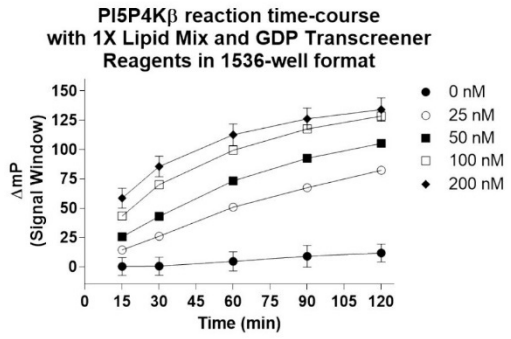
%GTP Conversion	[GDP], µM	Mean	SD	%CV	ΔmP	Z'-factor
0	0	334	8	2	-	-
1	0.1	299	10	3	37	-0.58
10	1	210	9	4	126	0.59
50	5	157	11	7	179	0.67
100	10	146	10	7	191	0.72

Supporting Figure 1. A. GDP antibody titration summary table. **B.** GDP antibody titration summary curve. **C.** %GTP conversion table with Z'-factor determination.

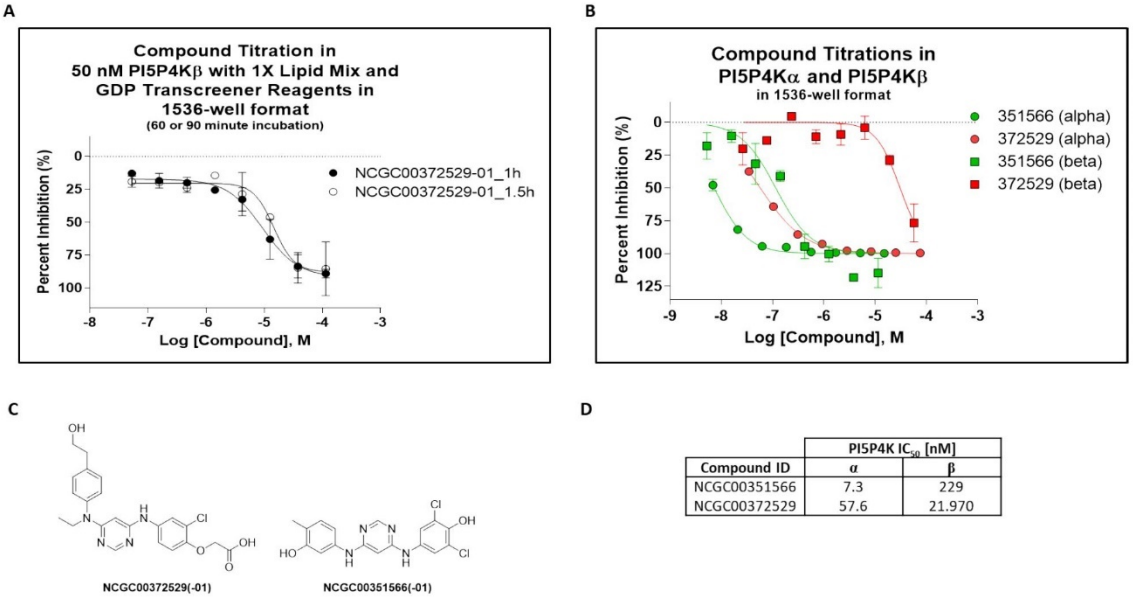
A



B

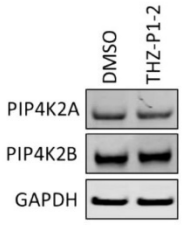


Supporting Figure 2. A. PI5P4Kβ reaction time-course curve (0-360 min). **B.** PI5P4Kβ reaction time-course curve (0-120 min).

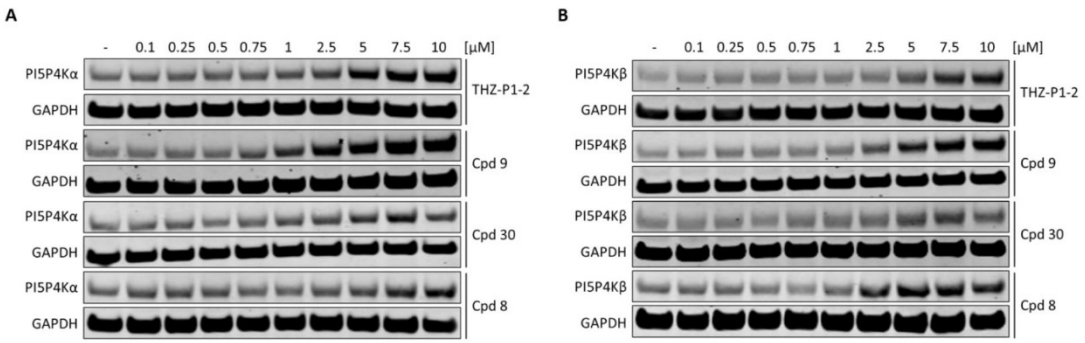


Compound ID	PI5P4K IC_{50} [nM] α	β
NCGC00351566	7.3	229
NCGC00372529	57.6	21.970

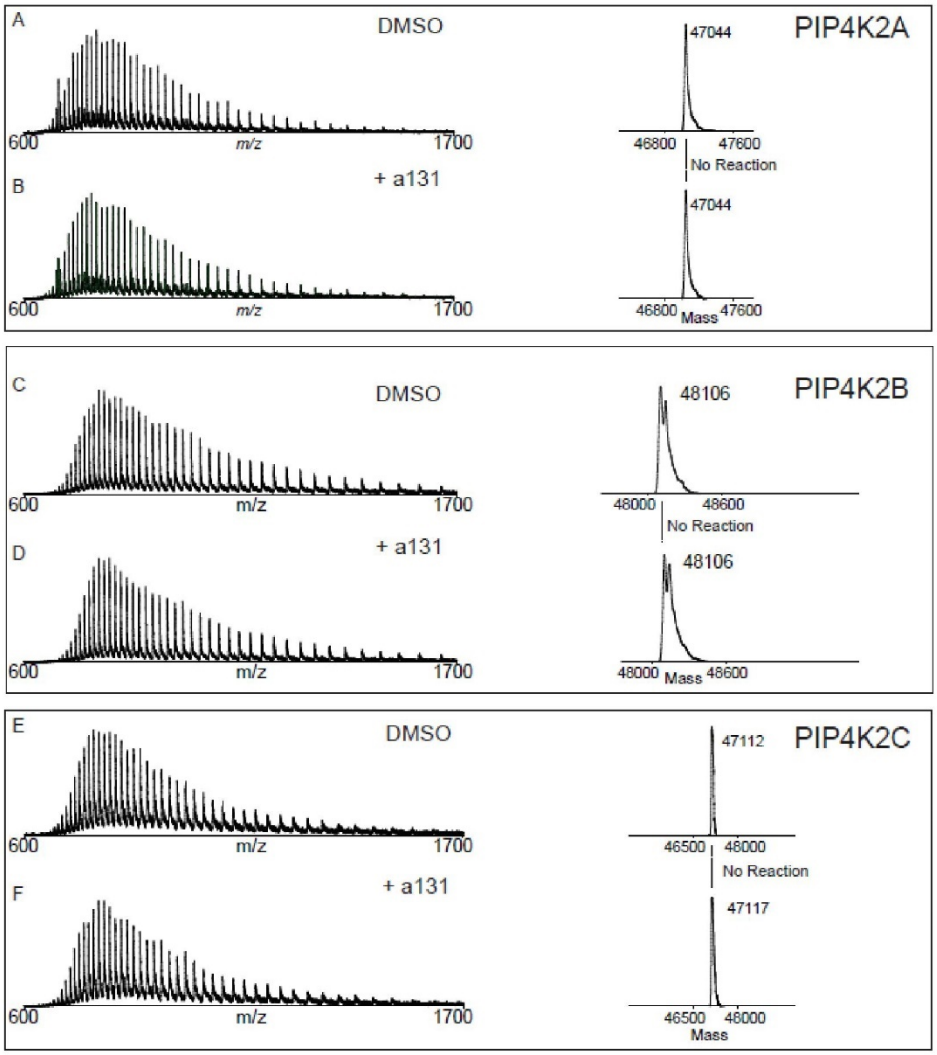
Supporting Figure 3. A. Dose-response curve (Transcreener assay) for PI5P4Kβ inhibition with reference compound NCGC00372529 at 1 h and 1.5 h. **B.** Combined dose-response curves for PI5P4Kα and PI5P4Kβ with reference compounds NCGC00351566 and NCGC00372529. **C.** Structures of reference compounds NCGC00372529, and NCGC00351566. **D.** IC_{50} summary for reference compounds NCGC00372529 and NCGC00351566.



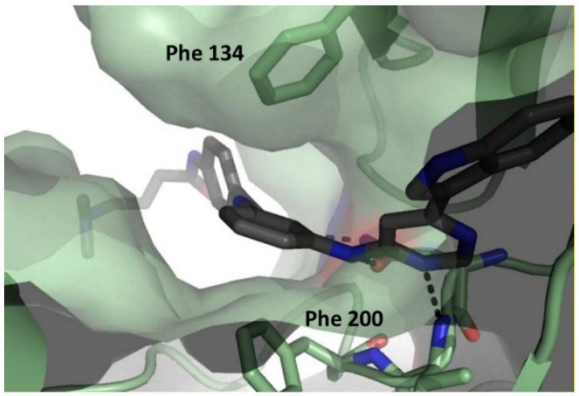
Supporting Figure 4. HEK 293T cells were treated with 10 μM compound or DMSO for 1 h at 37 °C.



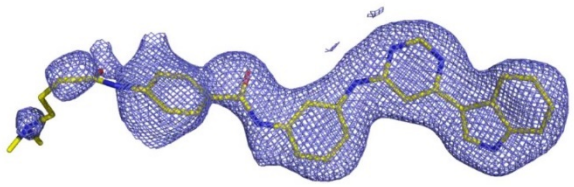
Supporting Figure 5. A. Representative western blot of isothermal CETSA for PI5P4Kα at 52 °C (HEK 293T cells) from three independent experiments. Dose-response curves were generated and are shown in Figure 4D. **B.** Representative western blot of isothermal CETSA for PI5P4Kβ at 52 °C (HEK 293T cells) from three independent experiments. Dose-response curves were generated and are shown in Figure 4E.



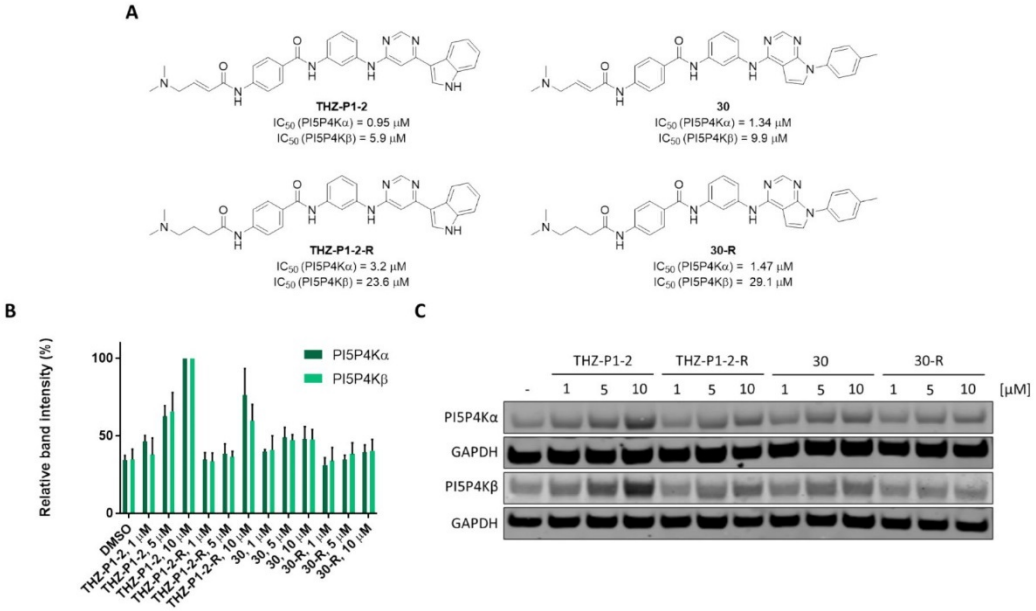
Supporting Figure 6. Mass spectra (left) and zero-charge mass spectra (right) obtained from recombinant PI5P4Kα (A, B), PI5P4Kβ (C, D) and PI5P4Kγ (E, F) treated with DMSO (A, C, E) or a 10-fold molar excess of a131 (B, D, F) for 2 hours at room temperature.



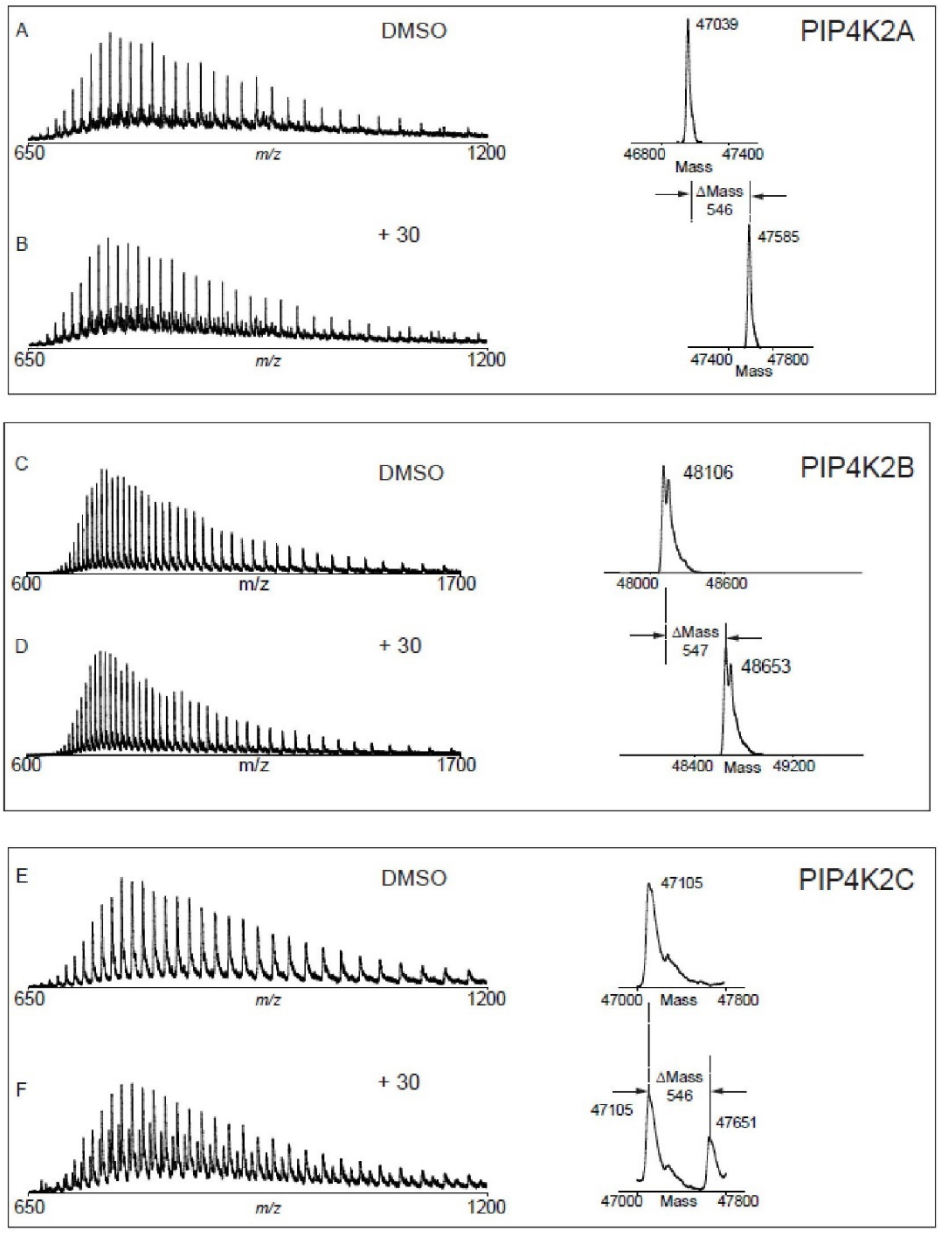
Supporting Figure 7. Phe134 and Phe200 form a narrow passage.[1]



Supporting Figure 8. Electron-density map for the ligand THZ-P1-2 in the cocrystal structure obtained for PI5P4Kα with THZ-P1-2.[1]



Supporting Figure 9. A. Chemical structures of THZ-P1-2 and compound **30** with their respective reversible analogs. **B.** Quantified relative band intensity (%) of isothermal CETSA experiment (Supporting Figure 8C). Mean of three independent experiments. **C.** Representative western blot of isothermal CETSA at 52 °C (HEK 293T cells) with THZ-P1-2, compound **30** and their reversible analogs at 1, 5, and 10 μM.



Supporting Figure 10. Mass spectra (left) and zero-charge mass spectra (right) obtained from recombinant PI5P4Kα (A, B), PI5P4Kβ (C, D) and PI5P4Kγ (E, F) treated with DMSO (A, C, E) or a 10-fold molar excess of compound **30** (B, D, F) for 2 hours at room temperature.

2. Experimental Procedures

2.1 Synthesis

2.1.1 General Methods

Reagents and solvent were purchased from commercial vendors and used without further purification, unless otherwise noted. Mass spectra were obtained on a Waters Acquity UPLC. Final purification steps were performed by preparative HPLC, using a Waters Sunfire C18 column (19 mm x 50 mm, 5 µM) and a gradient of 15-95% MeOH in water (0.05% TFA) over 22 min (28 min run time total) at a flow rate of 20 mL/min. Final compounds were isolated and tested as either TFA-salts or their free base. Their purities, determined by reverse-phase HPLC analysis, were greater than 95%, unless otherwise noted. Other purifications were performed, using a Teledyne ISCO CombiFlash R*f* system. ^{1}H NMR spectra were recorded on a 500 MHz Bruker Avance III spectrometer. Chemical shifts are reported in ppm (δ) downfield from TMS, coupling constants (*J*) are reported in Hz and spin multiplicities are described as s (singlet), d (doublet), t (triplet), q (quartet), dt (doublet of triplet) or m (multiplet).

General Procedure A

4,6-Dichloropyrimidine (1-2 eq.), boronic acid or pinacol boronic acid ester (1 eq.), $NaHCO_3$ (1 eq.) and $Pd(PPh_3)_2Cl_2$ (0.05 eq.) were dissolved in ACN/water (4:1, *v/v*). The reaction mixture was heated to 90-100 °C for 2.5-19 h under N_2. After cooling to room temperature, the mixture was diluted with EtOAc and the organic layer was washed with water and brine. The organic layer was further dried over Na_2SO_4 and purified by flash chromatography.

General Procedure B

Substituted 6-Chloropyrimidine (1 eq.) and phenylenediamine (5 eq.) were dissolved in NMP. DIEA (5-10 eq.) was added and the reaction mixture was stirred at 150-160 °C for 1.5 h - 2 d. The resulting mixture was allowed to cool to room temperature, diluted with EtOAc and the organic layer was washed with water and brine. The organic layer was further dried over Na_2SO_4 and purified by flash chromatography.

General Procedure C

Aniline (1 eq.) and 4-nitrobenzoyl chloride (1.2 eq.) were dissolved in pyridine and the reaction mixture was stirred at room temperature for 5 h. The solvent was removed *in vacuo* and it was continued without further purification.

General Procedure D

Aniline (1 eq.) and 4-nitrobenzoyl chloride (1.2 eq.) were dissolved in DCM. TEA (1.2 eq.) was added and the reaction mixture was stirred at room temperature for 1-5 h.

General Procedure E

Nitrobenzamide (1 eq.) and $SnCl_2 \cdot 2H_2O$ (10 eq.) were suspended in EtOAc/MeOH (5:2, *v/v*) and the reaction mixture was stirred at 80 °C for 4.5 h - 2 d. The reaction was quenched with saturated $NaHCO_3$-solution and the aqueous layer was extracted with DCM. The organic layer was then washed with water and brine, dried over Na_2SO_4 and purified by flash chromatography.

General Procedure F

Nitrobenzamide (1 eq.) and Pd/C (10%) were suspended in EtOAc/MeOH (1:1, *v/v*) or just MeOH and the

reaction mixture was stirred at room temperature for 16 h under H_2. The resulting mixture was filter over a pad of CELITE and purified by flash chromatography.

General Procedure G

Aminobenzamide (1 eq.) was dissolved in ACN or DCM and DIEA was added. A solution of 4-bromocrotonyl chloride in DCM was added dropwise to the reaction mixture at 0 °C until completion of reaction. The mixture was either purified by preparative HPLC or it was continued without further purification.

DMA in THF (2.0 M) was added to the mixture or purified intermediate and the reaction was stirred at room temperature for 1 h. The mixture was purified by preparative HPLC or flash chromatography.

General Procedure H

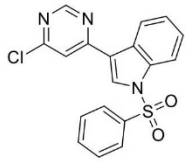
4-Dimethylaminocrotonic acid (1.2 eq.) and aminobenzamide (1 eq.) were dissolved in DMF or DCM. DIEA (3.5 eq.) and HATU (1.2 eq.) were added and the reaction mixture was stirred at room temperature for 12 h. The mixture was diluted with water and the aqueous layer was extracted with DCM. The organic layer was further washed with water and brine, dried over Na_2SO_4 and purified by flash chromatography.

General Procedure J

Protected indole or azaindole (1 eq.) was dissolved in 1.0 M NaOH/1,4-dioxane (1:1, *v/v*) and the reaction mixture was stirred at room temperature for 4-6 h. The mixture was purified by preparative HPLC.

2.1.2 Synthesis of THZ-P1-2 and compounds 6-11

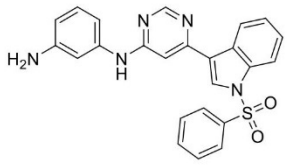
3-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-indole



2a

Compound **2a** was synthesized from 4,6-dichloropyrimidine (989 mg, 6.64 mmol) and (1-(phenylsulfonyl)-1*H*-indol-3-yl)boronic acid (1.0 g, 3.32 mmol) following General Procedure A. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (750 mg, 2.03 mmol, 61%) as an ochre solid. MS (ESI) *m/z* calc. for $C_{18}H_{12}ClN_3O_2S$ $[M+H]^+$ 370.04; found 370.05.

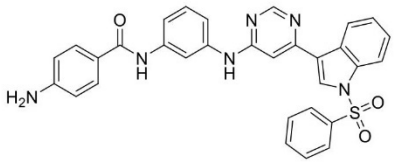
N^1-(6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)benzene-1,3-diamine



3a

m-phenylenediamine (0.82 g, 2.0 mmol) and 3-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-indole (**2a**) (1.4 g, 3.78 mmol) were used as described in General Procedure B to synthesize compound **3a**. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (1.6 g, 3.6 mmol, 95%) as an ochre solid. MS (ESI) *m/z* calc. for $C_{24}H_{19}N_5O_2S$ $[M+H]^+$ 442.13; found 441.97.

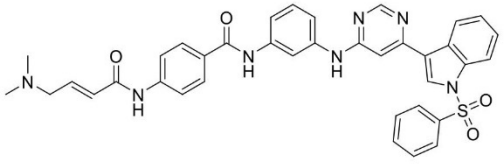
4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide



4a

N^1-(6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3a**) (0.25 g, 0.57 mmol) and 4-nitrobenzoyl chloride (137 mg, 0.74 mmol) were reacted as described in General Procedure C. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**4a**), following General Procedure E. The reaction mixture was purified by flash chromatography (1.75 M NH_3 in MeOH:DCM) to yield the product (177 mg, 0.32 mmol, 56% over 2 steps) as a white solid. MS (ESI) *m/z* calc. for $C_{31}H_{24}N_6O_3S$ $[M+H]^+$ 561.17; found 560.89.

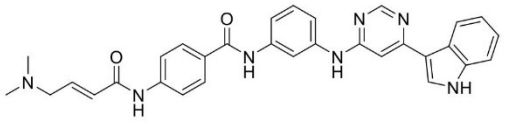
(*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide



5a

Compound **5a** was synthesized from 4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**4a**) (1.4 g, 2.5 mmol) following General Procedure G. The reaction mixture was diluted with water and the aqueous layer was extracted with DCM. The combined organic layers were washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (0.4 mg, 0.6 mmol, 24%).

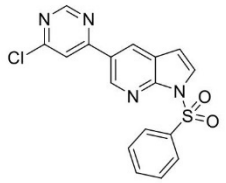
(*E*)-*N*-(3-((6-(1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



THZ-P1-2

In a final step, (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**5a**) (0.4 g, 0.6 mmol) was deprotected according to General Procedure J. The reaction mixture was extracted with chloroform/2-propanol (4:1, *v*/*v*) and the organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was purified by flash chromatography (MeOH:DCM) to yield the title compound (**THZ-P1-2**) (0.19 mg, 0.36 mmol, 60%). MS (ESI) *m/z* calc. for $C_{31}H_{29}N_7O_2$ $[M+H]^+$ 532.25; found 532.45. ^1H NMR (400 MHz, DMSO-*d6*) δ 11.73 (s, 1H), 10.50 (s, 1H), 10.19 (s, 1H), 9.55 (s, 1H), 8.63 (s, 1H), 8.27 (d, *J* = 8.8 Hz, 1H), 8.20 (s, 1H), 8.15 (d, *J* = 2.8 Hz, 1H), 8.00 (d, *J* = 8.7 Hz, 2H), 7.83 (d, *J* = 8.8 Hz, 2H), 7.59-7.44 (m, 2H), 7.37 (d, *J* = 8.3 Hz, 1H), 7.34-7.26 (m, 2H), 7.24-7.09 (m, 2H), 6.81 (dt, *J* = 15.4, 6.2 Hz, 1H), 6.40 (d, *J* = 15.4 Hz, 1H), 3.37 (d, *J* = 5.3 Hz, 2H), 2.39 (s, 6H).

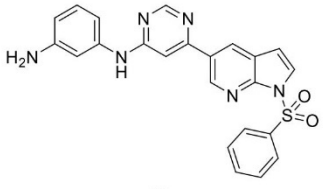
5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine



2b

Compound **2b** was synthesized from 4,6-dichloropyrimidine (0.38 g, 2.6 mmol) and 1-(phenylsulfonyl)-5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (1.0 g, 2.6 mmol) following General Procedure A. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (0.8 g, 2.2 mmol, 83%) as a slightly yellow solid. MS (ESI) *m/z* calc. for $C_{17}H_{11}ClN_4O_2S$ $[M+H]^+$ 371.04; found 371.81.

N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine

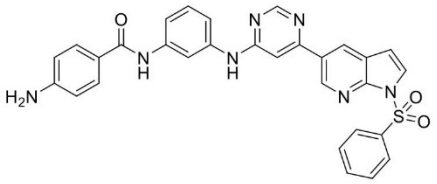


3b

m-phenylenediamine (1.1 g, 10.0 mmol) and 5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine (**2b**) (1.85 g, 5.0 mmol) were used as described in General Procedure B to

synthesize compound **3b**. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (1.0 g, 2.3 mmol, 46%) as a slightly yellow solid.

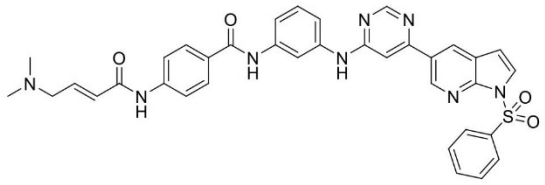
4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



4b

N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3b**) (0.3 g, 0.67 mmol) and 4-nitrobenzoyl chloride (151 mg, 0.81 mmol) were reacted as described in General Procedure D. The reaction mixture was diluted with DCM and the organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield 4-nitro-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.3 g, 0.51 mmol, 75%) as a slightly yellow solid. MS (ESI) *m/z* calc. for $C_{30}H_{21}N_7O_5S$ $[M+H]^+$ 592.14; found 592.60.
The intermediate nitrobenzamide (0.3 g, 0.50 mmol) was reduced to the respective aromatic amine (**4b**) according to General Procedure F. The reaction mixture was filtered through a pad of CELITE and the filtrate was concentrated in vacuo to yield the product (0.24 g, 0.43 mmol, 84%) as a slightly yellow solid. MS (ESI) *m/z* calc. for $C_{30}H_{23}N_7O_3S$ $[M+H]^+$ 562.17; found 562.60.

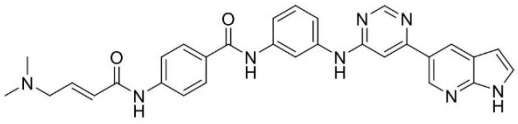
(*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



5b

Compound **5b** was synthesized from 4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**4b**) (0.24 g, 0.42 mmol) following General Procedure G. The crude was purified by flash chromatography to yield the product (0.2 g, 0.3 mmol, 71%).

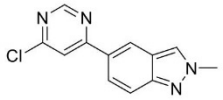
(*E*)-*N*-(3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



6

In a final step, (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**5b**) (0.10 g, 0.14 mmol) was deprotected according to General Procedure J. The pH of the reaction mixture was adjusted to 7 by addition of 1.0 M HCl. The solvent was removed *in vacuo* and the crude was by flash chromatography to yield the title compound (**6**) (6 mg, 0.011 mmol, 8%). MS (ESI) *m/z* calc. for $C_{30}H_{28}N_8O_2$ $[M+H]^+$ 533.24; found 533.60. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 11.89 (s, 1H), 10.67 (s, 1H), 10.23 (s, 1H), 9.75 (s, 1H), 8.91 (s, 1H), 8.73 (s, 1H), 8.61 (s, 1H), 8.20 (s, 1H), 8.01 (d, *J* = 8.8 Hz, 2H), 7.83 (d, *J* = 7.8 Hz, 2H), 7.57 (s, 1H), 7.51 (s, 1H), 7.44-7.37 (m, 2H), 7.32 (t, *J* = 8.1 Hz), 6.80 (dt, *J* = 14.7, 7.2 Hz, 1H), 6.59 (s, 1H), 6.52 (d, *J* = 15.3 Hz, 1H), 3.19 (s, 2H), 2.81 (s, 6H).

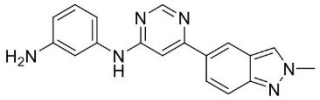
5-(6-chloropyrimidin-4-yl)-2-methyl-2*H*-indazole



2c

Compound **2c** was synthesized from 4,6-dichloropyrimidine (0.7 g, 4.65 mmol) and 2-methyl-5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-2*H*-indazole (1.2 g, 4.7 mmol) following General Procedure A. The crude was purified by flash chromatography (PE:EtOAc) to yield the product (0.8 g, 3.2 mmol, 70%).

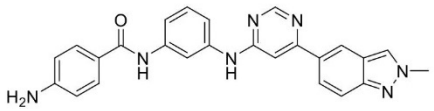
N^1-(6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)benzene-1,3-diamine



3c

m-phenylenediamine (0.7 g, 4.9 mmol) and 5-(6-chloropyrimidin-4-yl)-2-methyl-2*H*-indazole (**2c**) (0.8 g, 3.27 mmol) were used as described in General Procedure B to synthesize compound **3c**. After aqueous work-up, the solvent was removed *in vacuo* to yield the product (0.5 g, 1.6 mmol, 50%).

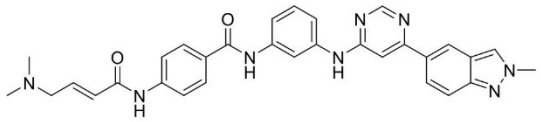
4-amino-*N*-(3-((6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



4c

N^1-(6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3c**) (0.5 g, 1.58 mmol) and 4-nitrobenzoyl chloride (0.45 g, 2.3 mmol) were reacted as described in General Procedure D. The reaction mixture was filtered, and the filtrate was washed with DCM to yield *N*-(3-((6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)amino)phenyl)-4-nitrobenzamide (0.6 g, 100% yield = 0.7 g).
The crude intermediate nitrobenzamide (0.6 g) was reduced to the respective aromatic amine (**4c**) according to General Procedure F to yield the product (0.6 g, 1.38 mmol, 87% over 2 steps).

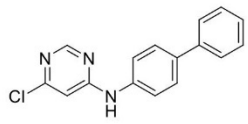
(*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



7

4-amino-*N*-(3-((6-(2-methyl-2*H*-indazol-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**4c**) (0.6 g, 1.76 mmol) was coupled with 4-dimethylaminocrotonic acid (0.35 g, 2.11 mmol) following General Procedure H. The crude was purified by flash chromatography to yield title compound (**7**) (0.2 g, 0.37 mmol, 30%). MS (ESI) *m/z* calc. for $C_{31}H_{30}N_8O_2$ $[M+H]^+$ 547.26; found 547.41. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 10.33 (s, 1H), 10.17 (s, 1H), 9.66 (s, 1H), 8.70 (d, *J* = 1.0 Hz, 1H), 8.51 (s, 2H), 8.17 (t, *J* = 2.0 Hz, 1H), 8.03-7.94 (m, 2H), 7.90 (dd, *J* = 9.1, 1.7 Hz, 1H), 7.83-7.76 (m, 2H), 7.71 (dd, *J* = 5.4, 4.6 Hz, 1H), 7.53 (d, *J* = 8.0 Hz, 1H), 7.40 (dd, *J* = 7.1, 1.9 Hz, 1H), 7.35-7.24 (m, 2H), 6.79 (dt, *J* = 15.4, 5.8 Hz, 1H), 6.31 (dt, *J* = 15.4, 1.6 Hz, 1H), 4.20 (s, 3H), 3.08 (dd, *J* = 5.9, 1.4 Hz, 2H), 2.19 (s, 6H).

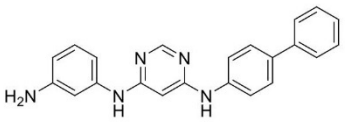
***N*-([1,1'-biphenyl]-4-yl)-6-chloropyrimidin-4-amine**



2d

Compound **2d** was synthesized from 4,6-dichloropyrimidine (723 mg, 4.85 mmol) and [1,1'-biphenyl]-4-amine (547 mg, 3.23 mmol) were dissolved in EtOH (5.0 mL). TEA (0.9 mL, 6.46 mmol) was added and the reaction mixture was stirred at 120 °C for 3 h. The mixture was poured into ice water. The resulting suspension was filtered to yield the product (899 mg, 3.19 mmol, 99%) as a brown-yellow solid. MS (ESI) *m/z* calc. for $C_{16}H_{12}ClN_3$ $[M+H]^+$ 282.08; found 282.17.

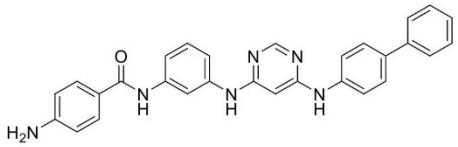
N^4-([1,1'-biphenyl]-4-yl)-N^6-(3-aminophenyl)pyrimidine-4,6-diamine



3d

m-phenylenediamine (664 mg, 6.14 mmol) and *N*-([1,1'-biphenyl]-4-yl)-6-chloropyrimidin-4-amine (**2d**) (346 g, 1.23 mmol) were used as described in General Procedure B to synthesize compound **3d**. For full conversion, an additional 300 mg *m*-phenylenediamine were added to the mixture. The crude was purified by flash chromatography (1.75 M NH_3 in MeOH:DCM) to yield the product (0.16 g, 0.45 mmol, 37%) as a brown solid. MS (ESI) *m/z* calc. for $C_{22}H_{19}N_5$ $[M+H]^+$ 354.17; found 353.96.

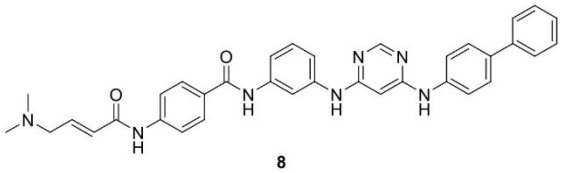
N-(3-((6-([1,1'-biphenyl]-4-ylamino)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide



4d

N^4-([1,1'-biphenyl]-4-yl)-N^6-(3-aminophenyl)pyrimidine-4,6-diamine (**3d**) (0.24 g, 0.68 mmol) and 4-nitrobenzoyl chloride (151 mg, 0.82 mmol) were reacted as described in General Procedure C. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**4d**), following General Procedure E. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield the product (232 mg, 0.49 mmol, 69% over 2 steps) as a brown solid. MS (ESI) *m/z* calc. for $C_{29}H_{24}N_6O$ $[M+H]^+$ 473.20; found 473.05.

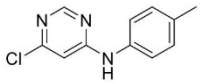
(*E*)-*N*-(3-((6-([1,1'-biphenyl]-4-ylamino)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



8

Compound **8** was synthesized from *N*-(3-((6-([1,1'-biphenyl]-4-ylamino)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide (**4d**) (0.1 g, 0.22 mmol) following General Procedure G. The crude was purified by reverse-phase HPLC to yield the product (TFA-salt, 3 mg, 0.004 mmol, 4%) as a white solid. MS (ESI) *m/z* calc. for $C_{35}H_{33}N_7O_2$ $[M+H]^+$ 584.28; found 584.45. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 10.67 (s, 1H), 10.23 (s, 1H), 9.87 (s, 1H), 9.45 (s, 1H), 9.40 (s, 1H), 8.39 (s, 1H), 8.11 (d, *J* = 2.2 Hz, 1H), 8.08-8.02 (m, 2H), 7.91-7.85 (m, 2H), 7.74-7.63 (m, 6H), 7.50 (t, *J* = 7.7 Hz, 2H), 7.45 (dd, *J* = 7.5, 2.0 Hz, 1H), 7.41-7.31 (m, 3H), 6.85 (dt, *J* = 14.9, 7.2 Hz, 1H), 6.56 (d, *J* = 15.3 Hz, 1H), 6.32 (s, 1H), 4.03 (d, *J* = 6.9 Hz, 2H), 2.88 (d, *J* = 3.4 Hz, 6H).

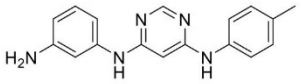
6-chloro-*N*-(*p*-tolyl)pyrimidin-4-amine



2e

To a solution of 4,6-dichloropyrimidine (1.39 g, 9.33 mmol) in BuOH (9.33 mL) was added *p*-toluidine (1.0 g, 9.33 mmol) and TEA (1.951 mL, 14.0 mmol). The reaction was stirred at 140 °C for 1 h, affording the initial adduct very cleanly. It was continued without further work-up.

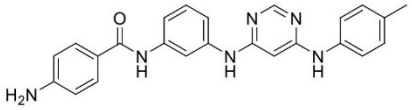
N^4-(3-aminophenyl)-N^6-(*p*-tolyl)pyrimidine-4,6-diamine



3e

To the mixture of compound **2e** was added *N*-(3-aminophenyl)acetamide (1.402 g, 9.33 mmol) and then concentrated HCl (5 mL) at which point (pH around 2.0) a solid precipitated. The reaction mixture was stirred at 140 °C for 4 h. After cooling to room temperature, the mixture was diluted with EtOAc and the organic layer was washed with $NaHCO_3$ and brine, dried over $MgSO_4$ and the solvent was removed *in vacuo*. The crude was partially dissolved in some EtOAc and triturated with *n*-hexanes and DCM to afford the product (**3e**) (1.05 g, 3.6 mmol, 39%) as a tan solid. MS (ESI) *m/z* calc. for $C_{17}H_{17}N_5$ $[M+H]^+$ 292.16; found 292.1.

4-amino-*N*-(3-((6-(*p*-tolylamino)pyrimidin-4-yl)amino)phenyl)benzamide

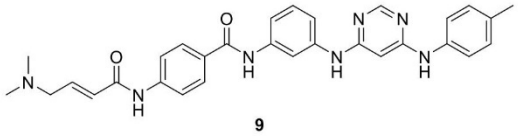


4e

A mixture of compound **3e** (1.05 g, 3.6 mmol) and 4-((*tert*-butoxycarbonyl)amino)benzoic acid (1.026 g, 4.32 mmol) in DMF (7.21 mL) was treated with HATU (1.781 g, 4.69 mmol) and Hunig's Base (1.888 mL, 10.81 mmol) at 23 °C and was stirred at this temperature for 1 h. The solution was concentrated over a stream of air.

The crude was treated with TFA/DCM (1:1, *v/v*) (10 mL) and stirred at 23 °C for 2 h. The solution was concentrated over a stream of air. The crude mixture was diluted with DMSO and purified by reverse-phase HPLC. The TFA-salt was neutralized with saturated $NaHCO_3$ (to pH 7-8) and the aqueous mixture was extracted with EtOAc. The organic layer was washed with $NaHCO_3$ and brine, dried over $MgSO_4$ and the solvent was removed *in vacuo* to yield the product (**4e**) (550 mg, 1.34 mmol, 37%). MS (ESI) *m/z* calc. for $C_{24}H_{22}N_6O$ $[M+H]^+$ 411.19; found 411.2.

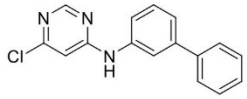
(*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(*p*-tolylamino)pyrimidin-4-yl)amino)phenyl)benzamide



9

To (*E*)-4-(dimethylamino)but-2-enoic acid, HCl (353 mg, 2.132 mmol) in THF (2.13 mL) was added oxalylic chloride (0.187 mL, 2.13 mmol). A drop of DMF was added and the reaction mixture was stirred at 23 °C for 1 h. The mixture was treated dropwise with a solution of 4-amino-*N*-(3-((6-(*p*-tolylamino)pyrimidin-4-yl)amino)phenyl)benzamide (**4e**) (350 mg, 0.853 mmol) in NMP (2.13 mL) and stirred at 23 °C for 1 h. The solvent was removed *in vacuo* and the crude was purified by reverse-phase HPLC. The TFA-salt was dissolved in EtOAc and the organic layer was washed with saturated $NaHCO_3$-solution and brine, dried over $MgSO_4$ and the solvent was removed *in vacuo* to yield the title compound (**9**) (28 mg, 0.054 mmol, 6%) as a light yellow solid. MS (ESI) *m/z* calc. for $C_{30}H_{31}N_7O_2$ $[M+H]^+$ 522.26; found 522.3. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 10.29 (s, 1H), 10.07 (s, 1H), 9.10 (s, 1H), 8.97 (s, 1H), 8.21 (s, 1H), 8.00-7.86 (m, 3H), 7.76 (d, *J* = 8.8 Hz, 2H), 7.37 (d, *J* = 8.2 Hz, 2H), 7.31 (t, *J* = 8.5 Hz, 2H), 7.22 (d, *J* = 8.0 Hz, 1H), 7.06 (d, *J* = 8.2 Hz, 2H), 6.75 (dt, *J* = 15.4, 5.8 Hz, 1H), 6.27 (d, *J* = 15.5 Hz, 1H), 6.13 (s, 1H), 3.04 (d, *J* = 6.2 Hz, 2H), 2.21 (s, 3H), 2.16 (s, 6H).

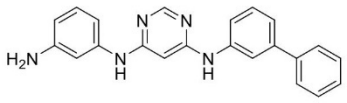
***N*-([1,1'-biphenyl]-3-yl)-6-chloropyrimidin-4-amine**



2f

Compound **2f** was synthesized from 4,6-dichloropyrimidine (500 mg, 3.36 mmol) and [1,1'-biphenyl]-3-amine (379 mg, 2.24 mmol), which were dissolved in EtOH (5.0 mL). TEA (0.62 mL, 4.48 mmol) was added and the reaction mixture was stirred at 120 °C for 4 h. The mixture was poured into ice water. The resulting suspension was filtered. The filtrate was re-dissolved in acetone and the solvent was removed *in vacuo* to yield the crude product (893 mg, 100% yield = 631 mg) as a dark brown oil. It was continued without further purification. MS (ESI) *m/z* calc. for $C_{16}H_{12}ClN_3$ $[M+H]^+$ 282.08; found 282.17.

***N*^{4}-([1,1'-biphenyl]-3-yl)-*N*^{6}-(3-aminophenyl)pyrimidine-4,6-diamine**

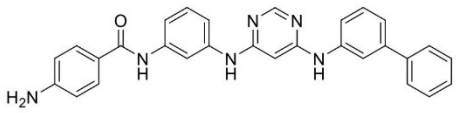


3f

m-phenylenediamine (1.15 g, 10.7 mmol) and *N*-([1,1'-biphenyl]-3-yl)-6-chloropyrimidin-4-amine (**2f**) (0.6 g, 2.13 mmol) were used as described in General Procedure B to synthesize compound **3f**. The mixture was purified by flash chromatography (1.75 M NH_3 in MeOH:DCM) to yield the crude product

(1.6 g, 100% yield = 0.75 g) as an ochre solid. MS (ESI) *m/z* calc. for $C_{22}H_{19}N_5$ $[M+H]^+$ 354.17; found 354.24.

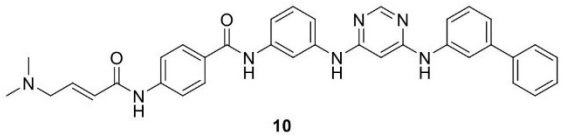
N-(3-((6-([1,1'-biphenyl]-3-ylamino)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide



4f

N^4-([1,1'-biphenyl]-3-yl)-N^6-(3-aminophenyl)pyrimidine-4,6-diamine (**3f**) (0.2 g, 0.57 mmol) and 4-nitrobenzoyl chloride (126 mg, 0.68 mmol) were reacted as described in General Procedure C. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**4f**), following General Procedure E. The reaction mixture was purified by flash chromatography (1.75 M NH_3 in MeOH:DCM) to yield the product (58 mg, 0.12 mmol, 22% over 2 steps) as an orange solid. MS (ESI) *m/z* calc. for $C_{29}H_{24}N_6O$ $[M+H]^+$ 473.21; found 473.38.

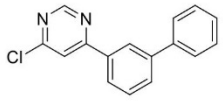
(*E*)-*N*-(3-((6-([1,1'-biphenyl]-3-ylamino)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



10

Compound **10** was synthesized from *N*-(3-((6-([1,1'-biphenyl]-3-ylamino)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide (**4f**) (58 mg, 0.12 mmol) following General Procedure G. The crude was purified by reverse-phase HPLC to yield the product (TFA-salt, 26 mg, 0.037 mmol, 69%) as a white solid. MS (ESI) *m/z* calc. for $C_{35}H_{33}N_7O_2$ $[M+H]^+$ 584.28; found 584.42. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 10.60 (s, 1H), 10.18 (s, 1H), 9.82 (s, 1H), 9.51 (s, 1H), 9.48 (s, 1H), 8.34 (s, 1H), 8.02-7.91 (m, 3H), 7.84-7.73 (m, 4H), 7.64 (d, *J* = 7.3 Hz, 2H), 7.54 (d, *J* = 8.2 Hz, 1H), 7.48 (t, *J* = 7.6 Hz, 2H), 7.44-7.35 (m, 2H), 7.34-7.23 (m, 4H), 6.78 (dt, *J* = 14.8, 7.2 Hz, 1H), 6.49 (d, *J* = 15.3 Hz, 1H), 6.25 (s, 1H), 3.96 (s, 2H), 2.81 (s, 6H).

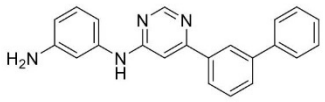
4-([1,1'-biphenyl]-3-yl)-6-chloropyrimidine



2g

Compound **2g** was synthesized from 4,6-dichloropyrimidine (1.5 g, 10.1 mmol) and [1,1'-biphenyl]-3-ylboronic acid (1.0 g, 5.1 mmol) following General Procedure A. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (0.708 g, 2.7 mmol, 53%) as an off-white solid. MS (ESI) *m/z* calc. for $C_{16}H_{11}ClN_2$ $[M+H]^+$ 267.07; found 267.04.

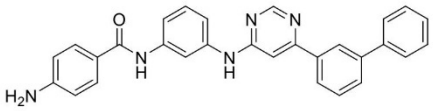
N^1-(6-([1,1'-biphenyl]-3-yl)pyrimidin-4-yl)benzene-1,3-diamine



3g

m-phenylenediamine (101 mg, 0.94 mmol), 4-([1,1'-biphenyl]-3-yl)-6-chloropyrimidine (**2g**) (250 mg, 0.94 mmol) and DIEA (170 µL, 0.94 mmol) were dissolved in NMP (5.0 mL) and stirred at 150 °C for 2 h. The reaction mixture was worked up as further described in General Procedure B. The mixture was purified by flash chromatography (MeOH:DCM) to yield compound **3g** (153 mg, 0.45 mmol, 48%) as a brown solid. MS (ESI) *m/z* calc. for $C_{22}H_{18}N_4$ $[M+H]^+$ 339.16; found 338.98.

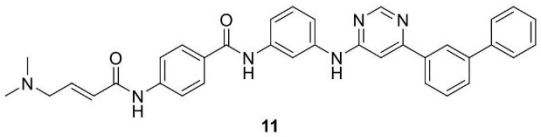
N-(3-((6-([1,1'-biphenyl]-3-yl)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide



4g

N^1-(6-([1,1'-biphenyl]-3-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3g**) (0.37 g, 1.1 mmol) and 4-nitrobenzoyl chloride (241 mg, 1.3 mmol) were reacted as described in General Procedure C. More 4-nitrobenzoyl chloride was added in small portions until the reaction was completed. The solvent was removed *in vacuo*. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**4g**), following General Procedure E. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield the product (255 mg, 0.56 mmol, 51% over 2 steps) as a brown solid. MS (ESI) *m/z* calc. for $C_{29}H_{23}N_5O$ $[M+H]^+$ 458.20; found 458.05.

(*E*)-*N*-(3-((6-([1,1'-biphenyl]-3-yl)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide

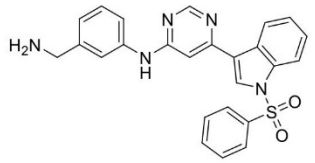


11

Compound **11** was synthesized from *N*-(3-((6-([1,1'-biphenyl]-3-yl)pyrimidin-4-yl)amino)phenyl)-4-aminobenzamide (**4g**) (0.1 g, 0.22 mmol) following General Procedure G. The crude was purified by reverse-phase HPLC to yield the product (TFA-salt, 8 mg, 0.011 mmol, 7%) as a white solid. MS (ESI) *m/z* calc. for $C_{35}H_{32}N_6O_2$ $[M+H]^+$ 569.27; found 568.98. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 10.60 (s, 1H), 10.22 (s, 1H), 9.83 (s, 1H), 9.75 (s, 1H), 8.78 (s, 1H), 8.31 (t, *J* = 1.9 Hz, 1H), 8.24 (t, *J* = 2.1 Hz, 1H), 8.06 (d, *J* = 7.8 Hz, 1H), 8.02 – 7.96 (m, 2H), 7.83 (dd, *J* = 8.8, 7.0 Hz, 3H), 7.76 (dd, *J* = 7.3, 1.6 Hz, 2H), 7.66 (t, *J* = 7.7 Hz, 1H), 7.53-7.47 (m, 3H), 7.46-7.38 (m, 3H), 7.33 (t, *J* = 8.0 Hz, 1H), 6.80 (dt, *J* = 14.8, 7.2 Hz, 1H), 6.50 (d, *J* = 15.3 Hz, 1H), 3.98 (t, *J* = 5.6 Hz, 2H), 2.83 (d, *J* = 4.2 Hz, 6H).

2.1.3 Synthesis of compounds 14-19

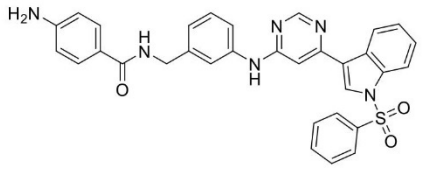
_N_-(3-(aminomethyl)phenyl)-6-(1-(phenylsulfonyl)-1_H_-indol-3-yl)pyrimidin-4-amine



12a

tert-butyl (3-aminobenzyl)carbamate (0.48 g, 2.17 mmol) and 3-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-indole (**2a**) (0.20 g, 0.54 mmol) were dissolved in NMP (2 mL) and DIEA (0.38 µL, 2.17 mmol) was added. The reaction mixture was stirred at 150 °C for 10 h. The reaction was allowed to cool to room temperature and was diluted with EtOAc. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*.
The crude was redissolved in DCM (2 mL) and TFA (2 mL) was added. The mixture was stirred at room temperature for 1 h. The solvent was removed *in vacuo* and the crude was further purified by flash chromatography (MeOH:DCM) to yield the product (**12a**) (0.21 g, 0.46 mmol, 84% over 2 steps) as a dark-brown solid. MS (ESI) *m/z* calc. for $C_{25}H_{21}N_5O_2S$ $[M+H]^+$ 456.15; found 456.00.

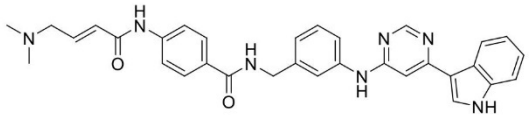
4-amino-_N_-(3-((6-(1-(phenylsulfonyl)-1_H_-indol-3-yl)pyrimidin-4-yl)amino)benzyl)benzamide



13a

N-(3-(aminomethyl)phenyl)-6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-amine (**12a**) (0.21 g, 0.46 mmol) and 4-nitrobenzoyl chloride (102 mg, 0.55 mmol) were reacted as described in General Procedure C. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**13a**), following General Procedure E. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield the product (61 mg, 0.11 mmol, 23% over 2 steps) as an ochre solid. MS (ESI) *m/z* calc. for $C_{32}H_{26}N_6O_3S$ $[M+H]^+$ 575.19; found 574.92.

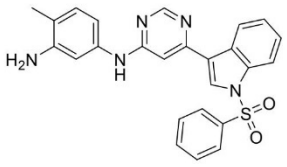
(*E*)-*N*-(3-((6-(1*H*-indol-3-yl)pyrimidin-4-yl)amino)benzyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



14

Compound **14** was synthesized from 4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)benzyl)benzamide (**13a**) (61 mg, 0.11 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by reversed-phase HPLC to yield (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)benzyl)benzamide (66 mg, 0.096 mmol, 91% over 2 steps) as a yellow solid.
The intermediate was further deprotected according to General Procedure J. The reaction mixture was quenched with 1.0 M HCl, the solvent was removed *in vacuo* and the crude was purified by reverse-phase HPLC and flash chromatography (MeOH:DCM) to yield the product (7 mg, 0.013 mmol, 14%) as an off-white solid. MS (ESI) *m/z* calc. for $C_{32}H_{31}N_7O_2$ $[M+H]^+$ 546.26; found 545.98. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 11.70 (s, 1H), 10.42 (s, 1H), 9.48 (s, 1H), 8.98 (s, 1H), 8.56 (s, 1H), 8.22 (d, *J* = 7.3 Hz, 1H), 8.11 (d, *J* = 3.1 Hz, 1H), 7.91 (d, *J* = 8.9 Hz, 2H), 7.76 (d, *J* = 8.9 Hz, 2H), 7.68 (d, *J* = 7.9 Hz, 1H), 7.61 (s, 1H), 7.48 (d, *J* = 7.0 Hz, 1H), 7.28 (t, *J* = 7.9 Hz, 1H), 7.24-7.14 (m, 3H), 6.96 (d, *J* = 7.6 Hz, 1H), 6.83-6.73 (m, 1H), 6.38 (d, *J* = 15.3 Hz, 1H), 4.49 (d, *J* = 5.8 Hz, 2H), 3.44 (s, 2H), 2.43 (s, 6H).

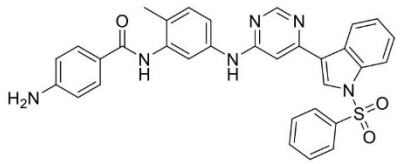
4-methyl-N^1-(6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)benzene-1,3-diamine



12b

tert-butyl (5-amino-2-methylphenyl)carbamate (0.5 g, 2.25 mmol) and 3-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-indole (**2a**) (0.21 g, 0.56 mmol) were dissolved in NMP (2 mL) and DIEA (0.39 µL, 2.25 mmol) was added. The reaction mixture was stirred at 150 °C for 6 h. The reaction was allowed to cool to room temperature and was diluted with EtOAc. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*.
The crude was redissolved in DCM (2 mL) and TFA (2 mL) was added. The mixture was stirred at room temperature for 2 h. The solvent was removed *in vacuo* and the crude was further purified by flash chromatography (MeOH:DCM) to yield the product (**12b**) (0.17 g, 0.37 mmol, 67% over 2 steps) as a dark-brown oil. MS (ESI) *m/z* calc. for $C_{25}H_{21}N_5O_2S$ $[M+H]^+$ 456.15; found 455.97.

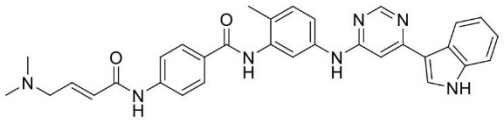
4-amino-*N*-(2-methyl-5-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide



13b

4-methyl-N^1-(6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)benzene-1,3-diamine (**12b**) (0.17 g, 0.38 mmol) and 4-nitrobenzoyl chloride (84 mg, 0.46 mmol) were reacted as described in General Procedure C. Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**13b**), following General Procedure E. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield the product (84 mg, 0.15 mmol, 39% over 2 steps) as a yellow-brown solid. MS (ESI) *m/z* calc. for $C_{32}H_{26}N_6O_3S$ $[M+H]^+$ 575.19; found 575.21.

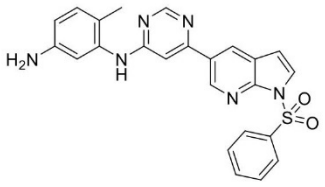
(*E*)-*N*-(5-((6-(1*H*-indol-3-yl)pyrimidin-4-yl)amino)-2-methylphenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



15

Compound **15** was synthesized from 4-amino-*N*-(2-methyl-5-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**13b**) (107 mg, 0.19 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by reversed-phase HPLC to yield (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(2-methyl-5-((6-(1-(phenylsulfonyl)-1*H*-indol-3-yl)pyrimidin-4-yl)amino)phenyl)benzamide.
The intermediate was further deprotected according to General Procedure J. The reaction mixture was quenched with 1.0 M HCl, the solvent was removed *in vacuo* and the crude was purified by reverse-phase HPLC to yield the product (TFA-salt, 4 mg, 0.006 mmol, 10%) as a white solid. MS (ESI) *m/z* calc. for $C_{32}H_{31}N_7O_2$ $[M+H]^+$ 546.26; found 546.02. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 12.14 (s, 1H), 10.63 (s, 1H), 10.47 (s, 1H), 9.98 (s, 1H), 9.85 (s, 1H), 8.78 (s, 1H), 8.30 (s, 1H), 8.07 (s, 1H), 8.02 (d, *J* = 8.5 Hz, 2H), 7.86-7.78 (m, 3H), 7.57 (d, *J* = 7.0 Hz, 1H), 7.49 (s, 1H), 7.35 (s, 1H), 7.34-7.24 (m, 3H), 6.79 (dd, *J* = 15.1, 7.5 Hz, 1H), 6.50 (d, *J* = 15.6 Hz, 1H), 3.97 (d, *J* = 7.0 Hz, 2H), 2.82 (s, 6H), 2.25 (s, 3H).

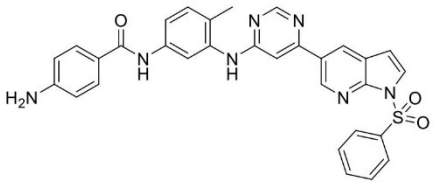
6-methyl-N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine



12c

2-methyl-5-nitroaniline (1.23 g, 8 mmol), 5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine (**2b**) (1.5 g, 4 mmol), Cs_2CO_3 (1.4 g, 4.4 mmol), $Pd_2(dba)_3$ (0.07 g, 0.08 mmol) and 2-dicyclohexylphosphino-2',4',6'-triisopropyl-1,1'-biphenyl (0.14 g, 0.24 mmol) were dissolved in 1,4-dioxane. The reaction mixture was stirred at 104 °C for 3 h under N_2. The reaction was allowed to cool to room temperature and was diluted with EtOAc. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography.
The intermediate (4.2 g, 8.63 mmol) redissolved in MeOH (200 mL) and $SnCl_2$ (16.4 g, 86.3 mmol) was added. The mixture was stirred at 80 °C for 8 h. The pH of the reaction mixture was adjusted to 7.0 using 1.0 M NaOH and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography to yield the product (**12c**) (2.4 g, 5.3 mmol, 61%) as a yellow solid.

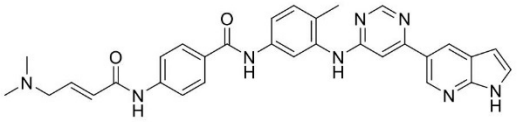
4-amino-*N*-(4-methyl-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



13c

6-methyl-N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine (**12c**) (0.91 g, 2.0 mmol), 4-((*tert*-butoxycarbonyl)amino)benzoic acid (0.71 g, 3.0 mmol) and HATU (1.14 g, 3.0 mmol) were dissolved in DMF (5.0 mL). DIEA (0.59 g, 4.0 mmol) was added and the mixture was stirred at room temperature for 30 min. The reaction mixture was diluted with EtOAc and the organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was purified by flash chromatography to yield *tert*-butyl (4-((4-methyl-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)carbamoyl)phenyl)carbamate (1.0 g, 1.48 mmol).
The intermediate (1.0 g, 1.48 mmol) was converted to the respective free amine (**13c**) *via* stirring in 4.0 M HCl in EtOAc at room temperature for 30 min. The reaction mixture was further diluted with EtOAc and the pH of the solution was adjusted to 7.0, using 1.0 M NaOH. The organic layer was then washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo* to yield the product (0.8 g, 1.39 mmol, 85%) as a white solid.

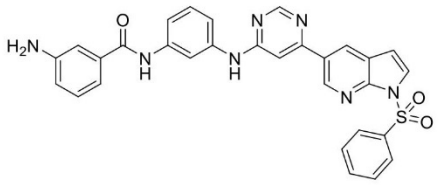
(*E*)-*N*-(3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)-4-methylphenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



16

Compound **16** was synthesized from 4-amino-*N*-(4-methyl-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**13c**) (0.4 g, 0.7 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(4-methyl-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.2 g, 0.29 mmol, 42%) as a slightly yellow solid.
The intermediate was further deprotected according to General Procedure J. The reaction mixture was quenched with 1.0 M HCl, the solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield the product (26 mg, 0.048 mmol, 16%) as a white solid. MS (ESI) *m/z* calc. for $C_{31}H_{30}N_8O_2$ $[M+H]^+$ 547.26; found 547.40. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 11.88 (s, 1H), 10.45 (s, 1H), 10.18 (s, 1H), 9.08 (s, 1H), 8.90 (s, 1H), 8.60 (d, *J* = 6.5 Hz, 2H), 7.97 (s, 3H), 7.81 (d, *J* = 7.8 Hz, 2H), 7.56 (s, 2H), 7.27 (d, *J* = 7.7 Hz, 1H), 7.21 (s, 1H), 6.79 (d, *J* = 15.0 Hz, 1H), 6.57 (s, 1H), 6.36 (d, *J* = 15.0 Hz, 1H), 3.22 (s, 2H), 2.28 (s, 6H), 2.24 (s, 3H).

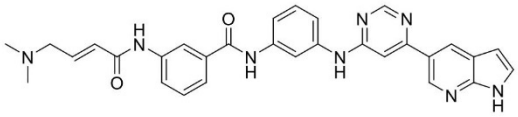
3-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



13d

N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3b**) (0.88 g, 2 mmol) and 3-nitrobenzoyl chloride (0.45 g, 2.4 mmol) were reacted as described in General Procedure D. The reaction mixture was diluted with DCM and the organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The reaction mixture was purified by flash chromatography to yield 3-nitro-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.8 g, 1.35 mmol, 68%) as a slightly yellow solid.
The intermediate nitrobenzamide (0.8 g, 1.35 mmol) was reduced to the respective aromatic amine (**13d**) according to General Procedure F. The reaction mixture was filtered through a pad of CELITE and the filtrate was concentrated in vacuo to yield the product (0.4 g, 0.71 mmol, 53%) as a slightly yellow solid.

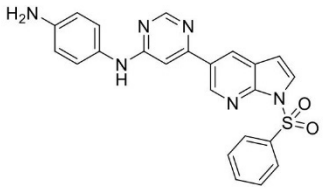
(*E*)-*N*-(3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)-3-(4-(dimethylamino)but-2-enamido)benzamide



17

Compound **17** was synthesized from 3-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**13d**) (0.4 g, 0.7 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield (*E*)-3-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.28 g, 0.42 mmol, 59%) as a slightly yellow solid.
The intermediate was further deprotected according to General Procedure J. The reaction mixture was quenched with 1.0 M HCl, the solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield the product (102 mg, 0.19 mmol, 46%) as a white solid. MS (ESI) *m/z* calc. for $C_{30}H_{28}N_8O_2$ $[M+H]^+$ 532.23; found 533.60. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 11.86 (s, 1H), 10.32 (s, 1H), 10.27 (s, 1H), 9.70 (s, 1H), 8.91 (d, *J* = 2.1 Hz, 1H), 8.72 (d, *J* = 0.9 Hz, 1H), 8.60 (d, *J* = 2.0 Hz, 1H), 8.18 (t, *J* = 1.9 Hz, 2H), 7.90 (s, 1H), 7.66 (d, *J* = 8.0 Hz, 1H), 7.58-7.55 (m, 1H), 7.52 (s, 1H), 7.48 (t, *J* = 7.9 Hz, 1H), 7.39 (s, 1H), 7.39-7.33 (m, 1H), 7.31 (d, *J* = 7.9 Hz, 1H), 6.78 (d, *J* = 15.4 Hz, 1H), 6.59 (dd, *J* = 3.4, 1.8 Hz, 1H), 6.30 (d, *J* = 15.4 Hz, 1H), 3.10 (d, *J* = 5.9 Hz, 2H), 2.21 (s, 6H).

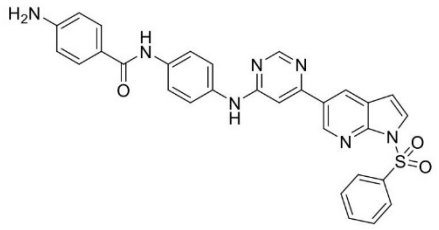
N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,4-diamine



12d

Benzene-1,4-diamine (0.81 g, 7.5 mmol) and 5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine (**2b**) (1.85 g, 5 mmol) were dissolved in NMP. DIEA (0.97 g, 7.5 mmol) was added and the reaction mixture was stirred at 150 °C for 16 h. The mixture was allowed to cool to room temperature and was diluted with EtOAc. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography (MeOH:DCM) to yield the product (**12d**) (1.3 g, 2.9 mmol, 60%).

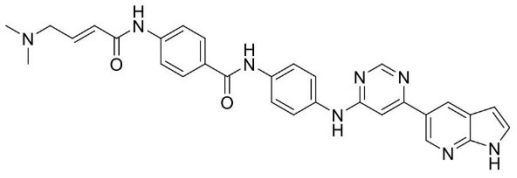
4-amino-*N*-(4-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide



13e

N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,4-diamine (**12d**) (0.6 g, 1.36 mmol), 4-nitrobenzoyl chloride (0.3 g, 1.2 mmol) were coupled as described in General Procedure D. The resulting suspension was filtered, and the filtrate was washed with DCM to 4-nitro-*N*-(4-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.65 g, 1.1 mmol, 85%).
The intermediate (0.65 g, 1.1 mmol) was converted to the respective free amine (**13e**) following General Procedure F to yield the product (0.4 g, 0.71 mmol, 60%).

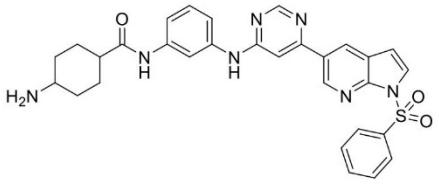
(*E*)-*N*-(4-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)benzamide



18

Compound **18** was synthesized from 4-amino-*N*-(4-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (**13e**) (0.11 g, 0.84 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by flash chromatography (MeOH:DCM) to yield (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(4-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)benzamide (0.17 g, 0.19 mmol, 30%).
The intermediate was further deprotected according to General Procedure J. The reaction mixture was extracted with chloroform/2-propanol (4:1, *v/v*). The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (22 mg, 0.038 mmol, 10%) as a white solid. MS (ESI) *m/z* calc. for $C_{30}H_{28}N_8O_2$ $[M+H]^+$ 533.24; found 533.42. ^{1}H NMR (400 MHz, DMSO-*d6*) δ 11.86 (s, 1H), 10.34 (s, 1H), 10.11 (s, 1H), 9.61 (s, 1H), 8.88 (s, 1H), 8.69 (s, 1H), 8.57 (s, 1H), 7.96 (d, J = 8.7 Hz, 2H), 7.90-7.77 (m, 4H), 7.67 (d, J = 9.1 Hz, 2H), 7.59-7.51 (m, 1H), 7.26 (s, 1H), 6.79 (m, 1H), 6.59 (m, 1H), 6.32 (d, J = 15.2 Hz, 1H), 3.13 (s, 2H), 2.28 (s, 6H).

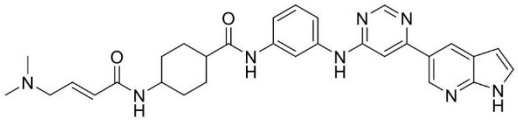
4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)cyclohexane-1-carboxamide



13f

N^1-(6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)benzene-1,3-diamine (**3b**) (1.5 g, 3.39 mmol), 4-((*tert*-butoxycarbonyl)amino)cyclohexane-1-carboxylic acid (0.9 g, 3.73 mmol) and HATU (1.55 g, 4.07 mmol) were dissolved in DCM. DIEA was added and the reaction mixture was stirred at room temperature for 3 h. The mixture was diluted with water and the aqueous layer was extracted with DCM. The organic layer was further washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was then purified by flash chromatography to yield *tert*-butyl (4-((3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)carbamoyl)cyclohexyl)carbamate (1.5 g, 2.25 mmol, 60%).
The intermediate was deprotected to the free amine (**13f**) by treating the respective MeOH-solution with 4.0 M HCl_{aq}. After completion of the reaction, the solvent was removed *in vacuo* to yield the product (HCl-salt, 1.2 g, 2.11 mmol, 90%).

(*E*)-*N*-(3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)-4-(4-(dimethylamino)but-2-enamido)cyclohexane-1-carboxamide

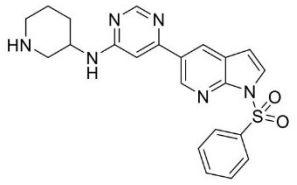


19

Compound **19** was synthesized from 4-amino-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)cyclohexane-1-carboxamide (**13f**) (1.0 g, 1.76 mmol) following General Procedure H to yield (*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)cyclohexane-1-carboxamide (0.5 g, 0.7 mmol, 30%).
The intermediate was further deprotected according to General Procedure J. The reaction mixture was extracted with chloroform/2-propanol (4:1, *v*/*v*). The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography to yield the product (0.15 g, 0.28 mmol, 30%) as a white solid. MS (ESI) *m/z* calc. for $C_{30}H_{34}N_8O_2$ $[M+H]^+$ 539.29; found 539.47. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 11.86 (s, 1H), 9.86 (s, 1H), 9.65 (s, 1H), 8.89 (d, J = 2.1 Hz, 1H), 8.69 (s, 1H), 8.59 (d, J = 2.1 Hz, 1H), 8.02 (s, 1H), 7.97 (d, J = 7.0 Hz, 1H), 7.57-7.52 (m, 1H), 7.38 (d, J = 7.6 Hz, 1H), 7.33 (s, 1H), 7.29-7.21 (m, 2H), 6.78 (dd, J = 3.4, 1.8 Hz, 1H), 6.54 (dt, J = 15.6, 6.1 Hz, 1H), 6.19 (d, J = 15.6 Hz, 1H), 3.90 (s, 1H), 2.96 (d, J = 6.1 Hz, 2H), 2.48-2.39 (m, 1H), 2.13 (s, 6H), 1.88-1.77 (m, 2H), 1.77-1.67 (m, 2H), 1.67-1.49 (m, 4H).

2.1.4 Synthesis of compound 22

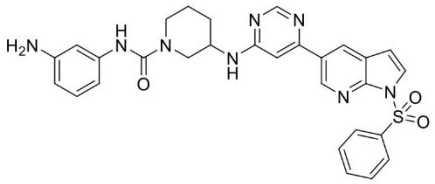
6-(1-(phenylsulfonyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-N-(piperidin-3-yl)pyrimidin-4-amine



20

Tert-butyl 3-aminopiperidine-1-carboxylate (7.2 g, 36 mmol) and 5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine (**2b**) (11 g, 30 mmol) were dissolved in NMP (200 mL). DIEA (4.3 g, 33 mmol) was added and the reaction mixture was stirred at 110 °C for 16 h. The mixture was allowed to cool to room temperature and poured into water/EtOAc mixture. The resulting suspension was filtered, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography to yield *tert*-butyl 3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)piperidine-1-carboxylate (8 g, 15 mmol, 50%).
The intermediate was redissolved in DCM (80 mL). TFA (40 mL) was added and the reaction mixture was stirred at room temperature for 16 h. The solvent was removed *in vacuo* to yield the product (**20**), which was used without further purification.

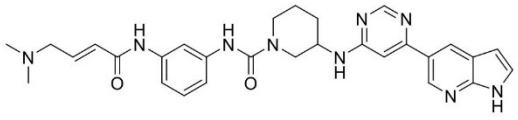
***N*-(3-aminophenyl)-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)piperidine-1-carboxamide**



21

3-nitroaniline (0.33 g, 2.4 mmol) in DCM (10 mL) was added to a solution of triphosgene (0.24 g, 0.8 mmol) and TEA (0.71 g, 7 mmol) in DCM (30 mL). The mixture was stirred at 0 °C for 2 h. A solution of 6-(1-(phenylsulfonyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-N-(piperidin-3-yl)pyrimidin-4-amine (**20**) (0.87 g, 2 mmol) in DCM was added dropwise to this mixture and the resulting reaction mixture was allowed to warm to room temperature and was stirred for 16 h. Filtration of the reaction mixture yields in *N*-(3-nitrophenyl)-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)piperidine-1-carboxamide (1 g, 1.67 mmol, 83%).
The intermediate (1.29 g, 2.15 mmol) was redissolved in EtOAc/MeOH (50 mL, 4:1, *v*/*v*) and $SnCl_2$ (3.3 g, 17.24 mmol) was added. The mixture was stirred at 80 °C for 16 h. The mixture was allowed to cool to room temperature and was diluted with chloroform/2-propanol (4:1, *v*/*v*). The organic layer was washed with saturated sodium bicarbonate solution and filtered through a pad of CELITE. The filtrate was dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography (1.75 M NH_3 in MeOH:DCM) to yield the product (**21**) (0.9 g, 1.58 mmol, 74%).

(*E*)-3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)-*N*-(3-(4-(dimethylamino)but-2-enamido)phenyl)piperidine-1-carboxamide

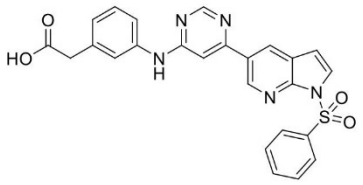


22

Compound **22** was synthesized from *N*-(3-aminophenyl)-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)piperidine-1-carboxamide (**21**) (0.32 g, 0.56 mmol) following General Procedure G. The solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield (*E*)-*N*-(3-(4-(dimethylamino)but-2-enamido)phenyl)-3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)piperidine-1-carboxamide (0.26 g, 0.38 mmol, 68%).
The intermediate was further deprotected according to General Procedure J. The reaction mixture was extracted with chloroform/2-propanol (4:1, *v*/*v*). The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was purified by flash chromatography (MeOH:DCM) to yield the product (0.06 g, 0.038 mmol, 30%) as a white solid. MS (ESI) *m/z* calc. for $C_{29}H_{33}N_9O_2$ $[M+H]^+$ 540.28; found 540.45. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 11.80 (s, 1H), 9.97 (s, 1H), 8.85 (s, 1H), 8.52 (s, 2H), 7.84 (s, 1H), 7.55-7.48 (m, 2H), 7.40 (s, 1H), 7.29-7.19 (m, 2H), 7.18-6.92 (m, 3H), 6.75-6.64 (m, 1H), 6.54 (s, 1H), 6.26 (d, *J* = 15.3 Hz, 1H), 4.21-3.80 (m, 2H), 3.04 (d, *J* = 5.2 Hz, 2H), 3.00-2.78 (m, 3H), 2.17 (s, 6H), 2.01 (s, 1H), 1.76 (s, 1H), 1.53 (s, 2H).

2.1.5 Synthesis of compound 25

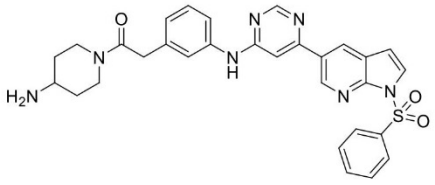
2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)acetic acid



23

2-(3-aminophenyl)acetic acid (1.83 g, 12 mmol) and 5-(6-chloropyrimidin-4-yl)-1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridine (**2b**) (3.7 g, 10 mmol) were dissolved in NMP (10 mL). The reaction mixture was stirred at 140 °C for 12 h. The mixture was allowed to cool to room temperature and diluted with EtOAc. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography to yield the product (**23**) (0.3 g, 0.62 mmol, 6%) as a slightly yellow solid.

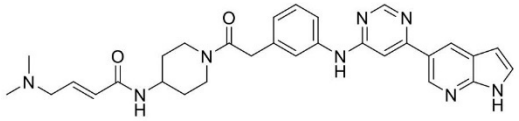
1-(4-aminopiperidin-1-yl)-2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)ethan-1-one



24

2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)acetic acid (**23**) (0.3 g, 0.62 mmol), *tert*-butyl piperidin-4-ylcarbamate (0.19 g, 0.93 mmol) and HATU (0.35 g, 0.93 mmol) were dissolved in DCM. DIEA (0.22 mL, 1.24 mmol) was added and the reaction mixture was stirred at room temperature for 30 min. The mixture was diluted with EtOAc. The organic layer was further washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude was then purified by flash chromatography to yield *tert*-butyl (1-(2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)acetyl)piperidin-4-yl)carbamate (0.21 g, 0.31 mmol, 51%) as a white solid.
The intermediate was deprotected to the free amine (**24**) by treating the respective EtOAc-solution with 4.0 M HCl_{aq}. After completion of the reaction, the pH of the solution was adjusted to 7.0 with 1.0 M NaOH. The organic layer was washed with water and brine, dried over Na_2SO_4 and the solvent was removed *in vacuo* to yield the product (0.15 g, 0.26 mmol, 90%).

(*E*)-*N*-(1-(2-(3-((6-(1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)acetyl)piperidin-4-yl)-4-(dimethylamino)but-2-enamide

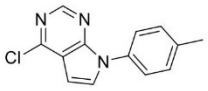


25

Compound **25** was synthesized from 1-(4-aminopiperidin-1-yl)-2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)ethan-1-one (**24**) (0.5 g, 0.88 mmol) following General Procedure H to yield (*E*)-4-(dimethylamino)-*N*-(1-(2-(3-((6-(1-(phenylsulfonyl)-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)pyrimidin-4-yl)amino)phenyl)acetyl)piperidin-4-yl)but-2-enamide (0.3 g, 0.44 mmol, 50%) as a white solid.
The intermediate was further deprotected according to General Procedure J. The pH of the reaction mixture was adjusted to 7.0 using 1.0 M HCl. The solvent was removed *in vacuo* and the crude was purified by flash chromatography to yield the product (180 mg, 0.33 mmol, 56%) as a white solid. MS (ESI) *m/z* calc. for $C_{30}H_{34}N_8O_2$ $[M+H]^+$ 539.29; found 539.46. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 11.87 (s, 1H), 9.66 (s, 1H), 8.90 (d, *J* = 2.1 Hz, 1H), 8.70 (s, 1H), 8.59 (d, *J* = 2.1 Hz, 1H), 7.95 (d, *J* = 7.6 Hz, 1H), 7.64 (d, *J* = 8.2 Hz, 1H), 7.56-7.54 (m, 2H), 7.36-7.23 (m, 2H), 6.92 (d, *J* = 7.6 Hz, 1H), 6.63-6.48 (m, 2H), 6.01 (d, *J* = 15.6 Hz, 1H), 4.26 (d, *J* = 13.3 Hz, 1H), 3.92 (d, *J* = 14.3 Hz, 1H), 3.87-3.79 (m, 1H), 3.73 (s, 2H), 3.19-3.12 (m, 1H), 2.99 (d, *J* = 6.0 Hz, 2H), 2.87-2.77 (m, 1H), 2.15 (s, 6H), 1.78-1.69 (m, 2H), 1.27-1.16 (m, 1H), 0.86 (t, *J* = 7.3 Hz, 1H).

2.1.6 Synthesis of compound 30

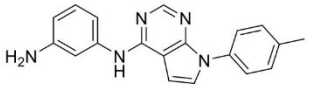
4-chloro-7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine



27

4-chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (0.25 g, 1.63 mmol), 4-iodotoluidine (0.39 g, 1.79 mmol), *trans*-1,2-cyclohexanediamine (59 µL, 0.49 mmol), K_3PO_4 (0.86 g, 4.07 mmol) and CuI (31 mg, 0.16 mmol) were dissolved in toluene (10 mL) and the reaction mixture was stirred for 17 h at 110 °C. After cooling to room temperature, the mixture was diluted with DCM, filtered and the solvent was removed *in vacuo*. The crude was further purified by flash chromatography (MeOH:DCM) to yield the product (**27**) (0.19 g, 0.78 mmol, 48%) as a slightly yellow solid. MS (ESI) *m/z* calc. for $C_{13}H_{10}ClN_3$ $[M+H]^+$ 244.06; found 244.03.

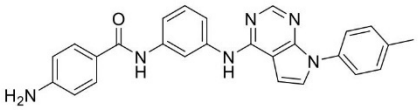
N^1-(7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)benzene-1,3-diamine



28

Compound **28** was synthesized from *m*-phenylenediamine (0.83 g, 7.7 mmol) and 4-chloro-7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine (**27**) (0.37 g, 1.5 mmol) following General Procedure B to yield the product (0.8 g, 100% yield = 0.5 g) as a brown solid. MS (ESI) *m/z* calc. for $C_{19}H_{17}N_5$ $[M+H]^+$ 316.16; found 316.05.

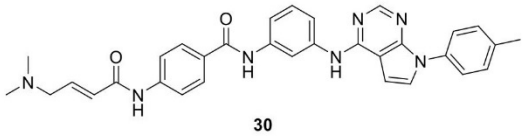
4-amino-*N*-(3-((7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)amino)phenyl)benzamide



29

N^1-(7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)benzene-1,3-diamine (**28**) (0.48 g, 1.5 mmol) and 4-nitrobenzoyl chloride (341 mg, 1.8 mmol) were reacted as described in General Procedure C. Another 0.2 eq. of 4-nitrobenzoyl chloride were added to complete the reaction.
Without further purification, the resulting nitrobenzamide was reduced to the respective aromatic amine (**29**), following General Procedure E. The reaction mixture was purified by flash chromatography (MeOH:DCM) to yield the product (268 mg, 0.62 mmol, 40% over 2 steps) as a slightly brown solid. MS (ESI) *m/z* calc. for $C_{26}H_{22}N_6O$ $[M+H]^+$ 435.19; found 435.25.

(*E*)-4-(4-(dimethylamino)but-2-enamido)-*N*-(3-((7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)amino)phenyl)benzamide



30

Compound **30** was synthesized from 4-amino-*N*-(3-((7-(*p*-tolyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)amino)phenyl)benzamide (**29**) (0.1 g, 0.23 mmol) following General Procedure G. The crude was purified by reverse-phase HPLC to yield the product (TFA-salt, 4 mg, 0.006 mmol, 4%) as a white solid. MS (ESI) *m/z* calc. for $C_{32}H_{31}N_7O_2$ $[M+H]^+$ 546.26; found 546.00. ^{1}H NMR (500 MHz, DMSO-*d6*) δ 10.61 (s, 1H), 10.20 (s, 1H), 9.90 (s, 1H), 9.64 (s, 1H), 8.39-8.32 (m, 2H), 8.00 (d, *J* = 8.9 Hz, 2H), 7.82 (d, *J* = 8.5 Hz, 2H), 7.75-7.63 (m, 4H), 7.42 (s, 1H), 7.39-7.29 (m, 3H), 7.07 (d, *J* = 3.7 Hz, 1H), 6.78 (s, 1H), 6.50 (d, *J* = 15.3 Hz, 1H), 3.96 (s, 2H), 2.82 (s, 6H), 2.38 (s, 3H).

2.2 X-ray Crystallography

THZ-P1-2 was co-crystallized with PI5P4Kα as previously described[1].

2.3 Mass Spectrometry

PI5P4Kα and γ were incubated with compound **30** and analyzed as previously described[1].

2.4 Biochemical Assay Protocols

2.4.1 Materials and Methods

Compounds

All compounds were initially sourced from the National Center for Advancing Translational Studies (NCATS)/National Institutes of Health (NIH). All compounds shipped from NCATS were subjected to quality control by LC/UV, LC/MS, or HRMS unless otherwise noted. All compounds exhibited >90% purity by peak area or *m/z*[2].

qHTS Data Analysis and Statistics

Data from each assay were normalized plate-wise to corresponding intra-plate controls (neutral control DMSO and positive control as noted) as described previously[3]. The same controls were also used for the calculation of the Z'-factor, a measure of assay quality control, as previously described[4]. Percent activity

was derived, using in-house software[3]. Concentration-response curves were fitted, classified, and IC_{50} values were determined as described previously[5,6]. Otherwise, concentration-response curves were fitted and IC_{50} values were calculated using Prism software (version 7.04, GraphPad Software, Inc. San Diego, CA), sigmoidal dose-response (variable slope) equation.

2.4.2 ADP-Glo™ Kinase Assay (PI5P4Kα)

Screen
Assay buffer conditions were adopted from previously described methods[7,8] incorporating the ADP-Glo™ Kinase Assay (Promega, Madison, WI). Briefly, 2 μL of PI5P4Kα and "Lipid Mix" (final concentrations of 20 nM and "1X" mixture of 52.8 μM DPPS (1,2-dipalmitoyl-sn-glycero-3-phosphoserine) and 20.9 μM D-myo (Dipalmitoyl Phosphatidylinositol 5-phosphate), respectively; columns 1-2, 5-48) in "reaction buffer" (40 mM HEPES pH 7.5, 0.25 mM EGTA, 20 mM $MgCl_2$ with 0.1% CHAPS; columns 3-4), or 2 μL of reaction buffer, were dispensed into a 1,536-well solid-bottom white plate (Greiner Bio One, Monroe, NC). 16 nL of DMSO or compound (final concentration range 884 pM to 52.2 μM) or intraplate control compounds NCGC00351566 or NCGC00372529 (final concentration range 4.7 nM to 10 μM or 23.9 nM to 52.2 μM, respectively) were transferred *via* Wako Pin-tool (Wako Automation, Richmond, VA). Samples were incubated for 15 min (room temperature, protected from light), followed by a 1 μL dispense (final assay volume of 3 μL) to all wells of ATP (final concentrations of 5 μM; in "ATP buffer": 20 mM HEPES, pH 7.5, 60 mM $MgCl_2$ with 0.1% CHAPS). Plates were incubated for 1 h (room temperature, protected from light) followed by 2 μL of ADP-Glo™ Reagent to all wells and a room temperature incubation (protected from light) of 40 min, to allow enough time for ATP to convert to ADP. Following the incubation, 4 μL of Kinase Detection Reagent was dispensed to all wells, incubated for 30 min (room temperature, protected from light), and read on a ViewLux imager (Perkin Elmer, Waltham, MA) using luminescence optics. The luminescence intensity (RLU) was normalized against DMSO and no-enzyme controls and the resulting percent inhibition data were fitted to a 4-parameter Hill equation.

Counterscreen
Assay buffer conditions were adopted from previously described methods[8]. Briefly, 2 μL of reaction buffer with 7.5% DMSO (final concentration of 5%), were dispensed into a 1,536-well solid-bottom white plate (Greiner Bio One, Monroe, NC). 16 nL of DMSO or compound (final concentration range 884 pM to 52.2 μM) were transferred *via* Wako Pin-tool (Wako Automation, Richmond, VA). Samples were incubated for 15 min (room temperature, protected from light), followed by a 1 μL dispense (final assay volume of 3 μL) of ATP/ADP mixture (final concentrations of 3 μM and 12 μM, respectively, in ATP buffer; representing 20% conversion levels; columns 1-2, 5-48) or ATP (final concentrations 15 μM in ATP buffer; columns 3 and 4). Plates were incubated for 1 h (room temperature, protected from light) followed by 2 μL of ADP-Glo™ Reagent to all wells and a room temperature incubation (protected from light) of 40 min. Following the incubation, 4 μL of Kinase Detection Reagent was dispensed to all wells, incubated for 30 min (room temperature, protected from light), and read on a ViewLux imager (Perkin Elmer, Waltham, MA) using luminescence optics. The luminescence intensity (RLU) was normalized against DMSO and no-ADP controls and the resulting percent inhibition data were fitted to a 4-parameter Hill equation.

2.4.3 GDP Transcreener Assay (PI5P4Kβ)

Screen
Briefly, 2 µL of PI5P4Kβ and "Lipid Mix" (final concentrations of 50 nM and "1X", respectively) in assay buffer (40 mM HEPES, pH 7.5, 0.25 mM EGTA, 20 mM $MgCl_2$ with 0.1% CHAPS), or 2 µL of assay buffer were dispensed into a 1,536-well solid-bottom black plate (Greiner Bio One, Monroe, NC). 34 to 46 nL of DMSO, compound (final concentration range 1.3 nM to 114 µM) or intraplate control compounds NCGC00351566 or NCGC00372529 (final concentration range 7.7 nM to 17 µM or 38.5 µM, respectively) were transferred *via* Wako Pin-tool (Wako Automation, Richmond, VA). Samples were incubated (room temperature, protected from light) for 15 min, followed by a 2 µL dispense to all wells of GDP Transcreener mixture (final concentrations of 10 µM, 2 nM, and 10 µg/mL of GTP, Alexa633, and GDP antibody, respectively). Plates were centrifuged (1,000 rpm), then read immediately on an EnVision microplate reader (PerkinElmer, Waltham, MA) equipped with fluorescence polarization (FP) optics (620(45) nm excitation, 688(45) nm emission, S and P filters; Mirror = D658fp/D688). The plate was then incubated (room temperature, protected from light) for 90 min (~90% conversion) and then read for a second time, using the same optics. The change in fluorescence polarization (ΔmP) over the 120-minute reaction period was normalized against no-inhibitor and no-enzyme controls and the resulting percent inhibition data were fitted to a 4-parameter Hill equation.

Counterscreen
Briefly, 2 µL of 5% DMSO in assay buffer, were dispensed into a 1,536-well solid-bottom black plate. 34 to 46 nL of DMSO of compound (final concentration range 1.3 nM to 114 µM) were transferred *via* Wako Pin-tool. Samples were incubated (room temperature, protected from light) for 15 min, followed by a 2 µL dispense to columns 1-2, 5-48 of the counterscreen GDP Transcreener mixture (final concentrations of 2.5 µM, 7.5 µM, 2 nM, and 10 µg/mL of GDP, GTP, Alexa633, and GDP antibody, respectively) or 2 µL to columns 3 and 4 of GDP Transcreener mixture. Plates were processed, read and analyzed as described above.

2.4.4 Assay Optimization and Miniaturization

PI5P4Kα
Using the protocol described above, we tested 20 nM PI5P4Kα vs. NCGC00351566 and NCGC00372529, where we observed a Z'-factor of 0.8 (ΔRLU ~940; S:B ~24) resulting in IC_{50} values of 7.27 nM and 57.6 nM, respectively (Supporting Figure 3B), confirming the assay was adopted correctly from the previously published method.

PI5P4Kβ
Assay design was based on previous work[9].
Assay buffer conditions were adopted from previously described methods[7,8] and modified to 40 mM HEPES, pH 7.5, 0.25 mM EGTA, 20 mM $MgCl_2$ with 0.1% CHAPS.
To measure the rate of GDP formation, we purchase the Transcreener® GDP FP Assay (Bellbrook Labs, Madison, Wisconsin).

We first performed a GDP antibody titration as recommended in their technical manual[10]. To simulate assay conditions, we added 5% DMSO to the buffer. In a 1,536-well plate, we performed a GDP antibody titration (assay volume = 4 µL) at 500 µg/mL, 1:2 dilution, 24-points (n = 4), containing 2 µM Alexa633

Tracer and 10 μM GTP, followed by a 1 h incubation (room temperature, protected from light). The plate was read on an EnVision microplate reader (PerkinElmer, Waltham, MA) equipped with fluorescence polarization (FP) optics (620(45) nm excitation, 688(450) nm emission, S and P filters; Mirror = D658fp/D688). Using Prism software (version 7.04, GraphPad Software, Inc. San Diego, CA), we calculated an EC_{85} value of 10.5 μg/mL, and chose to move forward with an antibody concentration of 10 μg/mL (Supporting Figure 1).

Next, we sought to determine the signal window of the assay by performing a GDP titration at various % GTP conversions of 0, 1, 10, 50, and 100%, or GDP concentrations of 0, 0.1, 1, 5, and 10 μM. We mixed these respective GDP concentrations with our detection mixture of 10 μM GTP, 10 μg/mL GDP antibody, and 2 nM Alexa633 tracer, incubated for 1 h (room temperature, protected from light) and the read for FP, using the above parameters. As shown in Supporting Figure 1, we observed acceptable assay statistics (subtracting the signal at the various GDP concentrations from the zero concentration wells) at 10% GTP conversion (1 μM GDP) or greater, with a signal window (ΔmP) range of 126-191, and Z'-factor range of 0.59 to 0.72.

Having confirmed the Transcreener GDP FP assay components work in 1,536-well format, we next determined its compatibility with our enzyme and substrate. Substrate solutions were prepared as previously described. Briefly, 1,2-Dipalmitoyl-sn-glycero-3-phosphoserine, sodium salt (DPPS; Echelon Biosciences, Salt Lake City, Utah) was weighed out in a glass vial and suspended in DMSO at concentration of 3 mg/mL (4 mM) and sonicated. Dipalmitoyl Phosphatidylinositol 5-phosphate (D-myo; Echelon Biosciences) was similarly weighed out in a glass vial and suspended in DMSO at a concentration of 3 mg/mL (3.1 mM) and sonicated. The DPPS and D-myo were combined ("Lipid Mix") at a ration of 2:1 in DMSO, for working concentrations of 106 μM and 41.8 μM, respectively. The Lipid Mix (final concentration of 52.8 μM and 20.9 μM for DPPS and D-myo, respectively, or "1X") was then mixed with enzyme (final concentrations of 0, 25, 50, 100, and 200 nM; n = 64/condition) or assay buffer in a glass vial, dispensed at 2 μL into a 1,536-well plate. The plate was then incubated at room temperature for 15 min, followed by a 2 μL addition of the GDP Transcreener solution of GTP, Alexa633, and GDP antibody at final concentrations of 10 μM, 2 nM, and 10 μg/mL, respectively, and a final percent DMSO of 5%. The plate was centrifuged at 1,000 rpm and immediately read, using the optics described above. The plate was then read in 15-min intervals over 120 min (Supporting Figure 2). For signal window calculation, we chose the 15, 30, 60, 90, and 120 min timepoints, and calculated the ΔmP for each to determine optimum assay performance (Supporting Figure 2).

We observed a ΔmP ranges of 14-71, 25-93, 43-117, and 58-122 for the 25, 50, 100, and 200 nM enzyme concentrations, respectively, over the 120-minute time-course. We observed Z'-factor ranges of (-2.4)-0.2, (-0.8)-0.5, (-0.2)-0.5, and 0.2-0.6 for the 25, 50, 100, and 200 nM enzyme concentrations, respectively, over the 120-minute time-course. Our goal was to have as low an enzyme concentration as possible, while maintaining acceptable assay statistics. With this in mind, balancing assay performance, reagent consumption, and percent conversion, we chose 50 nM enzyme and a 90 min incubation (~90% conversion) for compound testing, where we observed a ΔmP of 84 and Z'-factor of 0.3.

With these conditions in place, we initially prepared intraplate titrations of NCGC00372529 and NCGC00357560 (inactive compound) (final concentration ranges of 52 nM to 114 μM). Briefly, 2 μL of PI5P4Kβ and "Lipid Mix" (final concentrations of 50 nM and "1X", respectively) in assay buffer (40 mM HEPES, pH 7.5, 0.25 mM EGTA, 20 mM $MgCl_2$ with 0.1% CHAPS), or 2 μL of assay buffer, were dispensed into a 1,536-well solid-bottom black plate (Greiner Bio One, Monroe, NC). 46 nL of DMSO or intraplate control compounds NCGC00351566 or NCGC00372529 (final concentration range 7.7 nM to 17 μM or 38.5 nM to 84 μM, respectively) were transferred *via* Wako Pin-tool (Wako Automation,

Richmond, VA). Samples were incubated (room temperature, protected from light) for 15 min, followed by a 2 μL dispense to all wells of GDP Transcreener mixture (final concentrations of 10 μM, 2 nM, and 10 μg/mL of GTP, Alexa633, and GDP antibody, respectively). Plates were centrifuged (1,000 rpm), then read immediately on an EnVision microplate reader (PerkinElmer, Waltham, MA) equipped with fluorescence polarization (FP) optics (620(45) nm excitation, 688(45) nm emission, S and P filters; Mirror = D658fp/D688). The plate was then incubated (room temperature, protected from light) for 90 min (~90% conversion) and then read for a second time, using the same optics. The change in fluorescence polarization (ΔmP) over the 90-minute reaction period was normalized against no-inhibitor and no-enzyme controls and the resulting percent inhibition data were fitted to a 4-parameter Hill equation. We observed a Z'-factor of ~0.3 (ΔmP ~74), resulting in IC_{50} values of 14.6 μM and inactive (>114 μM) for NCGC00372529 and NCGC00357560, respectively (Supporting Figure 3A).

Using the above protocol, we then compared NCGC00351566 and NCGC00372529, where we observed a Z'-factor of 0.3 (ΔmP ~86), resulting in IC_{50} values of 0.229 μM and 22.0 μM, respectively (Supporting Figure 3B).

2.5 Kinome Profiling

Kinome profiling was performed, using KINOMEscan at compound concentrations of 1 μM. Protocols are available from DiscoverRx.

TREE*spot* images were generated, using TREE*spot*™ Software Tool, and reprinted with permission from KINOMEscan®, a division of DiscoverRx Corporation, ©DISCOVERX CORPORATION 2010.

2.6 CETSA Protocol

2.6.1 Cell culture

HEK 293T cells were cultured as recommended by ATCC and were kept frozen, so only freshly thawed cells at early passage were used. No authentication was done by the authors. Cells were maintained in DMEM media, which was supplemented with 10% FBS and 100 μL/mL penicillin-streptomycin and incubated at 37 °C and 5% CO_2.

2.6.2 Compounds and Reagents

Anti-PI5P4Kα (5527) and anti-PI5P4Kβ (9694) were purchased from Cell Signaling Technology (Hanover, MA). Anti-GAPDH (MAB374) was obtained from EMD Millipore Corporation (Merck KGaA, Darmstadt, Germany).

2.6.3 Cellular Thermal Shift Assay

The following procedures were performed in accordance to reported protocols[11], unless otherwise noted.

T_m Determination (melting curve)

A suspension of 20 Mio. HEK 293T cells in 15 mL DMEM media were plated in 15 cm culture dishes

and incubated overnight. The confluent cells were treated with either 15 µL of 10 mM THZ-P1-2 stock solution in DMSO or DMSO and were incubated for 1 h at 37 °C. The media was aspirated, and the cells were detached from the culture dish, using trypsin. The cell suspensions were spun down (1,000 rpm, 3 min) and the media/trypsin mixture was aspirated. After the cells were washed with PBS buffer, the cell pellets were resuspended in 1.05 mL PBS buffer, supplemented with a cOmplete™ Mini Protease Inhibitor Cocktail tablet (Roche). The cell suspensions were split into aliquots of 100 µL and transferred into 10 PCR tubes per condition. The leftover cell suspension was checked for cell viability.

The cell suspensions were heated in a SimpliAmp™ thermal cycler (AppliedBiosystems®, Life Technologies) for 3 min at different temperatures, ranging from 40-67 °C. After the heating step, the tubes were allowed to cool down to room temperature over a period of 3 min and were then immediately snap-frozen, using liquid nitrogen. The cells were further lysed by freeze-thawing the tubes twice using liquid nitrogen and the thermal cycler (at 25 °C). The tubes were vortexed between each step. The resulting lysate was further kept on ice.

Isothermal Dose-Response
2 mL of a 1 Mio cells/mL cell suspension of HEK 293T cells were transferred into wells of a 6-well plate. The cells were either treated with different concentrations of inhibitor (ranging from 100 nM to 10 µM) or DMSO and were incubated for 1 h at 37 °C. The media was aspirated, and the cells were detached from the culture dish, using trypsin. The cell suspensions were spun down (1,000 rpm, 3 min) and the media/trypsin mixture was aspirated. After the cells were washed with PBS buffer, the cell pellets were resuspended in 100 µL PBS buffer, supplemented with a cOmplete™ Mini Protease Inhibitor Cocktail tablet (Roche).

The cell suspensions were heated in a SimpliAmp™ thermal cycler (AppliedBiosystems®, Life Technologies) for 3 min at 52 °C. After the heating step, the tubes were allowed to cool down to room temperature over a period of 3 min and were then immediately snap-frozen, using liquid nitrogen. The cells were further lysed by freeze-thawing the tubes twice using liquid nitrogen and the thermal cycler (at 25 °C). The tubes were vortexed between each step. The resulting lysate was kept on ice from this point on.

2.6.4 Western Blot Analysis

The cell lysates were clarified by centrifugation at 14,000 rpm for 15 min at 4 °C. The supernatants were normalized to have equal protein content, using a BCA assay (Pierce), and diluted to a total volume of 75 µL. 25 µL 4x LDS/BME sample buffer were added to each sample, which were boiled at 95 °C for 5 min.

Imaging was performed by detection of fluorescently labeled infrared secondary antibodies (IRDye) on the Odyssey CLx Imager (LI-COR). The bands observed were quantified using Image Studio™ Lite software. The relative band intensities (%) were plotted as a function of temperature or ligand concentration using Prism software (version 7.04, GraphPad Software, Inc. San Diego, CA).

3. Abbreviations

ACN, acetonitrile; ADP, adenosine diphosphate; ATP, adenosine triphosphate; ATCC, American type culture collection; BCA, bicinchoninic acid; BME, β-mercaptoethanol; calc., calculated; cpd, compound; CETSA, cellular thermal shift assay; CHAPS, 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate; DCM, dichloromethane; DIEA, *N*,*N*-diisopropylethylamine; DMA, dimethylamine; DMEM, Dulbecco's modified eagle medium; DMF, dimethylformamide; DMSO, dimethyl sulfoxide; D-myo, dipalmitoyl phosphatidylinositol 5-phosphate; DPPS, 1,2-dipalmitoyl-sn-glycero-3-hosphoserine; FP, fluorescence polarization; EGTA, ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid; ESI, electrospray ionization; FBS, fetal bovine serum; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GDP, guanosine diphosphate; GTP, guanosine triphosphate; HATU, 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxide hexafluorophosphate; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HRMS, high-resolution mass spectrometry; HPLC, high-performance liquid chromatography; LC, liquid chromatography; LDS, lithium dodecyl sulfate; MS, mass spectrometry; NCATS, National Center for Advancing Translational Studies; NIH, National Institutes of Health; NMR, nuclear magnetic resonance; NMP, *N*-methyl-2-pyrrolidone; PBS, phosphate-buffered saline; $Pd_2(dba)_3$, Tris(dibenzylideneacetone)dipalladium(0); PI5P4K, phosphatidylinositol 5-phosphate 4-kinase; PIKFYVE, phosphatidylinositol 3-hosphate 5-kinase; qHTS, quantitative high-throughput screen; TEA, triethylamine; TFA, trifluoro acetic acid; THF, tetrahydrofuran; TMS, tetramethylsilane; UPLC, ultra-performance liquid chromatography; UV, ultraviolet; X-Phos, 2-Dicyclohexylphosphino-2',4',6'-triisoropylbiphenyl.

4. References

(1) Sivakumaren, S. C.; Shim, H.; Zhang, T.; Ferguson, F. M.; Lundquist, M. R.; Browne, C. M.; Seo, H.-S.; Paddock, M. N.; Manz, T. D.; Jiang, B.; Hao, M.-F.; Krishnan, P.; Wang, D. G.; Yang, J.; Kwiatkowski, N. P.; Ficarro, S. B.; Cunningham, J. M.; Marto, J. A.; Dhe-Paganon, S.; Cantley, L. C.; Gray, N. S. Targeting the PI5P4K lipid kinase family in cancer using novel covalent inhibitors. *bioRxiv*, published online October 25, 2019; DOI: 10.1101/819961.

(2) https://ncats.nih.gov/preclinical/core/chemistry

(2) Seethala, R.; Zhang, L. Handbook of drug screening. In *Informa Healthcare*, New York, 2009.

(3) http://tripod.nih.gov/curvefit/

(4) Zhang, J.-H.; Chung, D. Y.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **1999**, *4*, 67-73.

(5) Inglese, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *PNAS* **2006**, *103* (31), 11473-11478.

(6) Wang, Y.; Jadhav, A.; Southal, N.; Huang, R.; Nguyen, D.-T. A Grid Algorithm for High Throughput Fitting of Dose-Response Curve Data. *Curr. Chem. Genomics* **2010**, *4*, 57-66.

(7) Davis, M. I.; Sasaki, A. T.; Simeonov, A. Method for Assaying the Lipid Kinase Phosphatidylinositol-5-phosphate 4-kinase α in a Quantitative High-Throughput Screening (qHTS) Bioluminescent Format. *Methods Mol. Biol.* **2016**, *1376*, 1-9.

(8) Davis, M. I.; Sasaki, A. T.; Shen, M.; Emerling, B. M.; Thorne, N.; Michael S.; Pragani, R.; Boxer, M.; Sumita, K.; Takeuchi, K.; Auld, D. S.; Li, Z.; Cantley, L. C.; Simeonov, A. A Homogeneous, high-Throughput Assay for Phosphatidylinositol 5-Phosphate 4-Kinase with a Novel, Rapid Substrate Preparation. *PLoS ONE* **2013**, *8* (1): e54127.

(9) Sumita, K.; Lo, Y.-H.; Takeuchi, K.; Senda, M.; Kofuji, S.; Ikeda, Y.; Terakawa, J.; Sasaki, M.; Yoshino, H.; Majd, N.; Zheng, Y.; Kahoud, E. R.; Yokota, T.; Emerling, B. M.; Asara, J. M.; Ishida, T.; Locosale, J. W.; Daikoku, T.; Anastasiou, D.; Senda, T.; Sasaki, A. T. The lipid kinase PI5P4Kβ is an intracellular GTP sensor for metabolism and tumorigenesis. *Mol. Cell.* **2016**, *61* (2), 187-198.

(10) https://www.bellbrooklabs.com/wp-content/uploads/2018/04/Tech-Manual-GDP-FP-v041118-1.pdf

(11) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundbaeck, T.; Nordlund, P.; Molina, D. M. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **2014**, *9* (9), 2100-2122.

5.4 Publication C: *Discovery and Structure-Activity Relationship Study of (Z)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors*

5.4.1 Specific Aims

In another attempt to discover novel PI5P4K inhibitor scaffolds, a high-throughput screen of an in-house kinase inhibitor library, interrogating PI5P4Kα inhibitory activity, was performed. As a result of this screen, CVM-05-002 was identified as a potent PI5P4Kα inhibitor. Consequently, the aim of this study was the characterization of CVM-05-002 as well as the study of its SAR in regard to selectivity and potency. Most importantly, the cellular potency of the scaffold was expected to be poor due to the low LogD value of the compound. Thus, an important goal of the SAR study was the investigation of compounds with higher LogD values, while retaining on-target potency and selectivity.

Besides the design and synthesis of a small compound library of CVM-05-002 analogs based on an available co-crystal structure of the inhibitor with PI5P4Kα, characterization in biochemical and cellular assays was to be performed. Due to the reversible binding mode of the presented compounds as well as their relatively small size, labeling of the scaffold with a detection tag or photoaffinity warhead for the study of cellular target engagement would potentially interfere with the inhibitors' target binding. Thus, the label-free isothermal CETSA methodology was to be applied to confirm direct target engagement of the reversible kinase inhibitors in HEK293T cells.

5.4.2 Main Outcomes and Context

The presented publication shows the discovery, characterization, and optimization of the pan-PI5P4K inhibitor CVM-05-002. While discovered as a PI5P4Kα inhibitor, CVM-05-002 exhibited potent biochemical inhibition of both PI5P4Kα and β isoforms and was quite selective regarding other kinases. Crystallographic analysis confirmed a reversible binding mode of the compound at the ATP-binding site of PI5P4Kα and inspired designs for further CVM-05-002 analogs.

Interestingly, exchange of the (*Z*)-5-methylenethiazolidin-4-one moiety with an indole group (compound **13**) dialed out additional off-targets of the parent compound, while improving the LogD value of the compound from 0.66 for CVM-05-002 to 2.5 for compound **13**. Even though compound **13** loses some biochemical PI5P4K activity, this analog turned out to be a more potent PI5P4K binder in the cellular context.

Taken together, the study introduces and characterizes a novel PI5P4K inhibitor scaffold and provides an attractive chemical probe for the investigation of PI5P4K function in cells and potentially *in vivo*. It also underlines the importance of characterizing a probe in the cellular context in addition to biochemical assays, and thus the value of label-free methods for the study and characterization of reversible protein (kinase) inhibitors.

Details of the chemical synthesis and assay conditions are found in the Experimental section at the end of the main publication.

5.4.3 Published Article

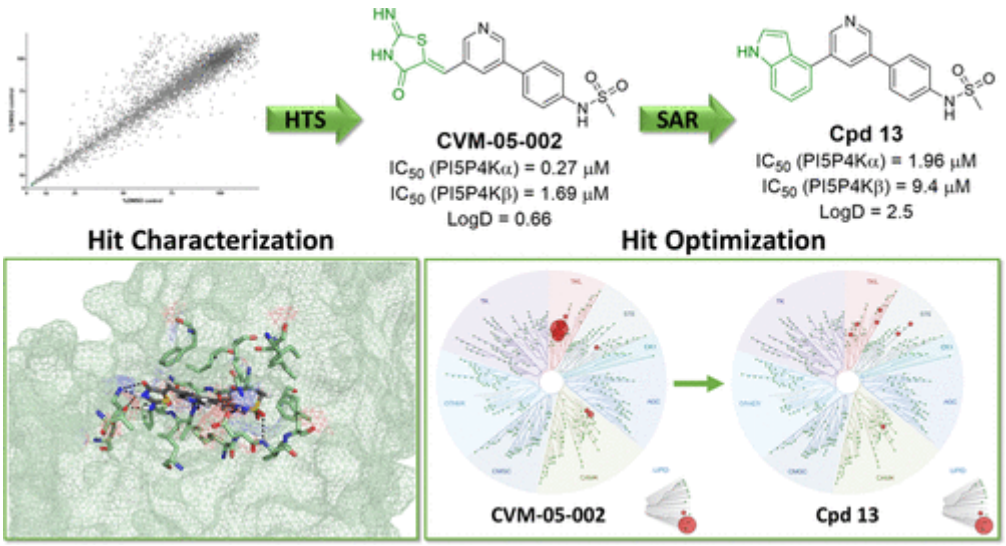
The following peer-reviewed publication was originally published as:

Discovery and Structure-Activity Relationship Study of (*Z*)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz, Sindhu C. Sivakumaren, Fleur M. Ferguson, Tinghu Zhang, Adam Yasgar, Hyuk-Soo Seo, Scott B. Ficarro, Joseph D. Card, Hyeseok Shim, Chandrasekhar V. Miduturu, Anton Simeonov, Min Shen, Jarrod A. Marto, Sirano Dhe-Paganon, Matthew D. Hall, Lewis C. Cantley, and Nathanael S. Gray.

J. Med. Chem. **2020**, *63* (9), 4880-4895.

DOI: https://doi.org/10.1021/acs.jmedchem.0c00227.



This article, including Supporting Information, was reproduced with kind permission from the *Journal of Medicinal Chemistry* © 2019 American Chemical Society.

Journal of Medicinal Chemistry

pubs.acs.org/jmc Article

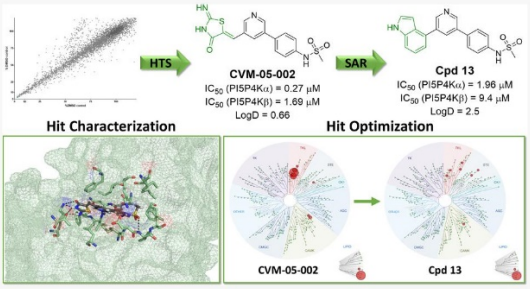
Discovery and Structure–Activity Relationship Study of (*Z*)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz,◆ Sindhu Carmen Sivakumaren,◆ Fleur M. Ferguson, Tinghu Zhang, Adam Yasgar, Hyuk-Soo Seo, Scott B. Ficarro, Joseph D. Card, Hyeseok Shim, Chandrasekhar V. Miduturu, Anton Simeonov, Min Shen, Jarrod A. Marto, Sirano Dhe-Paganon, Matthew D. Hall, Lewis C. Cantley, and Nathanael S. Gray*

Cite This: *J. Med. Chem.* 2020, 63, 4880–4895 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Due to their role in many important signaling pathways, phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) are attractive targets for the development of experimental therapeutics for cancer, metabolic, and immunological disorders. Recent efforts to develop small molecule inhibitors for these lipid kinases resulted in compounds with low- to sub-micromolar potencies. Here, we report the identification of CVM-05-002 using a high-throughput screen of PI5P4Kα against our in-house kinase inhibitor library. CVM-05-002 is a potent and selective inhibitor of PI5P4Ks, and a 1.7 Å X-ray structure reveals its binding interactions in the ATP-binding pocket. Further investigation of the structure–activity relationship led to the development of compound **13**, replacing the rhodanine-like moiety present in CVM-05-002 with an indole, a potent pan-PI5P4K inhibitor with excellent kinome-wide selectivity. Finally, we employed isothermal cellular thermal shift assays (CETSAs) to demonstrate the effective cellular target engagement of PI5P4Kα and -β by the inhibitors in HEK 293T cells.



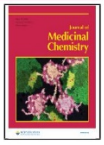
INTRODUCTION

Phosphoinositides, such as phosphatidylinositol 4,5-bisphosphate (PI-4,5-P_2), are key players in many cell-regulating and -signaling processes. Inositide function is controlled by lipid kinases, which alter their phosphorylation pattern, thus generating a range of signaling molecules with different downstream effects. Phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) represent a family of these lipid kinases, which have been shown to phosphorylate phosphatidylinositol 5-phosphate (PI5P) on its 4-position to produce PI-4,5-P_2.[1] In healthy cells, PI5P4Ks play a role in regulating cellular metabolism, stress response, and immunological processes,[2−7] while the abnormal function of these lipid kinases has been linked to diseases, such as diabetes and cancer.[8−11] As a result, PI5P4K inhibitors pose potentially important therapeutic agents, for example, in oncology. To date, several small molecule PI5P4K inhibitors with low- or sub-micromolar potency have been reported in the literature,[12−17] including our recently reported phenylamino pyrimidine-based, covalent pan-PI5P4K inhibitors, THZ-P1-2 and compound **32** (labeled compound "**30**" in the original publication; Supporting Figure 1).[16,17]

Here, we report our efforts to find chemically distinct PI5P4K inhibitors that can serve as chemical probes for further drug development efforts. We focused on developing reversible inhibitors that would not be susceptible to potentially acquired resistance through mutation of the active site cysteine residue, which is modified by irreversible PI5P4K inhibitors.[16,17] A prominent example of this type of resistance mechanism is the EGFR C797S mutation that develops in nonsmall cell lung cancer following treatment with irreversible small molecule EGFR inhibitor AZD9291 and renders tumors resistant.[18]

To find novel, reversible PI5P4K inhibitor scaffolds, we performed a high-throughput screen of our internal kinase inhibitor library, containing approximately 6,000 structurally diverse compounds developed and synthesized in our lab against PI5P4Kα, using an ADP-Glo assay to measure the PI5P4Kα kinase activity. This screen identified potent hit

Received: February 7, 2020
Published: April 16, 2020



Downloaded via HARVARD UNIV on June 4, 2020 at 16:24:35 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © 2020 American Chemical Society

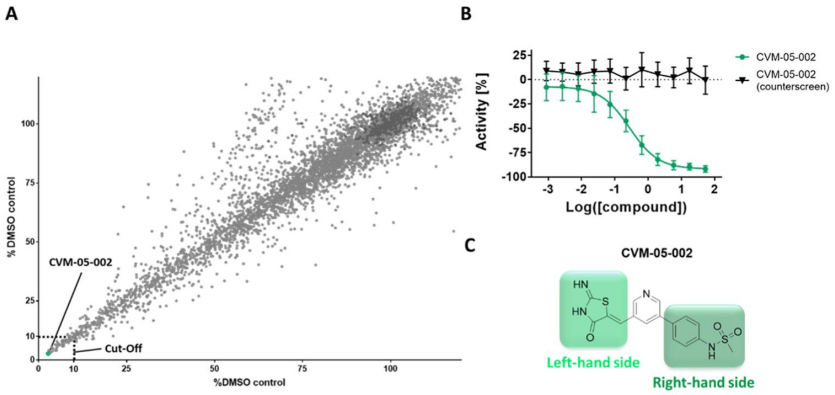


Figure 1. (A) Scatter plot of high-throughput screen for the PI5P4Kα activity of in-house small molecule library with ~6000 compounds (shown are hits at <120% DMSO control). The results of two replicates are plotted against each other (normalized to DMSO). Negative controls (DMSO) are shown in dark gray, CVM-05-002 is depicted in green, and the 10% DSMO control cutoff (90% inhibition) is labeled. (B) Concentration–response curves of CVM-05-002, including counterscreen results. (C) Structure of identified lead compound, CVM-05-002.

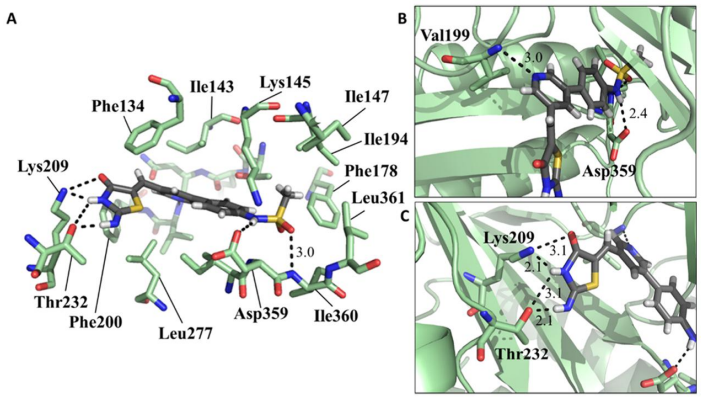


Figure 2. (A) Binding site of cocrystal structure of PI5P4Kα with bound inhibitor CVM-05-002 (dark gray). Key hydrogen bond interactions are depicted by dashed lines (numbers indicate distances in Å). (B) Key hydrogen bond interactions of the hinge-binding site are shown as dashed lines (numbers indicate distances in Å). (C) Proposed hydrogen bond network of the thiazolidinone moiety of CVM-05-002 with Lys209 and Thr232 (interactions are shown as dashed lines; numbers indicate distances in Å). PDB-ID: 6UX9.

compound CVM-05-002 (IC_{50} 270 nM). CVM-05-002 contains a distinct (*Z*)-5-methylenethiazolidin-4-one moiety, which can be described as a rhodanine-like structure. Although rhodanine-type compounds can function as pan-assay interference compounds (PAINS) in AlphaScreen assays,[19−22] in other settings, rhodanines and their derivatives elaborated through focused medicinal chemistry campaigns[23−26] have been advanced into clinical trials.[27] In this study, we describe the discovery, validation, and structure−activity relationship (SAR) of compounds based on this scaffold as potent, specific pan-PI5P4K inhibitors.

■ RESULTS AND DISCUSSION

High-Throughput Screen. Since reliable ADP-Glo-based assays for PI5P4Kα inhibition have been previously described,[12] we adapted this assay to establish a protocol for a miniaturized high-throughput screen. We optimized the assay to perform in a 384-well format with respect to PI5P4Kα enzyme concentration and ADP-Glo reagent volumes to obtain a robust signal-to-background ratio, a stable *Z*′-factor (0.85), and a low coefficient of variation (CV = 0.098) (Supporting Figure 2). Our in-house compound library, consisting of diverse kinase inhibitor scaffolds, was screened at a concentration of 66 μM. This relatively high concentration was chosen based on previous reports describing low potency inhibition of PI5P4K by known kinase inhibitor scaffolds, indicating that the identification of a highly potent inhibitor for this subfamily of lipid kinases may be more challenging than for the protein kinase family. Our screen resulted in the identification of approximately 50 hits exhibiting >90% inhibition (Figure 1A and Supporting Figures 3 and 4). With a mean of 2.7% dimethyl sulfoxide (DMSO) control, CVM-05-002 showed the strongest effects on the screen. All identified hits were analyzed regarding the diversity of the scaffolds, and historical data relating to kinase selectivity was examined. Highly promiscuous or potentially covalent scaffolds were excluded from the set. Based on this, a subset of hit compounds was tested in an 11-point concentration−response experiment, which confirmed CVM-05-002 as the most potent hit for PI5P4Kα inhibitory activity (IC_{50} 0.27 μM; Figure 1B,C). We verified that the PI5P4Kα activity was not due to compound interactions with assay reagents. This was done by

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

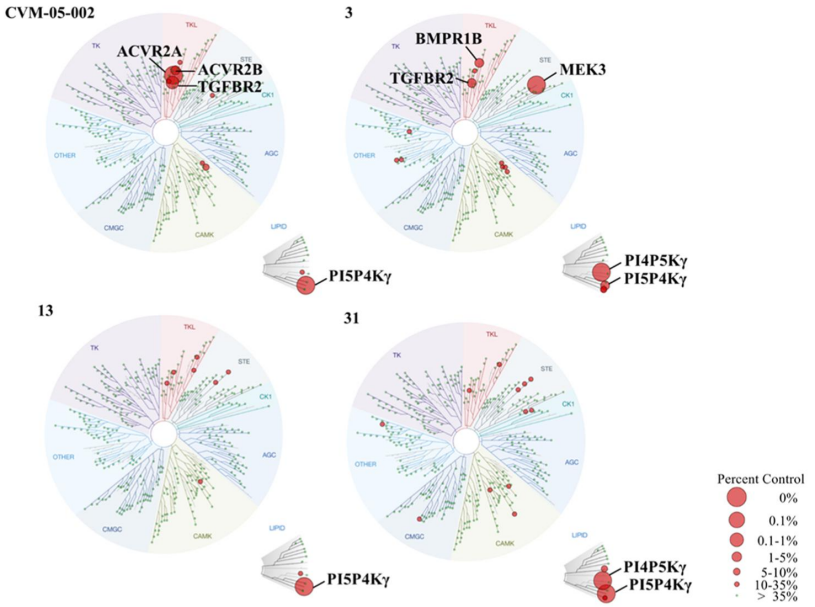


Figure 3. Kinome-wide selectivity profile for compounds CVM-05-002, **3**, **13**, and **31** (excluding mutant, atypical, and pathogen panels). The inhibitors were tested at 1 μM on a panel of 468 kinases. The results are displayed as red circles with their sizes correlating with the inhibitor's binding affinity (percent DMSO control).

counterscreening at a fixed ADP/ATP concentration ratio representing 20% conversion but excluding the PI5P4Kα enzyme, as well as the lipid substrate, confirming no assay interference with CVM-05-002 (Figure 1B).

Interestingly, several structurally similar molecules to CVM-05-002 were present in the library, while only four analogues were found among the top hits identified in the high-throughput screen. Subsequent testing of these hits in our ADP-Glo-based PI5P4Kα assay revealed that these analogues are ~30-fold less potent than CVM-05-002, or more (data not shown). Furthermore, the closest analogues of CVM-05-002, only differing in the 3,5-substituted pyridine moiety of the hit compound, showed strong PIM kinase off-targets (Supporting Table 1). This was not surprising since similar scaffolds have been shown to be potent PIM kinase inhibitors.[26] However, since CVM-05-002 was the only compound of this scaffold without any PIM kinase off-target activity, carrying a unique 3,5-substituted pyridine moiety, while showing potent on-target activity, we were highly motivated to further optimize this screening hit and explore its SAR.

Finally, knowing that the scaffold of CVM-05-002 could potentially have nonspecific effects due to its similarity to compounds with reported reactivity,[22] we performed mass spectrometry (MS) analysis of recombinant PI5P4Kα and -β, incubated with CVM-05-002. PI5P4Kα showed no detectable reaction after 2 h, while the β-isoform of the lipid kinase was modified only at a very low level (~14%). Under the same conditions, our previously reported covalent PI5P4K inhibitor, THZ-P1-2,[17] labeled ~63−87% of both isoforms (Supporting Figures 5 and 6).

Structural Basis of the CVM−PI5P4K Interaction. Recombinant PI5P4Kα was cocrystallized in complex with CVM-05-002 at a 1.7 Å resolution, confirming a noncovalent binding mode of the inhibitor in the active site of the lipid kinase (Figure 2) (PDB-ID: 6UX9).

The obtained cocrystal structure revealed general hydrophobic interactions of CVM-05-002 with the ATP-binding pocket. For example, Phe178, Ile147, Ile194, and Leu361 form a nonpolar pocket encasing the methyl group of the sulfonamide (Figure 2A). In addition, five key hydrogen-bonding interactions could be identified. At the hinge, the pyridine nitrogen anchors the inhibitor *via* a hydrogen bond with the backbone of Val199 (Figure 2B). The thiazolidinone moiety interacts with the side chains of Thr232 and Lys209 on the left-hand side of the scaffold, while the sulfonamide engages in two hydrogen-bonding interactions with the backbone of Ile360 and the side chain of Asp359 on the right-hand side (Figure 2A). Furthermore, Asp359 and Lys145 form a salt-bridge, potentially improving the hydrogen-bonding interaction with the sulfonamide-NH. Likewise, it should be noted that the thiazolidinone moiety is positioned in a way that permits the formation of hydrogen-bonding interactions of the imine, the carbonyl, and the cyclic-NH groups with Thr232 and Lys209, together forming a fork-like hydrogen bond network on the left-hand side of the inhibitor with the binding site. While the interaction of the pyridine moiety seemed to be essential for the selectivity of the molecule (Supporting Figure 1), the strong left- and right-hand-side hydrogen bond interactions are likely contributing to the relatively high potency of this scaffold (Figure 2A,C).

Selectivity Screening. The kinome-wide binding selectivity of CVM-05-002 at a 1 μM concentration was determined, using the commercial KINOMEscan profiling platform, consisting of 468 kinases. CVM-05-002 showed an excellent selectivity profile (Figure 3; selectivity score for hits with a percent control < 35, S_{35} = 0.02), providing further evidence that this inhibitor is a selective and potent pan-PI5P4K

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

inhibitor. Taken together, these findings were encouraging to pursue further medicinal chemistry efforts to better understand the SAR around this promising hit.

Chemistry. Given the above, the presented SAR study focuses mainly on changes regarding the thiazolidinone and the *N*-phenylmethanesulfonamide groups, decorating the hinge-binding pyridine moiety and being referred to as left- and right-hand sides of the scaffold, respectively (Figure 1C). Hence, two main series of CVM-05-002 analogues were designed as summarized in Tables 1 and 2, with changes made on either side of the molecule.

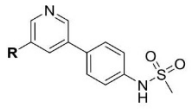
In general, CVM-05-002 and analogues were synthesized as summarized in Scheme 1, starting from a 3-bromopyridine analogue (**1**, **11**, **16**, **28a,b**) and by introducing different aromatic ring systems *via* Suzuki couplings (**2a−c**, **6−8**, **12a−c**, **13−15**, **17a−j**, and **29a,b**). (*Z*)-Methylenethiazolidinone and its analogues were accessed from nicotinaldehyde derivatives, which were coupled with either 2-iminothiazolidin-4-one, oxazolidine-2,4-dione, or maleimide, using different Knoevenagel condensation reaction conditions (CVM-05-002, **3−5**, **18−27**, and **30−31**). Compounds **9** and **10** were yielded *via* S_N2 reactions with different 3-substituted *N*-phenylacetamide analogues.

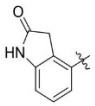
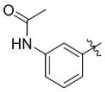
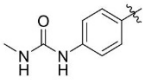
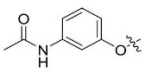
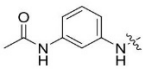
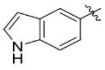
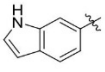
Biological Evaluation. The biochemical inhibition of PI5P4Kα and -β was determined as previously described.[12,17] Due to its low enzymatic activity,[28,29] we were unable to determine the PI5P4Kγ activity for our compounds; thus, this discussion focuses solely on the PI5P4Kα and -β isoforms. It should be noted that the structural alignment of all three PI5P4Ks confirmed that the binding pocket is highly conserved within this lipid kinase family (Supporting Figure 7). The amino acids, highlighted in Figure 2, are identical among all three isoforms, with the exception of Ile147 and Ile194 in PI5P4Kα, which are being replaced by valines in PI5P4Kβ or a valine and a leucine, respectively, in PI5P4Kγ. Val199 in PI5P4Kα is a methionine in PI5P4Kγ, which should not affect the hinge binding of CVM-05-002, since the interaction of the pyridine is with the backbone of the amino acid.

In this biochemical evaluation, our initial hit compound, CVM-05-002, showed consistent IC_{50} values of 270 nM and 1.69 μM for PI5P4Kα and -β, respectively, hence representing—to the best of our knowledge—the most potent reported pan-PI5P4K inhibitor to date. However, CVM-05-002 is a polar compound with a log *D* value of 0.66, which limits its use in cells and, subsequently, as a useful probe to study PI5P4Ks *in vivo*. The balance of lipophilicity and PI5P4K inhibitory activity was therefore another crucial criterion of this SAR study.

With a first set of compounds, summarized in Table 1, we introduced different substituents on the left-hand side of the scaffold, altering or replacing the methylenethiazolidinone moiety. For example, changing the imine of the thiazolidinone to a carbonyl (compound **3**) slightly improved the inhibitory activity with respect to PI5P4Kα by about 2-fold to an IC_{50} of 133 nM, while the β-activity did not change. This is not surprising since the carbonyl-O and the imine-NH can both function as hydrogen bond acceptors. However, an oxazolidine-2,4-dione in compound **4** led to an inactive compound on both isoforms, highlighting the importance of the electrochemical properties of this five-membered ring system for the PI5P4K activity, since the electronegativity of oxygen (compound **4**) is much higher as compared to that of sulfur (compound **3**). This is further supported by the fact that introducing a carbon (compound **5**), having similar electronegativity as sulfur, in position of the oxygen (compound **4**) led to a moderately active compound with IC_{50}s of 6.1 and 27

Table 1. IC_{50} Evaluation of Compounds CVM-05-002, 3−15[a]

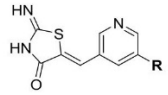


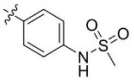
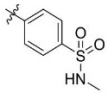
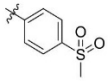
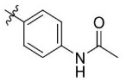
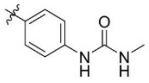
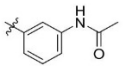
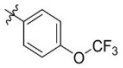
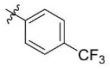
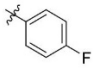
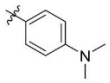
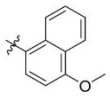
Cpd	R	IC_{50} [μM][a] PI5P4K -α	n	-β	n	LogD
CVM-05-002		0.27 ± 0.035	3	1.7 ± 0.50	3	0.66
3		0.13 ± 0.032	3	1.9 ± 0.42	3	-0.05
4		21 ± 1.4	3	>50	2	-1.0
5		6.1 ± 0.0	3	27 ± 13	2	-0.14
6		8.9 ± 2.9	3	15 ± 9.0	3	1.5
7		5.3 ± 0.72	3	36 ± 4.1	3	1.7
8		3.7 ± 0.24	3	43 ± 21	2	1.6
9		9.0 ± 0.61	3	25 ± 8.9	3	1.5
10		14 ± 0.96	3	45 ± 28	2	1.5
13		2.0 ± 0.45	3	9.4 ± 7.1	2	2.5
14		2.1 ± 0.53	3	22 ± 8.4	2	2.5
15		3.2 ± 0.21	3	29 ± 4.1	2	2.5

[a]Average ± SD of n = 2 or 3 experiments, each using an 11-point titration.

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

Table 2. IC_{50} Evaluation of Compounds 18–27[a]



Cpd	R	IC_{50} [μM][a] PI5P4K -α	n	-β	n	LogD
CVM-05-002		0.27 ± 0.035	3	1.7 ± 0.50	3	0.66
18		1.4 ± 0.18	3	4.3 ± 1.1	2	1.1
19		35 ± 57	3	>50	2	1.1
20		11 ± 4.2	3	>50	3	1.5
21		1.8 ± 0.12	3	8.6 ± 1.9	2	1.4
22		11 ± 0.76	3	>50	3	1.5
23		11 ± 5.2	3	>50	2	3.7
24		>50	3	26 ± 3.3	3	3.2
25		>50	3	28 ± 3.2	3	2.4
26		8.8 ± 5.8	3	>50	2	2.4
27		3.9 ± 0.82	3	>50	2	3.1

[a]Average ± SD of n = 2 or 3 experiments, each using an 11-point titration.

μM for PI5P4Kα and -β, respectively. It should be noted that this equals a 1–2-log-fold drop in the PI5P4K activity of compound **5** compared to that of compound **3**, pointing toward an essential role of the cyclic sulfur for the PI5P4K activity of the scaffold. One major difference between the five-membered ring systems of compounds **3** and **5** is the aromaticity. Whereas compound **3** shows aromatic features due to the lone electron pairs of the sulfur, compound **5** introduces an sp^3-hybridized carbon atom, disrupting the aromatic π-system of the ring.

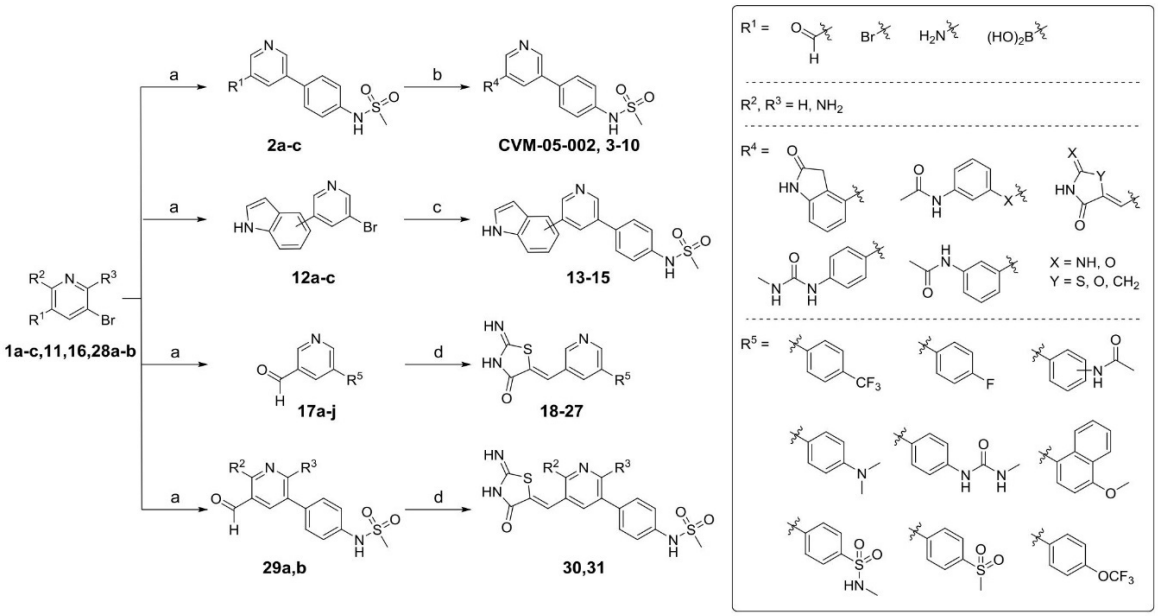
All compounds mentioned above have lower log D values than CVM-05-002, with compounds **3** and **5** having a log D of around 0, while compound **4** has the lowest log D value of the set with −1. Consequently, we next focused on replacing the entire methylenethiazolidinone moiety to improve lipophilicity of the inhibitors. Strategies for this replacement were inspired by the cocrystal structure, which suggests that the (Z)-5-methylenethiazolidin-4-one could be cyclized, by linking the methylene-H with the cyclic carbonyl to form an aromatic six-membered ring. An initial version of this is compound **6**, carrying an indolinone moiety on the left-hand side. This structure is closely related to compound **5** since the five-membered ring systems of both compounds only differ in the carbonyl oxygen in the 2-position, which is missing in compound **6**. Here, an aromatic six-membered ring, directly connecting to the pyridine, was introduced instead. The compound showed IC_{50}s of 8.9 and 15 μM for PI5P4Kα and -β, respectively, which is comparable to compound **5**'s PI5P4K activity, while improving log D by 1.5-log-folds to 1.5. Interestingly, the loss of the carbonyl interaction did not seem to affect the overall PI5P4K activity, which might indicate a positive effect of the aromatic six-membered ring system of compound **6** for the PI5P4K activity, potentially compensating for the loss of the carbonyl hydrogen bond interaction.

Due to the overall connected π-electron system in compound **6**, the molecule is likely rather flat, as compared to CVM-05-002, which shows a slight bend at the methylene linker between the pyridine and thiazolidinone moiety in the cocrystal structure (Figure 2). Thus, the left-hand side residues with higher flexibility, especially in the hydrogen-bond-forming groups to compensate for this, were of interest to us. Ring opening of the pyrrolidine-2-one ring, resulting in a more flexible amide in compound **7**, led to minor improvements in the PI5P4Kα activity while showing a slight drop in -β activity with IC_{50}s of 5.3 and 36 μM, respectively. This trend was further pronounced in compound **8**, introducing a methylated urea motif in the para-position of the aromatic ring system, with IC_{50}s of 3.7 μM for PI5P4Kα and 43 μM for PI5P4Kβ. Compound **8** offers more potential hydrogen bond interactions than the acetamide of compound **7**, which could be an explanation for its slightly higher potency on the α-isoform of the lipid kinase. It should be noted, however, that, even though there seems to be a drop in the PI5P4Kβ activity of over 2-fold from compound **6** to compound **8**, the relatively high standard deviation values are causing the differences in IC_{50} values to be statistically nonsignificant.

Introduction of an O- or NH-linker to compound **7** (compounds **9** and **10**, respectively) is further increasing the flexibility of the acetamide to form hydrogen bond interactions with the lipid kinase. Both compounds showed a slight loss in the PI5P4Kα activity by about 2–3-fold compared to compound **7**, while there was no significant difference in the PI5P4Kβ activity. In summary, comparing the overall lipophilic efficiency (LipE) of these molecules, the flatness of the SAR for compounds **6–10** becomes obvious. LipE is an important feature in drug discovery since it sets the lipophilicity of a molecule in relation to its activity, here, log D to pIC_{50}. Since compounds **6–10** have very similar log D values, plotting against pIC_{50} shows that the compounds depict almost identical PI5P4Kα and -β activity profiles with LipE values between 3 and 4 for PI5P4Kα and of values of about 3 for PI5P4Kβ (Supporting Figure 10), clearly showing no real improvement of the PI5P4K activity by more flexible hydrogen-bond-offering moieties. These findings indicate a

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880–4895

Scheme 1. Synthesis of Compound CVM-05-002 and Analogues[a]



[a]Reagents and conditions: (a) phenylboronic acids/respective pinacol boronic acid esters, $Pd(PPh_3)_2Cl_2$, Cs_2CO_3, 1,4-dioxane/H_2O, 100 °C, 1.5 h, 9−93% yield; (b) (*Z*)-methylenethiazolidinone derivative, *β*-alanine, AcOH, 100 °C, 0.5 h, 12−33% yield, or maleimide, PPh_3, dichloromethane (DCM), rt, overnight, 21% yield, or oxazolidine-2,4-dione, piperidine, EtOH, 100 °C, overnight, 4% yield, or phenylboronic acids/respective pinacol boronic acid ester, $Pd(PPh_3)_2Cl_2$, Cs_2CO_3, 1,4-dioxane/H_2O, 100 °C, 1.5 h, 1−37% yield, or *N*-(3-hydroxyphenyl)acetamide, CuI, K_2CO_3, *N*,*N*-dimethylformamide (DMF), 120 °C, 2 d, 3% yield, or *N*-(3-bromophenyl)acetamide, $Pd_2(dba)_3$, X-Phos, K_2CO_3, 1,4-dioxane/H_2O, 80 °C, overnight, 7% yield; (c) 4-(methanesulfonylamino)phenylboronic acid pinacol ester, $Pd(PPh_3)_2Cl_2$, Cs_2CO_3, 1,4-dioxane/H_2O, 100 °C, 1.5 h, 19−27% yield; (d) 2-iminothiazolidin-4-one, *β*-alanine, AcOH, 100 °C, 0.5 h, 4−22% yield.

limited benefit of the hydrogen bond interactions in Figure 2C while highlighting the gained PI5P4K activity through the introduced aromatic ring system.

When further simplifying the bicycle of compound **6** by introducing an indole on the left-hand side of the molecule (compound **13**), the activity on both PI5P4K isoforms could be improved to low micromolar IC_{50}s. While the slight improvement in the PI5P4K*β* activity is not significant, the PI5P4K*α* activity of compound **13** improves significantly to 2.0 *μ*M (*P* value = 0.049 by *t*-test) as compared to that of compound **6**. This further supports our findings that the aromaticity of the bicyclic system is more important than the hydrogen-bonding interaction by the lactam. Hence, removing the carbonyl of compound **6**, while increasing the aromaticity across both rings of the bicyclic moiety, did not result in loss in the potency of the compound, but actually improved the potency on PI5P4K*α*. Consequently, the higher lipophilicity with a log *D* value of 2.5, which is an improvement of over 1-log-fold as compared to that of compounds **6**−**10**, as well as the enlarged aromatic ring system of compound **13**, is beneficial for the PI5P4K*α* activity, but also likely improves the properties of the compound for the use in cells.

We assume that this benefit of aromatic moieties on the left-hand side could be indicating *π*−*π* interactions with either Phe200 or Phe134 (Figure 2A), which we had previously identified as a crucial interaction for our reported covalent pan-PI5P4K inhibitor series.[17] This was supported by the fact that further changes in the attachment site of the indole from the 4-position (compound **13**) to the 5- or 6- position (compounds **14** and **15**, respectively) led to no change in the PI5P4K*α* activity and only a minor 2−3-fold decrease in the *β*-activity. Even though compounds **13**−**15** show, thus, a very similar PI5P4K activity profile, compound **13**, the isomer closely related to compound **6**, showed the best-combined PI5P4K*α* and -*β* activities. As already mentioned, since all three compounds have improved log *D* values of 2.5, leading to LipE values of around 3 for PI5P4K*α* and values of 2−2.5 for PI5P4K*β*, the compounds should show a good cellular target engagement, considering the improved balance of lipophilicity and biochemical activity as compared to the initial hit compound, CVM-05-002 (Supporting Figure 10A,B).

In a next step, the SAR on the right-hand side of CVM-05-002 was investigated by introducing aromatic ring systems, carrying different substituents ranging from polar sulfonamides, amides, and urea motifs to nonpolar and bulkier moieties (Table 2). The sulfonamide moiety in CVM-05-002 significantly contributes to the compound's high hydrophilicity, which inspired designs that aim to create compounds of higher lipophilicity, while at the same time interrogating the necessity of the sulfonamide of the parent compound for the PI5P4K activity.

Compound **18**, a reversed sulfonamide analogue of CVM-05-002, lost activity against both PI5P4K isoforms by about 1-log-fold, most likely due to the changed hydrogen bond donor/acceptor configuration. Removing the sulfonamide-NH in compound **19** kept the log *D* of the compound comparable to that in compound **18** at 1.1; however, the overall polar surface area (PSA) of the compound decreased from 112 for

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

CVM-05-002 and compound **18** to 100 for compound **19**. Interestingly, losing this additional NH hydrogen-bonding interaction led to a complete loss of activity on PI5P4Ks. These results might indicate that the sulfonamide moiety of CVM-05-002 is crucial for the optimal PI5P4K activity.

Switching from a sulfonamide to an acetamide (compound **20**) increased the lipophilicity by about 1-log-fold, compared to that of CVM-05-002, to a log *D* value of 1.5. Nevertheless, the trigonal-planar conformation of the amide, in contrast to the tetrahedral conformation of the sulfonamide, led to an activity decrease of 2-log-folds or higher on the PI5P4K isoforms. The configuration of the amide in compound **20** makes it impossible for the hydrogen of the NH and the oxygen-O to point into the same direction, which is necessary to form both hydrogen bond interactions seen for CVM-05-002 with Asp359 and Ile360. This is further supported by compound **21**, carrying a methylated urea motif and showing improved biochemical potency on both PI5P4K isoforms to levels comparable to the reversed sulfonamide analogue **18**, while having a higher log *D* of around 1.4. The additional NH-moiety in compound **21** is likely forming similar interactions to the NH group of compound **18**, which would explain the similar potency, as well as the lack in the potency of compound **20**. Changing the acetamide from *p*- to *m*-position (compound **22**) did not influence the PI5P4K activity, underlining a certain tolerance in this substituent. The meta-position could either point toward the deeper pocket, or toward the thiazolidinone moiety, which would mean that both the NH and the carbonyl-*O* could potentially form hydrogen bond interaction with Asp359 or Ile360.

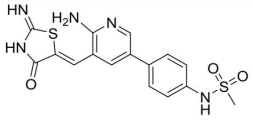
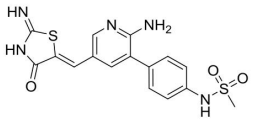
Interestingly, compound **23**, carrying a trifluoromethoxy substituent, exhibited the same activity profile as compound **20** with weak PI5P4Kα (11 μM) and no PI5P4Kβ activity. This points toward a tolerance for hydrogen bond donors, as well as acceptors for PI5P4Kα in this position, which is in accordance with the cocrystal structure, since Lys145 is in reach to form hydrogen bond interactions with the methoxy-*O* (2.4 Å). Removing the hydrogen bond acceptor, resulting in compound **24**, supports this hypothesis by showing the loss of PI5P4Kα activity. In addition to that, compounds **23** and **24** exhibit good lipophilicity for cellular treatments with log *D* values of 3.7 and 3.2, respectively, while the matched methoxy or methyl substituents would have had log *D* values of only 1.4 or 1.9 instead. Surprisingly, compound **24** gained back some PI5P4Kβ activity with an IC_{50} of 26 μM. However, this β-activity does not seem to be influenced by the trifluoromethyl moiety since a simple fluoro substituent (compound **25**) led to the same activity profile. Furthermore, comparing compounds **23** and **25**, it seems that the polar NH- or O-groups adjacent to the aromatic ring are beneficial for the PI5P4Kα activity, whereas it is unfavorable in the context of PI5P4Kβ activity, which is in contrast to the high similarity of the two binding sites (Supporting Figure 7).

Similarly, compounds **26** and **27** introduce either a dimethylamine or a methoxy group, respectively, in the para-position of the phenyl moiety on the right-hand side. These substituents pose hydrogen bond acceptors, leading to the activity pattern already seen above with PI5P4Kα, but no -β activity. Interestingly, the bulky naphthalene of compound **27** was well tolerated by PI5P4Kα and even led to a slight increase in IC_{50} (3.9 μM) as compared to that of compound **23** (11 μM). Despite the desirable lipophilicity of these two molecules with log *D* values of 2.4 and 3.1 for compounds **26** and **27**, respectively, the lack of PI5P4Kβ activity, as seen with other compounds of this batch of CVM-05-002 analogues, makes them unfit as pan-PI5P4K inhibitor probes. Taken together, it appears that trying to increase lipophilicity by changing the right-hand side of the molecule led to either poor PI5P4Kα or -β inhibitors, while even slight changes in the sulfonamide moiety of CVM-05-002 led to a drop in PI5P4K activity of at least 1-log-fold. These findings suggest that the sulfonamide of CVM-05-002 is indeed necessary to maintain both PI5P4Kα and -β activities.

In summary, the SAR analysis on the left- and right-hand sides of the PI5P4K inhibitor scaffold revealed that the key interactions seen in the cocrystal structure of PI5P4Kα with CVM-05-002 seem to be essential for the potency of the inhibitor. Removal of hydrogen bond donors or acceptors on either side of the molecule generally led to a minimum of 1-log-fold decrease in potency. Interestingly, replacing the (*Z*)-methylenethiazolidinone moiety with an aromatic bicyclic indole (compound **13**) retained good potency comparable to our previously reported pan-PI5P4K inhibitors (Supporting Figure 1),[17] while improving the lipophilicity of the scaffold by about 2-log-folds.

Although we had shown that the 3,5-substituted pyridine is likely essential for the CVM-05-002's kinome-wide selectivity (Supporting Table 1), we had noticed a small pocket extending from the 2-position of the pyridine at the hinge-binding region (Supporting Figure 8). Hence, we designed and synthesized the additional compounds **30** and **31** with an amino group at either the 2- or 6-position of the central pyridine of CVM-05-002. According to the cocrystal structure, the amino group of compound **30** should cause a steric clash with the protein surface at the hinge, whereas compound **31** should be able to fit the amino group into the adjacent pocket, potentially causing a shift in the compound's selectivity profile. The polar amino group was chosen since the backbone carbonyl of Arg197, as well as the side-chain hydroxyl group of Thr196, is in close proximity to the pyridine with distances under 4 Å (Supporting Figure 8). 2-Aminopyridine moieties are also well-known hinge-binding motifs of kinase inhibitors since they mimic the adenine binding interactions of ATP.[30−32] As expected, compound **30** lost activity on both, PI5P4Kα and -β, by almost 1.5-log-folds with IC_{50}s of 4.0 and 25 μM, respectively (Table 3). In contrast to this 15-fold decrease in

Table 3. IC_{50} Evaluation of Compounds 30 and 31[a]

Cpd	Structure	IC_{50} [μM][a] PI5P4K -α	PI5P4K -β	Log D
30		4.0 ± 0.73	25 ± 2.9	0.41
31		0.66 ± 0.040	2.7 ± 0.17	0.41

[a]Average ± SD of three experiments, each using an 11-point titration.

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

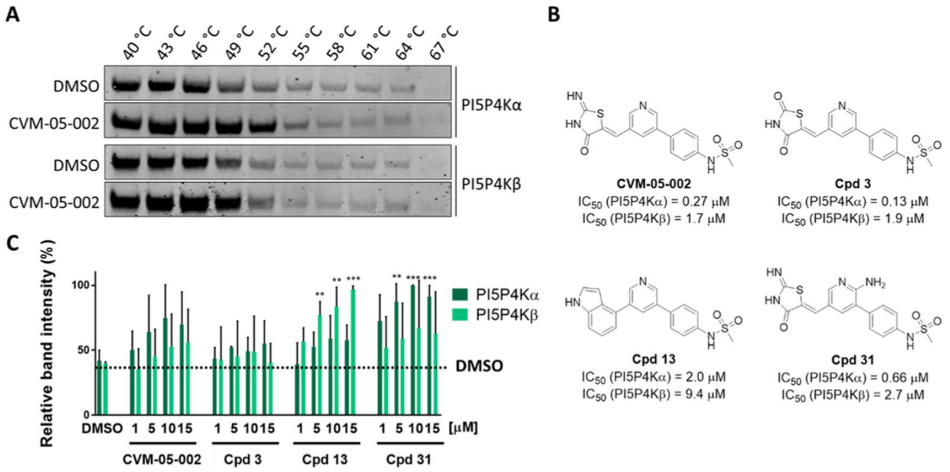


Figure 4. (A) Western blot of the melting curve determination of PI5P4Kα and -β. HEK 293T cells were either treated for 1 h with DMSO or 10 μM CVM-05-002. Representative Western blot of two independent repeats. (B) Structures of selected compounds of the presented SAR study and their biochemical activities. (C) Quantified relative band intensity (%) of the isothermal CETSA experiment (Supporting Figure 10). The mean of four independent experiments is shown. ** $p < 0.01$, *** $p < 0.001$ by t-test.

potency, compound **31** showed only a very minor decrease in potency with 660 nM and 2.7 μM potency on PI5P4Kα and -β, respectively. Naturally, both compounds are more polar than the parent inhibitor, CVM-05-002, with log D values of 0.4.

Selectivity Profiling. Selected compounds of our series were tested on a set of 468 kinases in Ambit's commercially available KINOMEscan platform. Besides our most potent pan-PI5P4K inhibitors (CVM-05-002, as well as compounds **3** and **31**), compound **13**, with the best balance between lipophilicity and biochemical activity, was selected as well. The compounds' binding activity was determined at 1 μM as percent DMSO control and is summarized in Figure 3. It should be noted that PI5P4Kα was not available in this profiling panel.

Our initial hit compound, CVM-05-002 (selectivity score for hits with a percent control < 35; S_{35} = 0.02), showed high selectivity among the kinases tested, with only 10 hits at a cutoff of 35% DMSO control, such as activin A receptor type 2A (ACVR2A) and TGFBR2, both of which are members of the TGFβ family receptors.[33−35] Within the lipid kinases, PI5P4Kγ is the only strongly bound target, suggesting excellent selectivity of the inhibitor within the lipid kinase families. PI5P4Kβ did not show up as a hit in this assay, potentially due to the weaker activity of CMV-05-002 toward this isoform. Consequently, we do not believe that the assay is a good predictor for identifying PI5P4Kβ activity in the micromolar range at this concentration, but it is still very useful to identify potential off-targets. This is true for all of the analogues tested in this assay format.

Surprisingly, even though compound **3** is similarly potent to CVM-05-002 in biochemical assays, the change from the 2-iminothiazolidin-4-one to the thiazolidine-2,4-dione led to a slight selectivity shift of the inhibitor. With a similar number of off-targets (S_{35} = 0.03), compound **3** was not more promiscuous than CVM-05-002; however, its activity shifted from ACVR2A toward MEK3. More importantly, among the lipid kinases, a major off-target, phosphatidylinositol 4-phosphate 5-kinase γ (PI4P5Kγ), showed strong binding affinity to compound **3**, while losing affinity for PI5P4Kγ. We further confirmed potential off-targets by commercially available biochemical activity assays (Supporting Table 2). This data validated that CVM-05-002 and compound **3** have potent off-target activity for activin A receptor type 2B (ACVR2B) and TGFBR2. Compound **3** showed slightly stronger PI4P5Kγ activity as compared to CVM-05-002 by almost 2-fold, which is not as pronounced as in the Ambit KINOMEscan profiling, nonetheless confirming this lipid kinase off-target. As mentioned before, there are no kinase activity assays for PI5P4Kγ; however, we have compared the binding affinity using K_D values (Supporting Table 2). This data supports our findings that CVM-05-002 is a strong binder of PI5P4Kγ with a K_D of 1.6 nM. Interestingly, compound **3** showed a slightly stronger K_D. It should be noted that we have added our previously reported pan-PI5P4K inhibitor, THZ-P1-2, as a positive control for PI5P4Kγ binding (K_D of 4.8 nM), where we have additionally shown covalent engagement of the lipid kinase.[17]

Comparing the selectivity profile of CVM-05-002 and compound **13**, replacing the (*Z*)-methylenethiazolidinone moiety with the bicyclic indole, the change on the left-hand side of the scaffold was able to dial out all seen off-targets of CVM-05-002 and resulted in a highly selective compound with a strong affinity to PI5P4Kγ (S_{35} = 0.02), confirmed with a K_D of 3.4 nM. The biochemical activity assays further showed lost off-target activity for ACVR2A/B and TGFBR2 of almost 2-log-fold changes or more. Likewise, the lipid kinase off-target PI4P5Kγ activity dropped by 4-fold. In contrast to that, even though we were able to gain selectivity against protein kinases with the additional hinge-binding amino group in compound **31**, this inhibitor became a pan-lipid kinase binder (S_{35} = 0.03), with a strong affinity for PI4P5Kγ, 200-times more active on this lipid kinase than compound **13**. Interestingly, this compound exhibited the strongest K_D for PI5P4Kγ with 23 pM.

Cellular Target Engagement. To determine the cellular target engagement of our inhibitors, we established a cellular thermal shift assay (CETSA) protocol, which was previously described.[17] HEK 293T cells were treated with either DMSO

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

or 10 μM of CVM-05-002 for 1 h. The collected cells were then aliquoted and treated at temperatures ranging from 40 to 67 °C. Finally, the treated samples were lysed and the soluble protein fraction was analyzed *via* Western blot. A representative Western blot is shown in Figure 4A, indicating that 49 °C is a suitable temperature for further isothermal CETSA experiments since the inhibitor-treated cells show a strong signal for both PI5P4Kα and -β, while the DMSO-treated cells show a visible loss in soluble PI5P4K kinases in comparison. This temperature was, hence, selected for the following isothermal CETSA experiments.

Summarized in Figure 4C are four independently performed isothermal CETSA experiments with the previously selected four PI5P4K inhibitors (Figure 4B and Supporting Figure 4). Compared to DMSO, all compounds, apart from compound **3**, show at least a slight stabilization of both PI5P4K isoforms in HEK 293T cells with increasing compound concentration. Although CVM-05-002 shows a visible stabilization of PI5P4Ks, the results are not statistically significant, which is in accordance with the compound's low permeability in a PAMPA assay setup (Supporting Table 4). This was to be expected due to the compound's low log D value of 0.66. It should, however, be noted, that the CETSA assay results were rather noisy, leading to relatively high error bars, as seen in Figure 4C, which in turn makes it difficult to get statistically significant results, even though the increase in soluble protein is visible in the Western blot, indicating some target engagement in HEK 293T cells. This lies in the nature of the assay itself since the outcome can be influenced by variations in the cell density of each repeat or within well plates, the duration of heating and cooling periods, and other factors.[36] Nonetheless, the CETSA assay remains a powerful tool to confirm the overall target engagement in cells.

As expected, compound **13** showed excellent permeability in the PAMPA assay (Supporting Table 4), which we were able to confirm in HEK 293T cells. In contrast to CMV-05-002 and compound **3**, compound **13** was able to cause statistically significant stabilization of soluble PI5P4Kβ ($p < 0.01$ by t-test). The visible but only slight effect of compound **13** on PI5P4Kα underlines that the biochemical PI5P4Kα and -β assays are not necessarily translating in cells since the IC_{50} values for PI5P4Kα were generally lower than those for the β-isoform. That said, compound **13** seemed to have a better stabilizing effect on PI5P4Kβ in HEK 293T cells.

Somewhat surprisingly, compound **31**, only differing from CVM-05-002 in an additional hinge-binding amino group, showed strong cellular stabilizing effects of PI5P4Kα ($p < 0.001$ by t-test). This was unexpected since the compound is even more polar than the initial hit compound and is consequently showing similarly bad permeability properties as CVM-05-002 (Supporting Table 4). Since all compounds tested showed good microsome stability (Supporting Table 4) and PAMPA assays do not include active cellular uptake or excretion, nor do we know the effects of the off-target binding of the inhibitors on the PI5P4K activity in cells, it is hard to explain this strong cellular effect of the compound. It should also be noted that PI5P4Ks are able to heterodimerize,[37] meaning the stabilization of PI5P4Kγ may stabilize PI5P4Kα through protein−protein interaction-mediated mechanism, further complicating the detailed interpretation of the CETSA results. We therefore focused on the general activity of the compounds in cells as a measure of the ability to function as a probe in the cellular context to investigate PI5P4Ks.

CONCLUSIONS

In this study, we have discovered a novel pan-PI5P4K inhibitor, CVM-05-002, with low- to sub-micromolar potency based on a thiazolidinone scaffold, using high-throughput screening methods of our in-house kinase inhibitor library. Even though the rhodanine-like scaffold can exhibit PAINS behavior, we were able to show that the identified hit compound is nonreactive, and a highly potent and selective ATP-site-directed inhibitor of the PI5P4K kinases in biochemical and cellular assays. A cocrystal structure of the initial hit with PI5P4Kα was used for the structure-guided elucidation of the structure−activity relationships to design further analogues based on this newly discovered PI5P4K-binding scaffold and explore its SAR. As a result, we were able to design compound **13**, which lacks the thiazolidinone chemotype and is a low micromolar PI5P4K inhibitor with excellent selectivity and cellular potency, posing a useful probe to study PI5P4Ks. In summary, we present not only a valuable new scaffold of PI5P4K inhibitors but also provide another example of the significance of rhodanine-based scaffolds for medicinal chemistry campaigns, when coupled with careful validation experiments.

EXPERIMENTAL SECTION

Chemistry Experimental. *General Methods.* Solvents and all chemicals were purchased from commercial sources and used without further purification. Purifications of intermediates and final compounds were performed, using a preparative high-performance liquid chromatography (HPLC) with a Waters Sunfire C18 column (19 mm × 50 mm, 5 μM) and a gradient of 15−95% MeOH in water (0.05% trifluoroacetic acid (TFA)) over 22 min (28 min total run time) at a flow rate of 20 mL/min. Mass spectra data were obtained on a Waters Acquity ultrahigh-pressure liquid chromatography (UPLC), which was also used to determine the purity of final compounds and is noted for each final compound individually. Proton (^{1}H) and carbon (^{13}C) NMR spectra were recorded on a 500 MHz Bruker Avance III spectrometer (500 MHz for ^{1}H; 126 MHz for ^{13}C). Chemical shifts in ppm (δ) downfield from TMS are reported. Coupling constants (J) are reported in Hz, and spin multiplicities are noted as s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublet), or m (multiplet).

General Procedure A: Suzuki Coupling. Boronic acids or respective boronic acid esters (1.1 equiv), bromobenzenes (1.0 equiv), and Cs_2CO_3 (5.0 equiv) were dissolved in a mixture of 1,4-dioxane and water (9:1). The mixture was purged with N_2 for about 10 min. $Pd(PPh_3)_2Cl_2$ (0.1 equiv) was added, and the reaction was stirred at 100 °C for 1.5 h. The resulting suspension was filtered through a pad of celite, and the solvent was removed *in vacuo*. The crude was further purified by preparative HPLC to yield the desired product.

General Procedure B: Knoevenagel Condensation. To a solution of thiazolidinone derivative (2.0 equiv) and aldehyde (1.0 equiv) in AcOH was added β-alanine (3.0 equiv). The resulting mixture was stirred at 100 °C for 30 min. After conversion, the precipitated product was filtered and purified by preparative HPLC to yield the desired product.

*N-(4-(5-Formylpyridin-3-yl)phenyl)methanesulfonamide (**2a**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**1a**) and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield a yellow-brown oil (84%). MS (ESI) m/z calcd for $C_{13}H_{12}N_2O_3S$ $[M + H]^+$ 277.06; found 277.26.

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

*N-(4-(5-Bromopyridin-3-yl)phenyl)methanesulfonamide (**2b**).* The intermediate was synthesized after general procedure A, starting from 3,5-dibromopyridine (**1b**) and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield a light-ochre solid (55%). MS (ESI) *m*/*z* calcd for $C_{12}H_{11}BrN_2O_2S$ $[M + H]^+$ 326.98; found 327.11.

*N-(4-(5-Aminopyridin-3-yl)phenyl)methanesulfonamide (**2c**).* The intermediate was synthesized after general procedure A, starting from 5-bromopyridin-3-amine (**1c**) and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield a white solid (206 mg, 100% = 187 mg). MS (ESI) *m*/*z* calcd for $C_{12}H_{13}N_3O_2S$ $[M + H]^+$ 264.08; found 264.27.

(Z)-N-(4-(5-((2-Imino-4-oxothiazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)methanesulfonamide (CVM-05-002). The title compound was synthesized after general procedure B, starting from intermediate **2a** and 2-iminothiazolidin-4-one to yield a light-yellow solid (33%). MS (ESI) *m*/*z* calcd for $C_{16}H_{14}N_4O_3S_2$ $[M + H]^+$ 375.06; found 375.20. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.98 (s, 1H), 9.57 (s, 1H), 9.24 (s, 1H), 8.88 (s, 1H), 8.77 (s, 1H), 8.15 (s, 1H), 7.76 (d, *J* = 8.5 Hz, 2H), 7.70 (s, 1H), 7.35 (d, *J* = 8.5 Hz, 2H), 3.06 (s, 3H). ^{13}C NMR (126 MHz, DMSO) δ 175.67, 149.25, 147.98, 139.45, 135.59, 133.82, 132.58, 132.02, 130.84, 128.45, 125.93, 120.41, 40.92; purity 100%.

*(Z)-N-(4-(5-((2,4-Dioxothiazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)methanesulfonamide (**3**).* The title compound was synthesized after general procedure B, starting from intermediate **2a** and thiazolidine-2,4-dione to yield a light-yellow solid (12%). MS (ESI) *m*/*z* calcd for $C_{16}H_{13}N_3O_4S_2$ $[M + H]^+$ 376.04; found 376.18. ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.75 (s, 1H), 9.99 (s, 1H), 8.93 (d, *J* = 1.8 Hz, 1H), 8.77 (s, 1H), 8.16 (s, 1H), 7.90 (s, 1H), 7.77 (d, *J* = 8.5 Hz, 2H), 7.36 (d, *J* = 8.5 Hz, 2H), 3.06 (s, 3H). ^{13}C NMR (126 MHz, DMSO) δ 167.96, 167.62, 149.44, 148.71, 139.54, 135.65, 134.25, 131.65, 129.93, 128.89, 128.52, 126.87, 120.32; purity 100%.

*(Z)-N-(4-(5-((2,4-Dioxo-oxazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)methanesulfonamide (**4**).* Intermediate **2a** (0.145 mmol, 1.0 equiv), oxazolidine-2,4-dione (0.174 mmol, 1.2 equiv), and piperidine (0.044 mmol, 0.3 equiv) were dissolved in EtOH (1 mL) and stirred at 100 °C overnight. The crude was purified by preparative HPLC to yield the title compound as a brown solid (4%). MS (ESI) *m*/*z* calcd for $C_{16}H_{13}N_3O_5S$ $[M + H]^+$ 360.06; found 360.48. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.96 (s, 2H), 8.94 (s, 1H), 8.53 (s, 1H), 8.03 (s, 1H), 7.77 (d, *J* = 8.6 Hz, 2H), 7.33 (d, *J* = 8.6 Hz, 2H), 3.05 (s, 3H); purity 90%.

*(E)-N-(4-(5-((2,5-Dioxopyrrolidin-3-ylidene)methyl)pyridin-3-yl)phenyl)methanesulfonamide (**5**).* Intermediate **2a** (0.145 mmol, 1.0 equiv), maleimide (0.145 mmol, 1.0 equiv), and triphenylphosphine (0.145 mmol, 1.0 equiv) were dissolved in DCM (1 mL) and stirred at room temperature overnight. The crude was purified by preparative HPLC to yield the title compound as an ivory solid (4%). MS (ESI) *m*/*z* calcd for $C_{17}H_{15}N_3O_4S$ $[M + H]^+$ 358.09; found 358.44. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.54 (s, 1H), 9.96 (s, 1H), 8.89 (s, 1H), 8.79 (s, 1H), 8.21 (s, 1H), 7.81 (d, *J* = 8.5 Hz, 2H), 7.49 (s, 1H), 7.35 (d, *J* = 8.5 Hz, 2H), 3.84 (d, *J* = 2.4 Hz, 2H), 3.05 (s, 3H); purity 98%.

*N-(4-(5-(2-Oxoindolin-4-yl)pyridin-3-yl)phenyl)methanesulfonamide (**6**).* The title compound was synthesized after general procedure A, starting from intermediate **2b** and (2-oxoindolin-4-yl)boronic acid pinacol ester to yield an orange-brown solid (1%). MS (ESI) *m*/*z* calcd for $C_{20}H_{17}N_3O_3S$ $[M + H]^+$ 380.11; found 380.22. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.53 (s, 1H), 9.94 (s, 1H), 8.89 (s, 1H), 8.76 (s, 1H), 8.20 (s, 1H), 7.82 (d, *J* = 8.6 Hz, 2H), 7.40–7.30 (m, 3H), 7.17 (d, *J* = 7.7 Hz, 1H), 6.90 (d, *J* = 7.5 Hz, 1H), 3.73 (s, 2H), 3.05 (s, 3H); purity 94%.

*N-(3-(5-(4-(Methylsulfonamido)phenyl)pyridin-3-yl)phenyl)acetamide (**7**).* The title compound was synthesized after general procedure A, starting from intermediate **2b** and (3-acetamidophenyl)-boronic acid to yield a white solid (37%). MS (ESI) *m*/*z* calcd for $C_{20}H_{19}N_3O_3S$ $[M + H]^+$ 382.12; found 382.28. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.15 (s, 1H), 10.04 (s, 1H), 9.03 (s, 1H), 8.93 (s, 1H), 8.49 (s, 1H), 8.03 (s, 1H), 7.88 (d, *J* = 8.7 Hz, 2H), 7.67 (d, *J* = 8.1 Hz, 1H), 7.54 (d, *J* = 7.8 Hz, 1H), 7.47 (t, *J* = 7.9 Hz, 1H), 7.37 (d, *J* = 8.7 Hz, 2H), 3.07 (s, 3H), 2.08 (s, 3H); purity 100%.

*N-(4-(5-(4-(3-Methylureido)phenyl)pyridin-3-yl)phenyl)methanesulfonamide (**8**).* The title compound was synthesized after general procedure A, starting from intermediate **2b** and (4-(3-methylureido)phenyl)boronic acid to yield a light-yellow solid (37%). MS (ESI) *m*/*z* calcd for $C_{20}H_{20}N_4O_3S$ $[M + H]^+$ 397.13; found 397.21. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.97 (s, 1H), 8.88 (s, 1H), 8.86 (s, 1H), 8.73 (s, 1H), 8.38 (s, 1H), 7.86 (d, *J* = 8.5 Hz, 2H), 7.76 (d, *J* = 8.9 Hz, 2H), 7.56 (d, *J* = 8.9 Hz, 2H), 7.35 (d, *J* = 8.5 Hz, 2H), 6.10 (d, *J* = 4.0 Hz, 1H), 3.06 (s, 3H), 2.66 (d, *J* = 3.2 Hz, 3H); purity 100%.

*N-(3-((5-(4-(Methylsulfonamido)phenyl)pyridin-3-yl)oxy)phenyl)acetamide (**9**).* Intermediate **2b** (0.122 mmol, 1.0 equiv), *N*-(3-hydroxyphenyl)acetamide (0.24 mmol, 2.0 equiv), K_2CO_3 (0.24 mmol, 2.0 equiv), and CuI (0.0122 mmol, 0.1 equiv) were dissolved in DMF (0.5 mL) and stirred at 120 °C for 2 days. The crude was purified by preparative HPLC to yield the title compound as a brown solid (3%). MS (ESI) *m*/*z* calcd for $C_{20}H_{19}N_3O_4S$ $[M + H]^+$ 398.12; found 398.29. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.04 (s, 1H), 9.95 (s, 1H), 7.98–7.68 (m, 4H), 7.42 (s, 1H), 7.38–7.29 (m, 5H), 6.78 (s, 1H), 3.03 (s, 3H), 2.01 (s, 3H); purity 92%.

*N-(3-((5-(4-(Methylsulfonamido)phenyl)pyridin-3-yl)amino)phenyl)acetamide (**10**).* Intermediate **2c** (0.152 mmol, 1.0 equiv), *N*-(3-bromophenyl)acetamide (0.182 mmol, 1.2 equiv), K_2CO_3 (0.38 mmol, 2.5 equiv), and X-Phos (0.023 mmol, 0.15 equiv) were dissolved in 1,4-dioxane (1.0 mL). The mixture was purged with N_2 for about 10 min, $Pd_2(dba)_3$ (0.0076 mmol, 0.05 equiv) was added, and the reaction was stirred at 80 °C overnight. The crude was purified by preparative HPLC to yield the title compound as a yellow solid (7%). MS (ESI) *m*/*z* calcd for $C_{20}H_{20}N_4O_3S$ $[M + H]^+$ 397.13; found 397.54. ^{1}H NMR (500 MHz, DMSO-*d*6) δ 9.96 (s, 1H), 9.92 (s, 1H), 8.71 (s, 1H), 7.78 (s, 1H), 7.71 (d, *J* = 8.5 Hz, 2H), 7.66 (s, 1H), 7.32 (d, *J* = 8.5 Hz, 2H), 7.24–7.18 (m, 1H), 7.11 (s, 1H), 7.05 (d, *J* = 8.2 Hz, 1H), 7.01 (s, 1H), 6.83 (d, *J* = 8.0 Hz, 1H), 3.05 (s, 3H), 2.04 (s, 3H); purity 100%.

*4-(5-Bromopyridin-3-yl)-1H-indole (**12a**).* The intermediate was synthesized after general procedure A, starting from 4-bromo-1*H*-indole and (5-bromopyridin-3-yl)boronic acid (**11**) to yield a yellow-brown oil (18%). MS (ESI) *m*/*z* calcd for $C_{13}H_9BrN_2$ $[M + H]^+$ 273.00; found 273.32.

*5-(5-Bromopyridin-3-yl)-1H-indole (**12b**).* The intermediate was synthesized after general procedure A, starting from 5-bromo-1*H*-indole and (5-bromopyridin-3-yl)boronic acid (**11**) to yield a yellow-brown oil (9%). MS (ESI) *m*/*z* calcd for $C_{13}H_9BrN_2$ $[M + H]^+$ 273.00; found 273.33.

*6-(5-Bromopyridin-3-yl)-1H-indole (**12c**).* The intermediate was synthesized after general procedure A, starting from 6-bromo-1*H*-indole and (5-bromopyridin-3-yl)boronic acid (**11**) to yield a yellow-orange solid (14%). MS (ESI) *m*/*z* calcd for $C_{13}H_9BrN_2$ $[M + H]^+$ 273.00; found 273.34.

*N-(4-(5-(1H-Indol-4-yl)pyridin-3-yl)phenyl)methanesulfonamide (**13**).* The title compound was synthesized after general procedure A, starting from intermediate **12a** and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield an ochre solid (19%). MS (ESI) *m*/*z* calcd for $C_{20}H_{17}N_3O_2S$ $[M + H]^+$ 364.11; found 364.14. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.39 (s, 1H), 9.97 (s, 1H), 8.94 (s, 1H), 8.90 (s, 1H), 8.35 (s, 1H), 7.85 (d, *J* = 8.9 Hz, 2H), 7.61 (d, *J* = 8.5 Hz, 1H), 7.51 (d, *J* = 7.3 Hz, 1H), 7.49–7.46 (m, 1H), 7.36 (d, *J* = 8.9 Hz, 2H), 7.30–7.23 (m, 1H), 6.61 (s, 1H), 3.06 (s, 3H); purity 94%.

*N-(5-(5-(1H-Indol-4-yl)pyridin-3-yl)phenyl)methanesulfonamide (**14**).* The title compound was synthesized after general procedure A, starting from intermediate **12b** and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield an orange solid (27%). MS (ESI) *m*/*z* calcd for $C_{20}H_{17}N_3O_2S$ $[M + H]^+$ 364.11; found 364.14. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.30 (s, 1H), 9.98 (s, 1H), 8.94 (s, 1H), 8.89 (s, 1H), 8.42 (s, 1H), 7.88 (d, *J* = 8.9 Hz, 2H), 7.84 (s, 1H), 7.70 (d, *J* = 8.2 Hz, 1H), 7.48 (d, *J* = 10.1 Hz, 1H),

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880–4895

7.46−7.41 (m, 1H), 7.37 (d, J = 8.5 Hz, 2H), 6.50 (s, 1H), 3.06 (s, 3H); purity 92%.

*N-(6-(5-(1H-Indol-4-yl)pyridin-3-yl)phenyl)methanesulfonamide (**15**).* The title compound was synthesized after general procedure A, starting from intermediate **12c** and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)methanesulfonamide to yield an off-white solid (21%). MS (ESI) m/z calcd for $C_{20}H_{17}N_3O_2S$ $[M + H]^+$ 364.11; found 364.13. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.28 (s, 1H), 9.99 (s, 1H), 8.97 (s, 1H), 8.89 (s, 1H), 8.50 (s, 1H), 8.07 (s, 1H), 7.90 (d, J = 8.5 Hz, 2H), 7.59 (d, J = 8.9 Hz, 1H), 7.55 (d, J = 8.2 Hz, 1H), 7.46−7.42 (m, 1H), 7.37 (d, J = 8.5 Hz, 2H), 6.54 (s, 1H), 3.07 (s, 3H); purity 95%.

*4-(5-Formylpyridin-3-yl)-N-methylbenzenesulfonamide (**17a**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(*N*-methylsulfamoyl)-phenyl)boronic acid to yield a colorless oil (76%). MS (ESI) m/z calcd for $C_{13}H_{12}N_2O_3S$ $[M + H]^+$ 277.06; found 277.40.

*5-(4-(Methylsulfonyl)phenyl)nicotinaldehyde (**17b**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(methylsulfonyl)phenyl)boronic acid to yield a colorless oil (74%). MS (ESI) m/z calcd for $C_{13}H_{11}NO_3S$ $[M + H]^+$ 262.05; found 262.38.

*N-(4-(5-Formylpyridin-3-yl)phenyl)acetamide (**17c**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-acetamidophenyl)boronic acid to yield a yellow oil (88%). MS (ESI) m/z calcd for $C_{14}H_{12}N_2O_2$ $[M + H]^+$ 241.10; found 241.27.

*1-(4-(5-Formylpyridin-3-yl)phenyl)-3-methylurea (**17d**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(3-methylureido)phenyl)boronic acid pinacol ester to yield a yellow oil (63%). MS (ESI) m/z calcd for $C_{14}H_{13}N_3O_2$ $[M + H]^+$ 256.11; found 256.46.

*N-(3-(5-Formylpyridin-3-yl)phenyl)acetamide (**17e**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (3-acetamidophenyl)boronic acid to yield a light-yellow oil (71%). MS (ESI) m/z calcd for $C_{14}H_{12}N_2O_2$ $[M + H]^+$ 241.10; found 241.27.

*5-(4-(Trifluoromethoxy)phenyl)nicotinaldehyde (**17f**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(trifluoromethoxy)phenyl)-boronic acid to yield a yellow oil (93%). MS (ESI) m/z calcd for $C_{13}H_8F_3NO_2$ $[M + H]^+$ 268.06; found 268.38.

*5-(4-(Trifluoromethyl)phenyl)nicotinaldehyde (**17g**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(trifluoromethyl)phenyl)boronic acid to yield a light-yellow oil (64%). MS (ESI) m/z calcd for $C_{13}H_8F_3NO$ $[M + H]^+$ 252.06; found 252.25.

*5-(4-Fluorophenyl)nicotinaldehyde (**17h**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-fluorophenyl)boronic acid to yield a light-yellow oil (86 mg, 100% = 85 mg). MS (ESI) m/z calcd for $C_{12}H_8FNO$ $[M + H]^+$ 202.07; found 202.18.

*5-(4-(Dimethylamino)phenyl)nicotinaldehyde (**17i**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-(dimethylamino)phenyl)boronic acid to yield a red-orange oil (114 mg, 100% = 92 mg). MS (ESI) m/z calcd for $C_{14}H_{14}N_2O$ $[M + H]^+$ 227.12; found 227.42.

*5-(4-Methoxynaphthalen-1-yl)nicotinaldehyde (**17j**).* The intermediate was synthesized after general procedure A, starting from 5-bromonicotinaldehyde (**16**) and (4-methoxynaphthalen-1-yl)boronic acid to yield a yellow-orange oil (71%). MS (ESI) m/z calcd for $C_{17}H_{13}NO_2$ $[M + H]^+$ 264.10; found 264.43.

*(Z)-4-(5-((2-Imino-4-oxothiazolidin-5-ylidene)methyl)pyridin-3-yl)-N-methylbenzenesulfonamide (**18**).* The title compound was synthesized after general procedure B, starting from intermediate **17a** and 2-iminothiazolidin-4-one to yield a white solid (19%). MS (ESI) m/z calcd for $C_{16}H_{14}N_4O_2S_3$ $[M + H]^+$ 375.06; found 375.42. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.59 (s, 1H), 9.26 (s, 1H), 8.97 (s, 1H), 8.86 (s, 1H), 8.27 (t, J = 2.0 Hz, 1H), 8.02 (d, J = 8.5 Hz, 2H), 7.91 (d, J = 8.5 Hz, 2H), 7.73 (s, 1H), 7.56 (q, J = 4.9 Hz, 1H), 2.46 (d, J = 5.2 Hz, 3H); purity 98%.

*(Z)-2-Imino-5-((5-(4-(methylsulfonyl)phenyl)pyridin-3-yl)-methylene)thiazolidin-4-one (**19**).* The title compound was synthesized after general procedure B, starting from intermediate **17b** and 2-iminothiazolidin-4-one to yield a white solid (9%). MS (ESI) m/z calcd for $C_{16}H_{13}N_3O_3S_2$ $[M + H]^+$ 360.05; found 360.38. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.59 (s, 1H), 9.27 (s, 1H), 8.99 (s, 1H), 8.88 (s, 1H), 8.28 (s, 1H), 8.11−8.01 (m, 4H), 7.74 (s, 1H), 3.29 (s, 3H); purity 97%.

*(Z)-N-(4-(5-((2-Imino-4-oxothiazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)acetamide (**20**).* The title compound was synthesized after general procedure B, starting from intermediate **17c** and 2-iminothiazolidin-4-one to yield an off-white solid (4%). MS (ESI) m/z calcd for $C_{17}H_{14}N_4O_2S$ $[M + H]^+$ 339.09; found 339.18. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.11 (s, 1H), 9.57 (s, 1H), 9.23 (s, 1H), 8.88 (d, J = 2.2 Hz, 1H), 8.75 (d, J = 2.1 Hz, 1H), 8.15 (t, J = 2.2 Hz, 1H), 7.76−7.71 (m, 4H), 7.70 (s, 1H), 2.08 (s, 3H); purity 100%.

*(Z)-1-(4-(5-((2-Imino-4-oxothiazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)-3-methylurea (**21**).* The title compound was synthesized after general procedure B, starting from intermediate **17d** and 2-iminothiazolidin-4-one to yield an ivory solid (18%). MS (ESI) m/z calcd for $C_{17}H_{15}N_5O_2S$ $[M + H]^+$ 354.10; found 354.42. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.56 (s, 1H), 9.23 (s, 1H), 8.87 (s, 1H), 8.75−8.67 (m, 2H), 8.12 (d, J = 2.1 Hz, 1H), 7.70 (s, 1H), 7.65 (d, J = 8.8 Hz, 2H), 7.56 (d, J = 8.8 Hz, 2H), 6.09 (t, J = 4.7 Hz, 1H), 2.66 (d, J = 4.5 Hz, 3H); purity 100%.

*(Z)-N-(3-(5-((2-Imino-4-oxothiazolidin-5-ylidene)methyl)pyridin-3-yl)phenyl)acetamide (**22**).* The title compound was synthesized after general procedure B, starting from intermediate **17e** and 2-iminothiazolidin-4-one to yield an off-white solid (4%). MS (ESI) m/z calcd for $C_{17}H_{14}N_4O_2S$ $[M + H]^+$ 339.09; found 339.16. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.12 (s, 1H), 9.58 (s, 1H), 9.26 (s, 1H), 8.82 (s, 2H), 8.09 (s, 1H), 7.95 (s, 1H), 7.72 (s, 1H), 7.66 (d, J = 8.5 Hz, 1H), 7.46 (t, J = 7.9 Hz, 1H), 7.41 (d, J = 7.6 Hz, 1H), 2.08 (s, 3H); purity 100%.

*(Z)-2-Imino-5-((5-(4-(trifluoromethoxy)phenyl)pyridin-3-yl)-methylene)thiazolidin-4-one (**23**).* The title compound was synthesized after general procedure B, starting from intermediate **17f** and 2-iminothiazolidin-4-one to yield a light-orange solid (11%). MS (ESI) m/z calcd for $C_{16}H_{10}F_3N_3O_2S$ $[M + H]^+$ 366.05; found 366.40. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.58 (s, 1H), 9.25 (s, 1H), 8.93 (s, 1H), 8.83 (s, 1H), 8.20 (t, J = 2.0 Hz, 1H), 7.95−7.88 (m, 2H), 7.72 (s, 1H), 7.54 (d, J = 7.9 Hz, 2H); purity 96%.

*(Z)-2-Imino-5-((5-(4-(trifluoromethyl)phenyl)pyridin-3-yl)-methylene)thiazolidin-4-one (**24**).* The title compound was synthesized after general procedure B, starting from intermediate **17g** and 2-iminothiazolidin-4-one to yield a white solid (12%). MS (ESI) m/z calcd for $C_{16}H_{10}F_3N_3OS$ $[M + H]^+$ 350.06; found 350.15. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.59 (s, 1H), 9.27 (s, 1H), 8.97 (s, 1H), 8.86 (s, 1H), 8.26 (t, J = 2.0 Hz, 1H), 8.02 (d, J = 8.2 Hz, 2H), 7.90 (d, J = 8.2 Hz, 2H), 7.73 (s, 1H); purity 100%.

*(Z)-5-((5-(4-Fluorophenyl)pyridin-3-yl)methylene)-2-iminothiazolidin-4-one (**25**).* The title compound was synthesized after general procedure B, starting from intermediate **17h** and 2-iminothiazolidin-4-one to yield a white solid (4%). MS (ESI) m/z calcd for $C_{15}H_{10}FN_3OS$ $[M + H]^+$ 300.06; found 300.15. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.58 (s, 1H), 9.25 (s, 1H), 8.89 (d, J = 2.1 Hz, 1H), 8.80 (s, 1H), 8.16 (t, J = 2.0 Hz, 1H), 7.88−7.77 (m, 2H), 7.71 (s, 1H), 7.39 (t, J = 8.9 Hz, 2H); purity 97%.

*(Z)-5-((5-(4-(Dimethylamino)phenyl)pyridin-3-yl)methylene)-2-iminothiazolidin-4-one (**26**).* The title compound was synthesized after general procedure B, starting from intermediate **17i** and 2-iminothiazolidin-4-one to yield an orange-red solid (13%). MS (ESI) m/z calcd for $C_{17}H_{16}N_4OS$ $[M + H]^+$ 325.11; found 325.43. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.57 (s, 1H), 9.25 (s, 1H), 8.87 (s, 1H), 8.69 (s, 1H), 8.14 (t, J = 2.1 Hz, 1H), 7.70 (s, 1H), 7.64 (d, J = 8.8 Hz, 2H), 6.86 (d, J = 8.8 Hz, 2H), 2.98 (s, 6H); purity 96%.

*(Z)-2-Imino-5-((5-(4-methoxynaphthalen-1-yl)pyridin-3-yl)-methylene)thiazolidin-4-one (**27**).* The title compound was synthesized after general procedure B, starting from intermediate **17j** and 2-

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

iminothiazolidin-4-one to yield an orange solid (21%). MS (ESI) m/z calcd for $C_{20}H_{15}N_3O_2S$ $[M + H]^+$ 362.10; found 362.44. ^1H NMR (500 MHz, DMSO-d_6) δ 9.54 (s, 1H), 9.18 (s, 1H), 8.89 (s, 1H), 8.68 (s, 1H), 8.33−8.24 (m, 1H), 7.99 (s, 1H), 7.79−7.75 (m, 1H), 7.75 (s, 1H), 7.61−7.56 (m, 2H), 7.50 (d, J = 7.9 Hz, 1H), 7.12 (d, J = 7.9 Hz, 1H), 4.05 (s, 3H); purity 100%.

*N-(4-(6-Amino-5-formylpyridin-3-yl)phenyl)-methanesulfonamide (**29a**).* The intermediate was synthesized after general procedure A, starting from 2-amino-5-bromonicotinaldehyde (**28a**) and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-methanesulfonamide (**1**) to yield a yellow oil (74%). MS (ESI) m/z calcd for $C_{13}H_{13}N_3O_3S$ $[M + H]^+$ 292.08; found 292.16.

*N-(4-(2-Amino-5-formylpyridin-3-yl)phenyl)-methanesulfonamide (**29b**).* The intermediate was synthesized after general procedure A, starting from 6-amino-5-bromonicotinaldehyde (**28b**) and *N*-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-methanesulfonamide (**1**) to yield a light-yellow solid (61%). MS (ESI) m/z calcd for $C_{13}H_{13}N_3O_3S$ $[M + H]^+$ 292.08; found 292.23.

*(Z)-N-(4-(6-Amino-5-((2-imino-4-oxothiazolidin-5-ylidene)-methyl)pyridin-3-yl)phenyl)methanesulfonamide (**30**).* The title compound was synthesized after general procedure B, starting from intermediate **29a** and 2-iminothiazolidin-4-one to yield a yellow solid (22%). MS (ESI) m/z calcd for $C_{16}H_{15}N_5O_3S_2$ $[M + H]^+$ 390.07; found 390.24. ^1H NMR (500 MHz, DMSO-d_6) δ 9.85 (s, 1H), 9.36 (s, 1H), 9.11 (s, 1H), 8.59 (d, J = 2.3 Hz, 1H), 8.24 (dd, J = 7.5, 2.3 Hz, 1H), 7.67−7.53 (m, 4H), 7.34−7.27 (m, 3H), 3.03 (s, 3H); purity 86%.

*(Z)-N-(4-(2-Amino-5-((2-imino-4-oxothiazolidin-5-ylidene)-methyl)pyridin-3-yl)phenyl)methanesulfonamide (**31**).* The title compound was synthesized after general procedure B, starting from intermediate **29b** and 2-iminothiazolidin-4-one to yield a yellow solid (10%). MS (ESI) m/z calcd for $C_{16}H_{15}N_5O_3S_2$ $[M + H]^+$ 390.07; found 390.23. ^1H NMR (500 MHz, DMSO-d_6) δ 9.96 (s, 1H), 9.39 (s, 1H), 9.03 (s, 1H), 8.26 (d, J = 2.1 Hz, 1H), 7.60 (d, J = 2.1 Hz, 1H), 7.53 (s, 1H), 7.49−7.42 (m, 2H), 7.35−7.32 (m, 2H), 7.07 (s, 2H), 3.06 (s, 3H); purity 100%.

Biochemical Experimental. *High-Throughput Screen.* Screenings were performed according to previously described protocols.[12] 1,2-dipalmitoyl-*sn*-glycero-3-phosphoserine, sodium salt (DPPS) and PI5P (Echelon Biosciences) (2:1) were dissolved in DMSO (333 mL/mg). Sixty-three microliters of DMSO was added to 1255 μL of buffer 1 (30 mM Hepes pH 7.4, 1 mM ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA), 0.1% 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS)) and 2868 μL of buffer 2 (46 mM Hepes pH 7.4, 0.1% CHAPS). This could be adjusted to sample numbers. PI5P4Kα enzyme was added to the buffer mixture at 32 nM concentration. Ten microliters was dispensed into white 384-well plates (Corning #3824), and 100 nL of compound solution (in DMSO) was transferred by a pintool (JANUS, PerkinElmer). The reaction was initiated with the addition of 5 μL of 15 μM ATP (Promega) and PI5P/DPPS (at 0.06 μg/μL and 0.12 μg/μL, respectively) in buffer 3 (20 mM Hepes pH 7.4, 60 mM $MgCl_2$, and 0.1% CHAPS). The final concentration of DMSO in the reaction was less than 5%. The resulting mixture was incubated at room temperature in the dark for 1 h. Five microliters of ADP-Glo reagent 1 was added to stop the reaction and remove the remaining ATP. After 45 min, 10 μL of ADP-Glo reagent 2 was added and incubated for 30 min. The luminescence intensity (RLU) was then read on an EnVision 2104 Multilabel Plate Reader (PerkinElmer). IC_{50} values were determined using GraphPad Prism 7.04 and nonlinear regression curve fit.

Protein Expression and Purification. Recombinant PI5P4Kα, β, and γ were obtained as described before.[16]

Mass Spectrometry. Recombinant PI5P4Kα/β was incubated with 10 μM CVM-05-002, THZ-P1-2 (positive control), or DMSO and analyzed by LC/MS as previously described,[16] except that all samples were incubated at room temperature for 2 h, and UniDec[38] was used to deconvolute the protein mass spectra.

X-ray Crystallography and Structure Determination. PI5P4Kα was cocrystallized with CVM-05-002 using similar conditions as previously described.[16] A 2-fold excess of inhibitor was mixed with 500 μM protein and crystallized by sitting-drop vapor diffusion as described. The buffer contained 2 M NH_4 (pH 6.5). Crystals were transferred briefly into crystallization buffer, containing 25% glycerol prior to flash-freezing in liquid nitrogen. Diffraction data from complex crystals were collected at beamline 24ID of the NE-CAT at the Advanced Photon Source (Argonne National Laboratory). Data sets were integrated and scaled using XDS (Kabsch, 2010). Structures were solved by molecular replacement using the program Phaser. The ligand was positioned manually and refined using Buster and Rhofit. Iterative manual model building and refinement using Phenix and Coot led to a model with excellent statistics, including a maximum diffraction of 1.71 Å (Table 4).

Table 4. Data Collection and Refinement Statistics

PI5P4Kα/CVM-05-002	
wavelength (Å)	0.9792
resolution range (Å)	45.37−1.71 (1.771−1.71)
space group	P 1 21 1
unit cell constants a, b, c, α, β, γ	44.23 Å, 88.58 Å, 105.78 Å, 90°, 92.91°, 90°
total reflections	291561 (29776)
unique reflections	85988 (8587)
multiplicity	3.4 (3.5)
completeness (%)	97.72 (97.59)
mean $I/\sigma(I)$	12.42 (1.34)
Wilson B-factor	32.73
R_{merge}	0.04003 (1.022)
R_{meas}	0.04773 (1.207)
R_{pim}	0.02566 (0.6364)
CC1/2	0.999 (0.754)
CC*	1 (0.927)
reflections used in refinement	85916 (8579)
reflections used for R_{free}	4198 (366)
R_{work}	0.1767 (0.3611)
R_{free}	0.2032 (0.3938)
CC (work)	0.966 (0.878)
CC (free)	0.947 (0.833)
no. of nonhydrogen atoms	5611
macromolecules	5203
ligands	50
solvent	358
protein residues	628
RMS (bonds)	0.016
RMS (angles)	1.73
Ramachandran favored (%)	97.39
Ramachandran allowed (%)	2.45
Ramachandran outliers (%)	0.16
rotamer outliers (%)	0.17
clashscore	10.58
average B-Factor	51.85
macromolecules	52.05
ligands	46.78
solvent	49.78
no. of TLS groups	15

Protein Structure Alignment. The above-described cocrystal structure of PI5P4Kα with CVM-05-002 was aligned with the PDB structures of PI5P4Kβ (PDB-ID: 3WZZ) and PI5P4Kγ (PDB-ID: 2GK9.A) using the alignment function in PyMOL 2.0.5, Incentive Product Copyright Schrodinger, LLC.

Biological Evaluation. Materials and Methods. All compounds were initially sourced from the National Center for Advancing Translational Studies (NCATS)/National Institutes of Health (NIH). Shipped compounds from NCATS were subjected to quality control

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

by LC/UV, LC/MS, or high-resolution mass spectrometry (HRMS), unless otherwise noted, and exhibited >90% purity by peak area or m/z^2.

Data analysis was performed as previously described.[17]

ADP-Glo Kinase Assay (PI5P4Kα). GDP Transcreener Assay (PI5P4Kβ). Protocols for screens and counterscreens were previously reported.[17]

Kinome Profiling. Protocols for the KINOMEscan are available from DiscoverX. Tree*spot* images were generated, using TREE*spot* Software Tool, and reprinted with permission from KINOMEscan, a division of DiscoverX Corporation, DISCOVERX CORPORATION 2010. All compounds were tested at a concentration of 1 μM.

IC_{50} Evaluation of Off-Targets. Biochemical IC_{50}s of identified off-targets were determined using the commercially available LanthaScreen or Adapta offered by Invitrogen (Thermo Fisher Scientific).

K_D Evaluation for PI5P4Kγ. The K_D values for PI5P4Kγ binding were determined by the commercially available KdELECT, and a respective protocol is available from DiscoverX.

Cell Culture. HEK 293T cells were cultured as recommended by American Type Culture Collection (ATCC). The cells were kept frozen, so only freshly thawed cells at early passage were used. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) media, which was supplemented with 10% fetal bovine serum (FBS) and 100 μL/mL penicillin−streptomycin and incubated at 37 °C and 5% CO_2. No authentication was done by the authors.

Compounds and Reagents. Antibodies purchased from Cell Signaling Technology (Hanover, MA) were anti-PI5P4Kα (5527) and anti-PI5P4Kβ (9694). Anti-GAPDH (MAB374) was obtained from EMD Millipore Corporation (Merck KGaA, Darmstadt, Germany).

Cellular Thermal Shift Assay (CETSA). Obtaining the melting curves, as well as performing the isothermal CETSA experiments and Western blot analysis were reported previously.[17]

Kinetic Solubility Assay. Solubility of selected compounds was determined according to methods as previously published.[39]

Rat Liver Microsome Stability Assay. Single time point microsomal stability was determined in a 96-well HTS format. Sample preparation was automated using a Tecan EVO 200 robot. High-resolution LC−MS (Thermo ZExactive) instrument was used to measure the percentage of compound remaining after incubation using a previously described method.[40] Six standard controls were tested in each run: buspirone and propranolol (for short half-life), loperamide and diclofenac (for short to medium half-life), and carbamazepine and antipyrine (for long half-life). Briefly, the incubation consisted of a 0.5 mg/mL microsomal protein, a 1.0 μM drug concentration, and NADPH regeneration system (containing 0.650 mM $NADP^+$, 1.65 mM glucose 6-phosphate, 1.65 mM $MgCl_2$, and 0.2 unit/mL G6PDH) in 100 mM phosphate buffer at pH 7.4. The incubation was carried out at 37 °C for 15 min. The reaction was quenched by adding 555 μL of acetonitrile (~1:2 ratio) containing 0.28 μM albendazole (internal standard). Sample acquisition and data analysis were done using a previously described method.[40]

Parallel Artificial Membrane Permeability Assay (PAMPA). Compound permeability was determined as previously described.[41]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.0c00227.

Overview of CVM-05-002 analogs, full KINOMEscan data sets, lipophilicity and early ADME overview of CVM-05-002 and analogs, structures of reported covalent pan-PI5P4K inhibitors, plate-well series plot with mean %DMSO control, and complete results of high-throughput screen (PDF)

SMILES molecular formula strings (CSV)

Accession Codes

PDB code for PI5P4Kα with bound CVM-05-002 is 6UX9. The authors will release the atomic coordinates and experimental data upon article publication.

■ AUTHOR INFORMATION

Corresponding Author

Nathanael S. Gray − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States;* orcid.org/0000-0001-5354-7403; Email: Nathanael_gray@dfci.harvard.edu

Authors

Theresa D. Manz − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States; Department of Pharmaceutical and Medicinal Chemistry, Saarland University, 66123 Saarbruecken, Germany*

Sindhu Carmen Sivakumaren − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States*

Fleur M. Ferguson − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States;* orcid.org/0000-0003-4091-7617

Tinghu Zhang − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States*

Adam Yasgar − *National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States;* orcid.org/0000-0001-7350-1402

Hyuk-Soo Seo − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States;* orcid.org/0000-0003-0646-2102

Scott B. Ficarro − *Department of Cancer Biology and Blais Proteomics Center and Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, United States*

Joseph D. Card − *Department of Cancer Biology and Blais Proteomics Center and Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, United States*

Hyeseok Shim − *Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, New York 10065, United States*

Chandrasekhar V. Miduturu − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and*

Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States

Anton Simeonov − *National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States*

Min Shen − *National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States*

Jarrod A. Marto − *Department of Cancer Biology and Blais Proteomics Center and Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, United States*

Sirano Dhe-Paganon − *Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, United States*

Matthew D. Hall − *National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States;* orcid.org/0000-0002-5073-442X

Lewis C. Cantley − *Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, New York 10065, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.0c00227

Author Contributions

◆T.D.M. and S.C.S. contributed equally to this work.

Author Contributions

T.D.M., S.C.S., T.Z., F.M.F., and N.S.G. conceived the project and designed the research strategy. S.C.S. performed the library screening and helped with planning experimental strategies. T.D.M., F.M.F., and T.Z. designed the compounds. C.V.M. synthesized the initial hit compound. T.D.M. synthesized the compounds presented. H.-S.S. and S.D.-P. generated the cocrystal structure. A.Y. and A.S. performed the biological evaluation of the compounds under the supervision of M.D.H. and M.S. J.D.C. and S.B.F. performed mass spectrometry analysis under the supervision of J.A.M. T.D.M. planned and performed the cellular assays. H.S. and L.C.C. were involved in the planning and provided expertise and feedback for the project. The manuscript was written by T.D.M. and the contributions of all authors. All authors have given approval to the final version of the manuscript. N.S.G. supervised the research.

Funding

The authors would like to thank the National Institutes of Health for their generous financial support through grants R01 CA197329 (to N.S.G., subcontracted to S.D.P.), R01 NS089815, and R03 CA250020 (to J.A.M.), as well as R35 CA197588, R01 GM041890, and U54 CA210184 (to L.C.C.). This work was further supported by the Breast Cancer Research Foundation (to L.C.C.) and the NIH intramural research program (NCATS).

Notes

The authors declare the following competing financial interest(s): N.S.G. is a founder, SAB member, and equity holder in Gatekeeper, Syros, Petra, C4, B2S, and Soltego. The Gray lab receives or has received research funding from Novartis, Takeda, Astellas, Taiho, Janssen, Kinogen, Voronoi, Her2llc, Deerfield, and Sanofi. L.C.C. is a founder and member of the Board of Directors (BOD) of Agios Pharmaceuticals and is a founder and receives research support from Petra Pharmaceuticals. These companies are developing novel therapies for cancer. J.A.M. serves on the SAB of 908 Devices. S.D.P. receives research support from Taiho Pharmaceuticals.

■ ACKNOWLEDGMENTS

The authors would like to thank M. Kostic for her help and expertise in writing the manuscript.

■ ABBREVIATIONS

ACVR2A, activin A receptor type 2A; ACVR2B, activin A receptor type 2B; ATCC, American Type Culture Collection; calcd, calculated; CETSA, cellular thermal shift assay; CHAPS, 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate; CV, coefficient of variation; DPPS, 1,2-dipalmitoyl-*sn*-glycero-3-phosphoserine, sodium salt; EGTA, ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MEK3, mitogen-activated protein kinase kinase 3; NCATS, National Center for Advancing Translational Studies; NIH, National Institutes of Health; PAINS, pan-assay interference compounds; PAMPA, parallel artificial membrane permeability assay; $Pd_2(dba)_3$, tris(dibenzylideneacetone)dipalladium(0); PI-4,5-P_2, phosphatidylinositol 4,5-bisphosphate; PI5P, phosphatidylinositol 5-phosphate; PI5P4K, phosphatidylinositol 5-phosphate 4-kinases; PI4P5K, phosphatidylinositol 4-phosphate 5-kinase; S_{35}, selectivity score for hits with percent control <35; RLU, relative light units; SD, standard deviation; TGFβ, transforming growth factor β; TGFBR2, transforming growth factor β receptor 2; UPLC, ultrahigh-pressure liquid chromatography; X-Phos, 2-dicyclohexylphosphino-2′,4′,6′-triisopropylbiphenyl

■ REFERENCES

(1) Rameh, L. E.; Tolias, K. F.; Duckworth, B. C.; Cantley, L. C. A new pathway for synthesis for phosphatidylinositol-4,5-bisphosphate. *Nature* **1997**, *390*, 192−196.

(2) Hu, A.; Zhao, X.-T.; Fu, T.; Wang, Y.; Liu, Y.; Shi, X.-J.; Luo, J.; Song, B.-L.; et al. PIP4K2A regulates intracellular cholesterol transport through modulating $PI(4,5)_2$ homeostasis. *J. Lipid Res.* **2018**, *59*, 507−514.

(3) Lamia, K. A.; Peroni, O. D.; Kim, Y.-B.; Rameh, L. E.; Kahn, B. B.; Cantley, L. C. Increased insulin sensitivity and reduced adiposity in phosphatidylinositol 5-phoshate 4-kinase $\beta^{-/-}$ mice. *Mol. Cell. Biol.* **2004**, *24*, 5080−5087.

(4) Shim, H.; Wu, C.; Ramsamooj, S.; Bosch, K. N.; Chen, Z.; Emerling, B. M.; Yun, J.; Liu, H.; Choo-Wing, R.; Yang, Z.; Wulf, G. M.; Kuchroo, V. K.; Cantley, L. C. Deletion of the gene *Pip4k2c*, a novel phosphatidylinositol kinase, results in hyperactivation of the immune system. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 7596−7601.

(5) Lundquist, M. R.; Goncalves, M. D.; Loughran, R. M.; Possik, E.; Vijayaraghavan, T.; Yang, A.; Pauli, C.; Ravi, A.; Verma, A.; Yang, Z.; Johnson, J. L.; Wong, J. C. Y.; Ma, Y.; Hwang, K. S.-K.; Weinkove, D.; Divecha, N.; Asara, J. M.; Elemento, O.; Rubin, M. A.; Kimmelman, A. C.; ause, A.; Cantley, L. C.; Emerling, B. M. Phosphatidylinositol-5-phosphate 4-kinases regulate cellular lipid metabolism by facilitating autophagy. *Mol. Cell* **2018**, *70*, 531−543.

(6) Bulley, S. J.; Droubi, A.; Clarke, J. H.; Anderson, K. E.; Stephens, L. R.; Hawkins, P. T.; Irvine, R. F. In B cells, phosphatidylinositol 5-phosphate 4-kinase-α synthesizes $PI(4,5)P_2$ to impact mTORC2 and Akt signaling. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 10571−10576.

(7) Keune, W.-J.; Jones, D. R.; Divecha, N. PtdIns5*P* and Pin1 in oxidative stress signaling. *Adv. Biol. Regul.* **2013**, *53*, 179−189.
(8) Al-Ramahi, I.; Giridharan, S. S. P.; Chen, Y.-C.; Patnaik, S.; Safren, N.; Hasegawa, J.; de Haro, M.; Wagner Gee, A. K.; Titus, S. A.; Jeong, H.; Clarke, J.; Krainc, D.; Zheng, W.; Irvine, R. F.; Barmada, S.; Ferrer, M.; Southall, N.; Weisman, L. S.; Botas, J.; Marugan, J. J. Inhibition of PIP4Kγ amerliorates the pathological effects of mutant huntingtin protein. *eLife* **2017**, No. e29123.
(9) Jude, J. G.; Spencer, G. J.; Huang, X.; Somerville, T. D. D.; Jones, D. R.; Divecha, N.; Somervaille, T. C. P. A targeted knockdown screen of genes coding for phosphoinositide modulators identifies *PIP4K2A* as required for acute myeloid leukemia cell proliferation and survival. *Oncogene* **2015**, *34*, 1253−1262.
(10) Luoh, S.-W.; Venkatesan, N.; Tripathi, R. Overexpression of the amplified *Pip4k2*β gene from 17q11-12 in breast cancer cells confers proliferation advantage. *Oncogene* **2004**, *23*, 1354−1363.
(11) Emerling, B. M.; Hurov, J. B.; Poulogiannis, G.; Tsukazawa, K. S.; Choo-Wing, R.; Wulf, G. M.; Bell, E. L.; Shim, H.-S.; Lamia, K. A.; Rameh, L. E.; Bellinger, G.; Sasaki, A. T.; Asara, J. M.; Yuan, X.; Bullock, A.; DeNicola, G. M.; Song, J.; Brown, V.; Signoretti, S.; Cantley, L. C. Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-null tumors. *Cell* **2013**, *155*, 844−857.
(12) Davis, M. I.; Sasaki, A. T.; Shen, M.; Emerling, B. M.; Thorne, N.; Michael, S.; Pragani, R.; Boxer, M.; Sumita, K.; Takeuchi, K.; Auld, D. S.; Li, Z.; Cantley, L. C.; Simeonov, A. A Homogeneous, high-throughput assay for phosphatidylinositol 5-phosphate 4-kinase with a novel, rapid substrate preparation. *PLoS One* **2013**, *8*, No. e54127.
(13) Kitagawa, M.; Liao, P.-J.; Lee, K. H.; Wong, J.; Shang, S. C.; Minami, N.; Sampetrean, O.; Saya, H.; Lingyun, D.; Prabhu, N.; Diam, G. K.; Sobata, R.; Larsson, A.; Nordlung, P.; McCormick, F.; Ghosh, S.; Estein, D. M.; Dymock, B. W.; Lee, S. H. Dual blockade of the lipid kinase PIP4Ks and mitotic pathways lead to cancer-selective lethality. *Nat. Commun.* **2017**, *8*, No. 2200.
(14) Voss, M. D.; Czechtizky, W.; Li, Z.; Rudolph, C.; Petry, S.; Brummerhop, H.; Langer, T.; Schiffer, A.; Schaefer, H.-L. Discovery and pharmacological characterization of a novel small molecule inhibitor of phosphatidylinositol-5-phosphate 4-kinase, type II, beta. *Biochem. Biophys. Res. Commun.* **2014**, *449*, 327−331.
(15) Clarke, J. H.; Giudici, M.-L.; Burke, J. E.; Williams, R. L.; Maloney, D. J.; Marugan, J.; Irvine, R. F. The function of phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ) explored using a specific inhibitor that targets the PI5P-binding site. *Biochem. J.* **2015**, *466*, 359−367.
(16) Sivakumaren, S. C.; Shim, H.; Zhang, T.; Ferguson, F. M.; Lundquist, M. R.; Browne, C. M.; Seo, H.-S.; Paddock, M. N.; Manz, T. D.; Jiang, B.; Hao, M.-F.; Krishnan, P.; Wang, D. G.; Yang, J.; Kwiatkowski, N. P.; Ficarro, S. B.; Cunningham, J. M.; Marto, J. A.; Dhe-Paganon, S.; Cantley, L. C.; Gray, N. S. Targeting the PI5P4K lipid kinase family in cancer using novel covalent inhibitors. *Cell Chem. Biol.* **2020**, DOI: 10.1016/j.chembiol.2020.02.003.
(17) Manz, T. D.; Sivakumaren, S. C.; Yasgar, A.; Hall, M. D.; Davis, M. I.; Seo, H.-S.; Card, J. D.; Shim, H.; Ficarro, S. B.; Dhe-Paganon, S.; Sasaki, A. T.; Boxer, M. B.; Simeonov, A.; Cantley, L. C.; Shen, M.; Zhang, T.; Ferguson, F. M.; Gray, N. S.; et al. Structure-activity relationship study of covalent pan-phosphatidylinositol 5-phosphate 4-kinase inhibitors. *ACS Med. Chem. Lett.* **2020**, *11*, 346−352.
(18) Thress, K. S.; Paweletz, C. P.; Felip, E.; Cho, B. C.; Stetson, D.; Dougherty, B.; Lai, Z.; Markovets, A.; Vivancos, A.; Kuang, Y.; Ercan, D.; Matthews, S.; Cantarini, M.; Barrett, J. C.; Jänne, P. A.; Oxnard, G. R. Acquired *EGFR* C797S mediates resistance to AZD9291 in advanced non-small cell lung cancer harboring *EGFR* T790M. *Nat. Med.* **2015**, *21*, 560−562.
(19) Kenny, P. W. Comment on the ecstasy and agony of assay interference compounds. *J. Chem. Inf. Model* **2017**, *57*, 2640−2645.
(20) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719−2740.
(21) Baell, J. B.; Nissink, J. W. M. Seven year itch: Pan-assay interference compounds (PAINS) in 2017 − utility and limitations. *ACS Chem. Biol.* **2018**, *13*, 36−44.
(22) Mendgen, T.; Steuer, C.; Klein, C. D. Privileged scaffolds or promiscuous binders: A comparative study on rhodanines and related heterocycles in medicinal chemistry. *J. Med. Chem.* **2012**, *55*, 743−753.
(23) Camps, M.; Rueckle, T.; Ji, H.; Ardissone, V.; Rintelen, F.; Shaw, J.; Ferrandi, C.; Chabert, C.; Gillieron, C.; Francon, B.; Martin, T.; Gretener, D.; Perrin, D.; Leroy, D.; Vitte, P.-A.; Hirsch, E.; Wymann, M. P.; Cirillo, R.; Schwarz, M. K.; Rommel, C. Blockade of PI3Kγ suppresses joint inflammation and damage in mouse models of rheumatoid arthritis. *Nat. Med.* **2005**, *11*, 936−943.
(24) Ye, X.; Zhou, W.; Li, Y.; Sun, Y.; Zhang, Y.; Ji, H.; Lai, Y. Darbufelone, a novel anti-inflammatory drug, induces growth inhibition of lung cancer cells both in vitro and in vivo. *Cancer Chemother. Pharmacol.* **2010**, *66*, 277−285.
(25) Ghatak, S.; Vyas, A.; Misra, S.; O'Brien, P.; Zambre, A.; Fresco, V. M.; Markwald, R. R.; Swamy, K. V.; Afrasiabi, Z.; Choudhury, A.; Khetmalas, M.; Padhye, S. Novel di-tertiary-butyl phenylhydrazones as dual cyclooxygenase-2/5-lipoxygenase inhibitors: Synthesis, COX/LOX inhibition, molecular modeling, and insights into their cytotoxicities. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 317−324.
(26) Dakin, L. A.; Block, M. H.; Chen, H.; Code, E.; Dowling, J. E.; Feng, X.; Ferguson, A. D.; Green, I.; Hird, A. W.; Howard, T.; Keeton, E. K.; Lamb, M. L.; Lyne, P. D.; Pollard, H.; Read, J.; Wu, A. J.; Zhang, T.; Zheng, X. Discovery of novel benzylidene-1,3-thiazolidine-2,4-diones as potent and selective inhibitors of the PIM-1, PIM-2, and PIM-3 protein kinases. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 4599−4604.
(27) Cortes, J.; Tamura, K.; DeAngelo, D. J.; de Bono, J.; Lorente, D.; Minden, M.; Uy, G. L.; Kantarjian, H.; Chen, L. S.; Gandhi, V.; Godin, R.; Keating, K.; McEachern, K.; Vishwanathan, K.; Pease, J. E.; Dean, E. Phase I studies of AZD1208, a proviral integration Moloney virus kinase inhibitor in solid and haematological cancers. *Br. J. Cancer* **2018**, *118*, 1425−1433.
(28) Mathre, S.; Reddy, K. B.; Ramya, V.; Krishnan, H.; Ghosh, A.; Raghu, P.Functional analysis of the biochemical activity of mammalian phosphatidylinositol 5 phosphate 4-kinase enzymes. *Biosci Rep.***2019**, *39* () DOI: 10.1042/BSR20182210.
(29) Clarke, J. H.; Emson, P. C.; Irvine, R. F. Localization of phosphatidylinositol phosphate kinase II gamma in kidney to a membrane trafficking compartment within specialized cells of the nephron. *Am. J. Physiol.: Renal Physiol.* **2008**, *295*, F1422−F1430.
(30) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28−39.
(31) Roskoski, R., Jr. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol. Res.* **2016**, *103*, 26−48.
(32) Zhao, Z.; Wu, H.; Wang, L.; Liu, Y.; Knapp, S.; Liu, Q.; Gray, N. S. Exploration of type II binding mode: A privileged approach for kinase inhibitor focused drug discovery? *ACS Chem. Biol.* **2014**, *9*, 1230−1241.
(33) Massagué, J.; Gomis, R. R. The logic of TGFβ signaling. *FEBS Lett.* **2006**, *580*, 2811−2820.
(34) Chung, H.; Young, D. J.; Lopez, C. G.; Le, T.-A. T.; Lee, J. K.; Ream-Robinson, D.; Huang, S. C.; Carethers, J. M. Mutation rates of TGFBR2 and ACVR2 coding microsatellites in human cells with defective DNA mismatch repair. *PLoS One* **2008**, *3*, No. e3463.
(35) Jung, B. H.; Beck, S. E.; Cabral, J.; Chau, E.; Cabrera, B. L.; Fiorino, A.; Smith, E. J.; Bocanegra, M.; Carethers, J. M. *Activin type 2 receptor* restoration in MSI-H colon cancer suppresses growth and enhances migration with activin. *Gastroenterology* **2007**, *132*, 633−644.
(36) Henderson, M. J.; Holbert, M. A.; Simeonov, A.; Kallal, L. A. High-throughput cellular thermal shift assays in research and drug discovery. *SLAS Discovery* **2020**, *25*, 137−147.

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

(37) Bulley, S. J.; Clarke, J. H.; Droubi, A.; Giudici, M.-L.; Irvine, R. F. Exploring phosphatidylinositol 5-phosphate 4-kinase function. *Adv. Biol. Regul.* **2015**, *57*, 193−202.
(38) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **2015**, *87*, 4370−4376.
(39) Sun, H.; Shah, P.; Nguyen, K.; Yu, K. R.; Kerns, E.; Kabir, M.; Wang, Y.; Xu, X. Predictive models of aqueous solubility of organic compounds built on a large dataset of high integrity. *Bioorg. Med. Chem.* **2019**, *27*, 3110−3114.
(40) Shah, P.; Kerns, E.; Nguyen, D.-T.; Obach, S.; Wang, A. Q.; Zakharov, A.; McKew, J.; Simeonov, A.; Hop, C. E. C. A.; Xu, X. An automated high-throughput metabolic stability assay using an integrated high-resolution accurate mass method and automated data analysis software. *Drug Metab. Dispos.* **2016**, *44*, 1653−1661.
(41) Sun, H.; Nguyen, K.; Kerns, E.; Yan, Z.; Yu, K. R.; Shah, P.; Jadhav, A.; Xu, X. Highly predictive and interpretable models for PAMPA permeability. *Bioorg. Med. Chem.* **2017**, *25*, 1266−1276.

https://dx.doi.org/10.1021/acs.jmedchem.0c00227
J. Med. Chem. 2020, 63, 4880−4895

Supporting Information

Discovery and Structure-Activity Relationship Study of (Z)-5-Methylenethiazolidin-4-one Derivatives as Potent and Selective Pan-phosphatidylinositol 5-Phosphate 4-Kinase Inhibitors

Theresa D. Manz,[†,‡,⊥,°] Sindhu Carmen Sivakumaren,[†,‡,°] Fleur M. Ferguson,[†,‡] Tinghu Zhang,[†,‡] Adam Yasgar,[Δ] Hyuk-Soo Seo,[†,‡] Scott B. Ficarro,[Ψ,Θ,φ] Joseph D. Card,[Ψ,Θ,φ] Hyeseok Shim,[§] Chandrasekhar V. Miduturu,[†,‡] Anton Simeonov,[Δ] Min Shen,[Δ] Jarrod A. Marto,[Ψ,Θ,φ] Sirano Dhe-Paganon,[†,‡] Matthew D. Hall,[Δ] Lewis C. Cantley,[§] and Nathanael S. Gray[†,‡,*]

[†]Department of Cancer Biology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, MA 02215, USA

[‡]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA

[⊥]Department of Pharmaceutical and Medicinal Chemistry, Saarland University, 66123 Saarbruecken, Germany

[Δ]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD 20850, USA

[Ψ]Department of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Ma 02215, USA

[Θ]Department of Oncologic Pathology, Dana-Farber Cancer Institute, 360 Longwood Avenue, Boston, MA 02215, USA

[φ]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA

[§]Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian Hospital, New York, NY 10065, USA

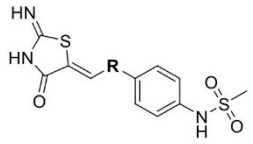
Table of Contents

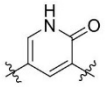
1. Supporting Tables and Figures 3
1.1 Supporting Tables 3
1.2 Supporting Figures 15
1.2.1 High-throughput Screen 15
1.2.2 Mass Spectrometry 17
1.2.3 Co-crystal Structure 19
1.2.4 Lipophilic Efficiency (LipE) 20
1.2.5 Isothermal CETSA 21
1.3 Compound Characterization 22
1.3.1 ^{1}H NMR 22
1.3.2 ^{13}C NMR 22
1.3.3 UPLC 23
2. Abbreviations 24
3. References 24

1. Supporting Tables and Figures

1.1 Supporting Tables

Supporting Table 1. Overview of CVM-05-002 Analogs



Compound ID	R	% DMSO control[a] (PI5P4Kα)	PIM off-target[b]
CVM-05-002		3	-
Analog 1		72	+
Analog 2		18	+
Analog 3		58	+
Analog 4		62	+

[a] Mean of two replicates of the high-throughput screen. [b] Based off of Ambit KINOMEscan data. "+" representing <10% DMSO control for all PIM isoforms tested (PIM1/2/3) at either 1 or 10 µM compound concentration. "-" representing 100% DMSO control for all PIM isoforms tested (PIM1/2/3) at either 1 or 10 µM compound concentration.

Supporting Table 2. Biochemical Activity of CVM-05-002, 3, 13, and 31 for Selected Off-targets[a]

Compound ID	ACVR2A		ACVR2B		TGFBR2		PI4P5Kγ		PI5P4Kγ	
	% DMSO control	IC_{50} [nM]	% DMSO control	IC_{50} [nM]	% DMSO control	IC_{50} [nM]	% DMSO control	IC_{50} [nM]	% DMSO control	K_D [nM]
CVM-05-002	0	3.3	2.5	5.0	0.75	2.1	24	93	0	1.6
3	59	N/A	10	4.4	2	1.1	0	53	2.4	0.58
13	49	1,190	33	344	50	348	26	398	0	3.4
31	100	N/A	86	2,490	97	4,960	0	1.9	0	0.023
THZ-P1-2	94	N/A	91	N/A	97	N/A	20	39	9	4.8

[a] %DMSO control represents the data from the Ambit KINOMEscan selectivity profiling. IC_{50} values were determined in commercially available activity assays by Invitrogen (Thermo Fisher Scientific). K_D values were determined by Ambit.

Supporting Table 3. Full KINOMEscan datasets

Kinase ID	Cpd conc. [μM]	Percent Control			
		CVM-05-002	Cpd 3	Cpd 13	Cpd 31
AAK1	1	96	95	100	86
ABL1(E255K)-phosphorylated	1	85	98	97	62
ABL1(F317I)-nonphosphorylated	1	94	100	100	99
ABL1(F317I)-phosphorylated	1	92	90	88	78
ABL1(F317L)-nonphosphorylated	1	100	95	100	81
ABL1(F317L)-phosphorylated	1	90	83	69	65
ABL1(H396P)-nonphosphorylated	1	83	98	83	47
ABL1(H396P)-phosphorylated	1	84	95	82	60
ABL1(M351T)-phosphorylated	1	85	79	71	57
ABL1(Q252H)-nonphosphorylated	1	82	84	84	59
ABL1(Q252H)-phosphorylated	1	97	100	73	72
ABL1(T315I)-nonphosphorylated	1	81	100	80	79
ABL1(T315I)-phosphorylated	1	100	81	100	96
ABL1(Y253F)-phosphorylated	1	90	98	77	50
ABL1-nonphosphorylated	1	76	60	70	51
ABL1-phosphorylated	1	91	67	85	61
ABL2	1	90	100	94	77
ACVR1	1	51	47	91	88
ACVR1B	1	83	93	100	96
ACVR2A	1	0	59	49	100
ACVR2B	1	2.5	10	33	86
ACVRL1	1	45	74	77	92
ADCK3	1	100	81	100	100
ADCK4	1	94	78	94	86
AKT1	1	100	100	100	100
AKT2	1	97	100	96	94
AKT3	1	98	100	98	100
ALK	1	92	100	90	74
ALK(C1156Y)	1	100	100	93	72
ALK(L1196M)	1	100	100	99	75
AMPK-alpha1	1	90	100	83	82
AMPK-alpha2	1	97	100	99	97
ANKK1	1	100	92	87	91
ARK5	1	93	100	78	87
ASK1	1	90	100	100	98
ASK2	1	100	100	99	100
AURKA	1	65	91	67	46
AURKB	1	100	53	78	26
AURKC	1	91	100	87	66
AXL	1	100	100	83	82
BIKE	1	96	100	92	89
BLK	1	66	100	95	65

BMPR1A	1	52	37	100	99
BMPR1B	1	21	3.3	64	94
BMPR2	1	70	70	79	78
BMX	1	100	71	100	100
BRAF	1	74	78	88	62
BRAF(V600E)	1	88	100	93	71
BRK	1	91	100	86	77
BRSK1	1	75	100	86	75
BRSK2	1	78	98	91	76
BTK	1	64	100	63	61
BUB1	1	83	97	85	69
CAMK1	1	100	95	100	96
CAMK1B	1	100	100	86	100
CAMK1D	1	100	100	100	98
CAMK1G	1	100	100	100	98
CAMK2A	1	100	100	100	100
CAMK2B	1	79	100	95	92
CAMK2D	1	98	100	100	98
CAMK2G	1	99	100	100	100
CAMK4	1	100	100	89	95
CAMKK1	1	99	100	100	99
CAMKK2	1	100	98	100	90
CASK	1	95	99	88	88
CDC2L1	1	95	100	100	94
CDC2L2	1	100	100	100	100
CDC2L5	1	95	66	92	94
CDK11	1	99	100	100	89
CDK2	1	89	100	98	98
CDK3	1	100	100	100	100
CDK4	1	100	100	72	91
CDK4-cyclinD1	1	88	98	69	83
CDK4-cyclinD3	1	100	100	87	96
CDK5	1	79	100	84	89
CDK7	1	91	100	87	81
CDK8	1	100	100	93	100
CDK9	1	97	100	96	87
CDKL1	1	97	76	99	93
CDKL2	1	77	100	65	75
CDKL3	1	100	100	100	97
CDKL5	1	98	90	95	100
CHEK1	1	89	100	93	92
CHEK2	1	100	100	100	100
CIT	1	75	93	84	81
CLK1	1	99	90	67	79
CLK2	1	94	100	98	100
CLK3	1	98	100	90	97

CLK4	1	100	100	100	81
CSF1R	1	81	100	88	78
CSF1R-autoinhibited	1	100	80	100	100
CSK	1	100	100	100	85
CSNK1A1	1	72	100	90	56
CSNK1A1L	1	95	100	94	90
CSNK1D	1	67	100	98	72
CSNK1E	1	71	100	87	54
CSNK1G1	1	65	100	100	30
CSNK1G2	1	83	95	100	40
CSNK1G3	1	66	45	66	27
CSNK2A1	1	99	30	95	94
CSNK2A2	1	91	27	95	76
CTK	1	100	100	78	100
DAPK1	1	60	20	100	47
DAPK2	1	59	15	99	51
DAPK3	1	70	55	100	63
DCAMKL1	1	96	84	99	91
DCAMKL2	1	100	100	100	100
DCAMKL3	1	55	91	93	21
DDR1	1	98	100	92	94
DDR2	1	96	96	79	90
DLK	1	93	85	78	80
DMPK	1	92	96	95	100
DMPK2	1	95	100	100	97
DRAK1	1	12	11	81	56
DRAK2	1	8	24	85	66
DYRK1A	1	93	72	84	97
DYRK1B	1	65	100	97	66
DYRK2	1	90	84	87	69
EGFR	1	100	91	100	86
EGFR(E746-A750del)	1	100	100	100	88
EGFR(G719C)	1	92	100	87	30
EGFR(G719S)	1	94	100	100	67
EGFR(L747-E749del, A750P)	1	100	92	100	100
EGFR(L747-S752del, P753S)	1	100	96	100	100
EGFR(L747-T751del,Sins)	1	83	100	92	89
EGFR(L858R)	1	100	94	100	100
EGFR(L858R,T790M)	1	97	78	86	81
EGFR(L861Q)	1	91	100	95	66
EGFR(S752-I759del)	1	82	100	100	87
EGFR(T790M)	1	86	92	94	72
EIF2AK1	1	100	100	97	95
EPHA1	1	89	100	97	80
EPHA2	1	100	77	100	100
EPHA3	1	95	100	100	98

EPHA4	1	100	78	94	100
EPHA5	1	82	100	96	88
EPHA6	1	87	100	94	80
EPHA7	1	90	100	93	88
EPHA8	1	96	100	92	91
EPHB1	1	97	100	98	91
EPHB2	1	100	100	97	99
EPHB3	1	88	100	95	83
EPHB4	1	92	100	94	89
EPHB6	1	92	100	89	40
ERBB2	1	100	100	100	99
ERBB3	1	76	100	83	75
ERBB4	1	99	100	100	73
ERK1	1	86	100	100	95
ERK2	1	91	100	91	89
ERK3	1	96	100	95	100
ERK4	1	96	100	100	93
ERK5	1	98	100	99	92
ERK8	1	100	99	58	83
ERN1	1	72	97	82	56
FAK	1	97	100	99	94
FER	1	97	96	100	100
FES	1	97	100	94	89
FGFR1	1	93	99	100	90
FGFR2	1	100	90	100	100
FGFR3	1	100	90	100	97
FGFR3(G697C)	1	96	100	96	82
FGFR4	1	70	100	66	52
FGR	1	85	100	92	73
FLT1	1	97	100	91	85
FLT3	1	97	100	77	80
FLT3(D835H)	1	97	100	96	100
FLT3(D835V)	1	59	100	29	63
FLT3(D835Y)	1	60	100	66	75
FLT3(ITD)	1	89	100	56	94
FLT3(ITD,D835V)	1	87	100	50	84
FLT3(ITD,F691L)	1	59	90	76	84
FLT3(K663Q)	1	73	100	88	79
FLT3(N841I)	1	81	89	100	56
FLT3(R834Q)	1	97	73	82	86
FLT3-autoinhibited	1	100	89	95	97
FLT4	1	79	97	90	87
FRK	1	65	100	52	49
FYN	1	89	81	91	91
GAK	1	100	97	85	62
GCN2(Kin.Dom.2,S808G)	1	94	73	94	97

GRK1	1	91	85	89	84
GRK2	1	100	94	98	97
GRK3	1	100	67	100	100
GRK4	1	90	91	67	80
GRK7	1	96	100	90	84
GSK3A	1	89	70	92	91
GSK3B	1	100	81	90	100
HASPIN	1	100	78	94	86
HCK	1	91	100	88	77
HIPK1	1	81	51	94	68
HIPK2	1	92	48	94	95
HIPK3	1	100	45	100	95
HIPK4	1	82	65	95	65
HPK1	1	95	100	100	85
HUNK	1	90	92	100	92
ICK	1	100	88	100	98
IGF1R	1	99	93	97	99
IKK-alpha	1	77	95	92	91
IKK-beta	1	79	79	94	92
IKK-epsilon	1	100	100	95	93
INSR	1	79	86	80	78
INSRR	1	100	94	100	100
IRAK1	1	92	92	92	95
IRAK3	1	81	100	77	60
IRAK4	1	90	83	86	91
ITK	1	100	100	99	96
JAK1(JH1domain-catalytic)	1	97	100	93	93
JAK1(JH2domain-pseudokinase)	1	100	79	99	96
JAK2(JH1domain-catalytic)	1	100	95	85	63
JAK3(JH1domain-catalytic)	1	74	100	78	76
JNK1	1	90	80	90	98
JNK2	1	88	74	84	98
JNK3	1	87	89	85	92
KIT	1	85	100	42	52
KIT(A829P)	1	77	87	76	70
KIT(D816H)	1	95	100	92	89
KIT(D816V)	1	89	95	100	100
KIT(L576P)	1	100	100	48	75
KIT(V559D)	1	93	100	34	46
KIT(V559D,T670I)	1	94	100	93	81
KIT(V559D,V654A)	1	98	96	92	98
KIT-autoinhibited	1	100	97	98	99
LATS1	1	90	96	90	97
LATS2	1	100	100	100	91
LCK	1	93	97	99	76
LIMK1	1	100	93	100	93

LIMK2	1	92	100	100	96
LKB1	1	83	100	100	86
LOK	1	97	95	94	75
LRRK2	1	87	100	85	97
LRRK2(G2019S)	1	87	100	88	83
LTK	1	84	100	83	60
LYN	1	88	100	92	86
LZK	1	100	100	97	97
MAK	1	100	100	100	100
MAP3K1	1	100	100	92	95
MAP3K15	1	100	85	93	93
MAP3K2	1	94	92	92	87
MAP3K3	1	96	100	96	92
MAP3K4	1	87	100	94	73
MAP4K2	1	63	85	100	100
MAP4K3	1	89	100	83	92
MAP4K4	1	80	100	42	14
MAP4K5	1	71	100	91	75
MAPKAPK2	1	94	100	73	100
MAPKAPK5	1	98	83	94	100
MARK1	1	84	91	80	87
MARK2	1	66	81	81	71
MARK3	1	86	80	93	100
MARK4	1	92	100	62	69
MAST1	1	100	100	100	100
MEK1	1	94	88	93	86
MEK2	1	94	84	91	90
MEK3	1	80	0	100	93
MEK4	1	100	98	98	46
MEK5	1	30	69	41	43
MEK6	1	96	100	100	94
MELK	1	100	99	100	100
MERTK	1	100	100	100	100
MET	1	88	99	92	93
MET(M1250T)	1	88	100	95	90
MET(Y1235D)	1	100	100	90	97
MINK	1	81	94	35	17
MKK7	1	87	97	70	86
MKNK1	1	87	74	82	89
MKNK2	1	88	76	33	73
MLCK	1	97	91	89	19
MLK1	1	89	100	90	84
MLK2	1	93	77	97	82
MLK3	1	87	91	94	75
MRCKA	1	96	64	90	89
MRCKB	1	100	100	91	100

MST1	1	85	100	76	77
MST1R	1	90	100	100	99
MST2	1	76	100	78	86
MST3	1	85	98	98	88
MST4	1	100	100	96	97
MTOR	1	99	37	94	19
MUSK	1	100	94	99	100
MYLK	1	100	88	97	59
MYLK2	1	88	100	99	66
MYLK4	1	93	89	83	66
MYO3A	1	100	100	100	56
MYO3B	1	100	100	88	69
NDR1	1	100	99	81	89
NDR2	1	100	57	93	100
NEK1	1	98	89	100	97
NEK10	1	100	18	100	100
NEK11	1	95	100	95	92
NEK2	1	100	98	98	97
NEK3	1	98	82	89	86
NEK4	1	98	100	88	92
NEK5	1	91	100	100	92
NEK6	1	99	100	93	93
NEK7	1	100	100	92	94
NEK9	1	97	100	98	98
NIK	1	100	100	93	73
NIM1	1	84	95	80	86
NLK	1	83	91	84	53
OSR1	1	100	95	99	83
p38-alpha	1	99	100	100	100
p38-beta	1	93	100	100	94
p38-delta	1	88	100	98	97
p38-gamma	1	43	100	72	30
PAK1	1	100	98	100	100
PAK2	1	97	100	100	95
PAK3	1	79	100	85	69
PAK4	1	100	100	100	99
PAK6	1	100	100	100	100
PAK7	1	97	98	100	100
PCTK1	1	90	93	62	97
PCTK2	1	97	100	98	93
PCTK3	1	100	100	100	99
PDGFRA	1	92	86	79	73
PDGFRB	1	83	100	52	43
PDPK1	1	92	100	100	95
PFCDPK1(P.falciparum)	1	82	95	67	47
PFPK5(P.falciparum)	1	100	69	79	89

PFTAIRE2	1	98	100	94	97
PFTK1	1	100	100	95	99
PHKG1	1	95	100	98	94
PHKG2	1	97	100	100	97
PIK3C2B	1	85	95	76	45
PIK3C2G	1	99	100	100	88
PIK3CA	1	77	91	78	67
PIK3CA(C420R)	1	96	100	88	73
PIK3CA(E542K)	1	87	68	81	87
PIK3CA(E545A)	1	92	100	87	80
PIK3CA(E545K)	1	93	68	86	93
PIK3CA(H1047L)	1	100	51	86	85
PIK3CA(H1047Y)	1	100	100	100	100
PIK3CA(I800L)	1	86	46	77	82
PIK3CA(M1043I)	1	79	100	88	57
PIK3CA(Q546K)	1	88	96	80	86
PIK3CB	1	93	97	88	96
PIK3CD	1	78	100	83	55
PIK3CG	1	67	42	67	53
PIK4CB	1	94	52	77	8.8
PIKFYVE	1	95	7.2	100	11
PIM1	1	100	52	100	93
PIM2	1	100	45	100	95
PIM3	1	100	50	100	97
PIP5K1A	1	88	92	94	77
PIP5K1C	1	24	0	26	0
PIP5K2B	1	90	96	89	100
PIP5K2C	1	0	2.4	0	0
PKAC-alpha	1	100	100	100	99
PKAC-beta	1	99	83	99	100
PKMYT1	1	94	88	93	93
PKN1	1	100	100	97	96
PKN2	1	96	98	88	93
PKNB(M.tuberculosis)	1	93	91	100	34
PLK1	1	85	66	85	91
PLK2	1	85	56	89	79
PLK3	1	90	87	90	81
PLK4	1	74	88	80	50
PRKCD	1	100	100	100	96
PRKCE	1	97	100	100	94
PRKCH	1	95	100	95	80
PRKCI	1	97	93	89	100
PRKCQ	1	93	90	100	94
PRKD1	1	72	97	80	85
PRKD2	1	99	100	97	100
PRKD3	1	88	100	86	91

PRKG1	1	97	87	95	96
PRKG2	1	97	100	98	97
PRKR	1	98	59	100	76
PRKX	1	84	100	88	79
PRP4	1	80	100	88	82
PYK2	1	100	100	96	97
QSK	1	100	78	91	100
RAF1	1	95	99	100	100
RET	1	98	100	86	78
RET(M918T)	1	100	100	90	91
RET(V804L)	1	100	100	94	100
RET(V804M)	1	91	100	93	83
RIOK1	1	100	100	78	89
RIOK2	1	100	81	97	85
RIOK3	1	94	100	93	93
RIPK1	1	93	100	97	99
RIPK2	1	34	52	19	24
RIPK4	1	64	74	87	86
RIPK5	1	98	90	92	77
ROCK1	1	99	89	84	92
ROCK2	1	93	100	94	89
ROS1	1	93	94	100	77
RPS6KA4(Kin.Dom.1-N-terminal)	1	100	76	93	96
RPS6KA4(Kin.Dom.2-C-terminal)	1	88	44	87	94
RPS6KA5(Kin.Dom.1-N-terminal)	1	92	100	94	82
RPS6KA5(Kin.Dom.2-C-terminal)	1	98	62	88	100
RSK1(Kin.Dom.1-N-terminal)	1	93	95	88	77
RSK1(Kin.Dom.2-C-terminal)	1	83	95	99	35
RSK2(Kin.Dom.1-N-terminal)	1	90	90	82	63
RSK2(Kin.Dom.2-C-terminal)	1	84	83	75	79
RSK3(Kin.Dom.1-N-terminal)	1	100	100	99	100
RSK3(Kin.Dom.2-C-terminal)	1	84	91	99	71
RSK4(Kin.Dom.1-N-terminal)	1	89	99	91	70
RSK4(Kin.Dom.2-C-terminal)	1	94	99	95	88
S6K1	1	95	100	75	51
SBK1	1	95	100	100	86
SGK	1	89	99	99	61
SgK110	1	98	99	95	100
SGK2	1	100	100	100	95
SGK3	1	100	75	87	91
SIK	1	99	100	95	97
SIK2	1	91	99	62	100
SLK	1	70	53	100	97
SNARK	1	100	82	96	88
SNRK	1	89	100	89	83
SRC	1	100	100	100	82

SRMS	1	99	95	83	91
SRPK1	1	53	95	60	62
SRPK2	1	100	99	86	100
SRPK3	1	100	89	54	93
STK16	1	100	100	100	92
STK33	1	100	100	96	86
STK35	1	77	99	81	83
STK36	1	91	100	99	93
STK39	1	93	69	94	76
SYK	1	90	91	95	94
TAK1	1	50	94	90	60
TAOK1	1	77	92	81	70
TAOK2	1	95	95	85	86
TAOK3	1	95	87	84	81
TBK1	1	85	95	96	80
TEC	1	96	100	95	91
TESK1	1	79	100	79	68
TGFBR1	1	90	96	100	100
TGFBR2	1	0.75	2	50	97
TIE1	1	100	99	100	93
TIE2	1	89	100	99	87
TLK1	1	87	75	97	96
TLK2	1	98	100	94	80
TNIK	1	57	43	23	12
TNK1	1	89	42	100	100
TNK2	1	94	96	94	62
TNNI3K	1	36	87	20	28
TRKA	1	91	100	80	74
TRKB	1	100	100	93	83
TRKC	1	94	100	85	84
TRPM6	1	95	100	71	87
TSSK1B	1	98	96	100	100
TSSK3	1	100	94	88	93
TTK	1	88	100	94	83
TXK	1	86	100	83	63
TYK2(JH1domain-catalytic)	1	63	90	92	100
TYK2(JH2domain-pseudokinase)	1	95	61	80	82
TYRO3	1	96	77	100	95
ULK1	1	100	87	98	93
ULK2	1	80	96	90	86
ULK3	1	100	83	93	70
VEGFR2	1	95	84	96	78
VPS34	1	93	88	92	89
VRK2	1	96	92	51	82
WEE1	1	97	100	100	100
WEE2	1	100	100	99	100

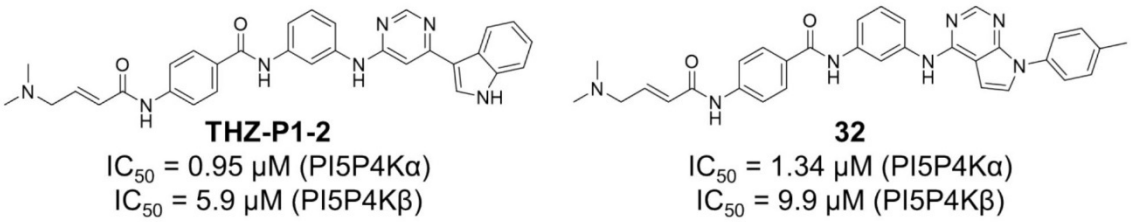
WNK1	1	100	100	94	100
WNK2	1	93	100	95	91
WNK3	1	100	100	93	96
WNK4	1	95	100	99	97
YANK1	1	85	100	85	77
YANK2	1	97	100	90	94
YANK3	1	98	98	97	96
YES	1	89	100	99	79
YSK1	1	98	100	100	97
YSK4	1	100	70	81	62
ZAK	1	72	95	32	58
ZAP70	1	85	96	74	77

Supporting Table 4. Lipophilicity and Early ADME Overview of CVM-05-002 and Analogs

Compound ID	LogD	LipE		Solubility [µg/mL]	Avg. Pe.[a] [10^{-6}cm/s]	Microsome Stability[b] ; $t_{1/2}$ [min]
		PI5P4Kα	PI5P4Kβ			
CVM-05-002	0.66	5.9	5.1	<1	<1	>30
Cpd 3	-0.05	6.9	5.8	>55	<1	>30
Cpd 13	2.52	3.2	2.5	5.0	high permeability	>30
Cpd 31	0.41	5.8	5.2	6.2	<1	>30

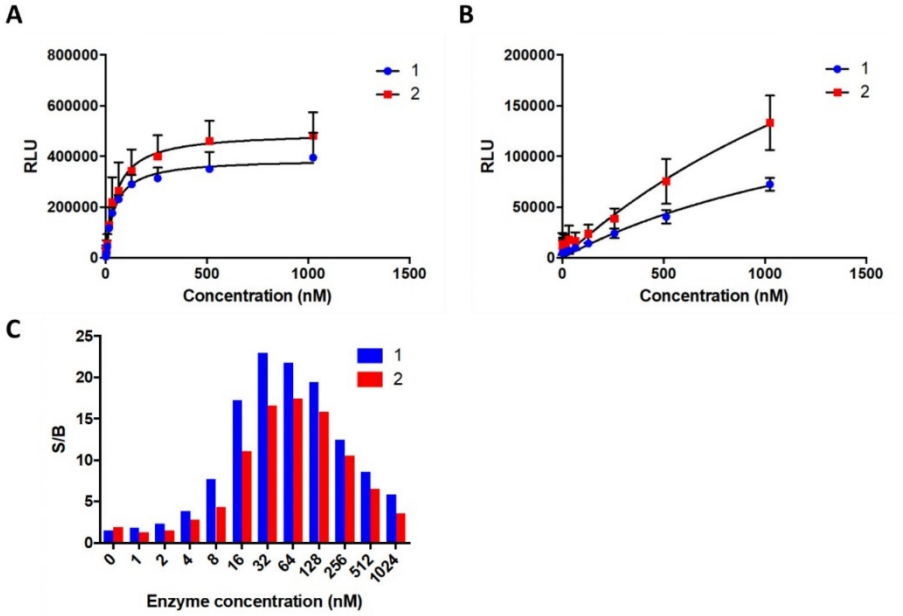
[a] PAMPA average permeability. [b] Rat liver microsome stability assay.

1.2 Supporting Figures

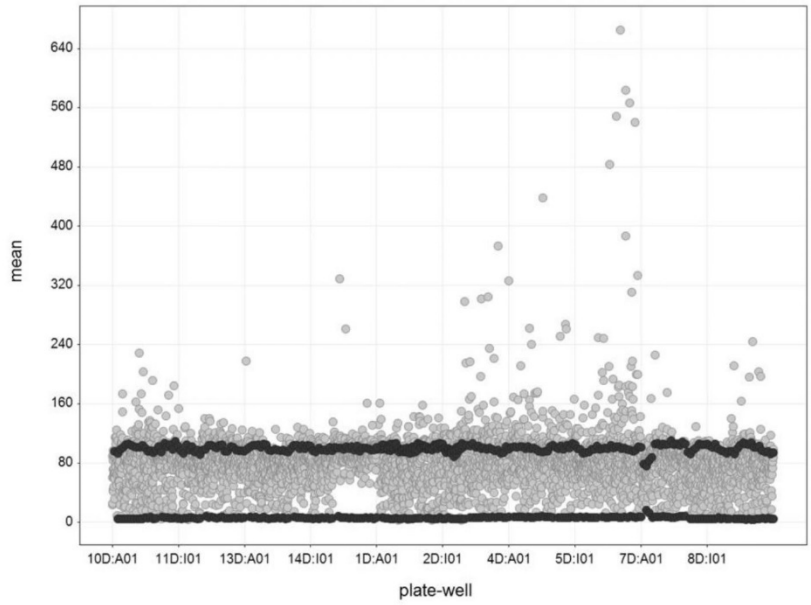


Supporting Figure 1. Structures of reported covalent pan-PI5P4K inhibitors THZ-P1-2 and compound **32** (named "compound **30**" in the original publication).[1]

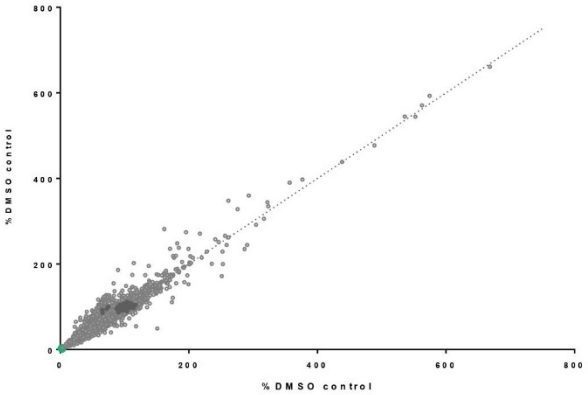
1.2.1 High-throughput Screen



Supporting Figure 2. Titration of PI5P4Kα with (**A**) or without (**B**) lipid and RLU read-out. Condition 1 represents addition of 10 µL ADP-Glo reagent 1 and 20 µL ADP-Glo reagent 2. Condition 2 represents addition of 5 µL ADP-Glo reagent 1 and 10 µL ADP-Glo reagent 2. **C**. Signal to Background ratio in response to enzyme concentration used (Background defined as 'in absence of lipid'). Condition 1 represents addition of 10 µL ADP-Glo reagent 1 and 20 µL ADP-Glo reagent 2. Condition 2 represents addition of 5 µL ADP-Glo reagent 1 and 10 µL ADP-Glo reagent 2.

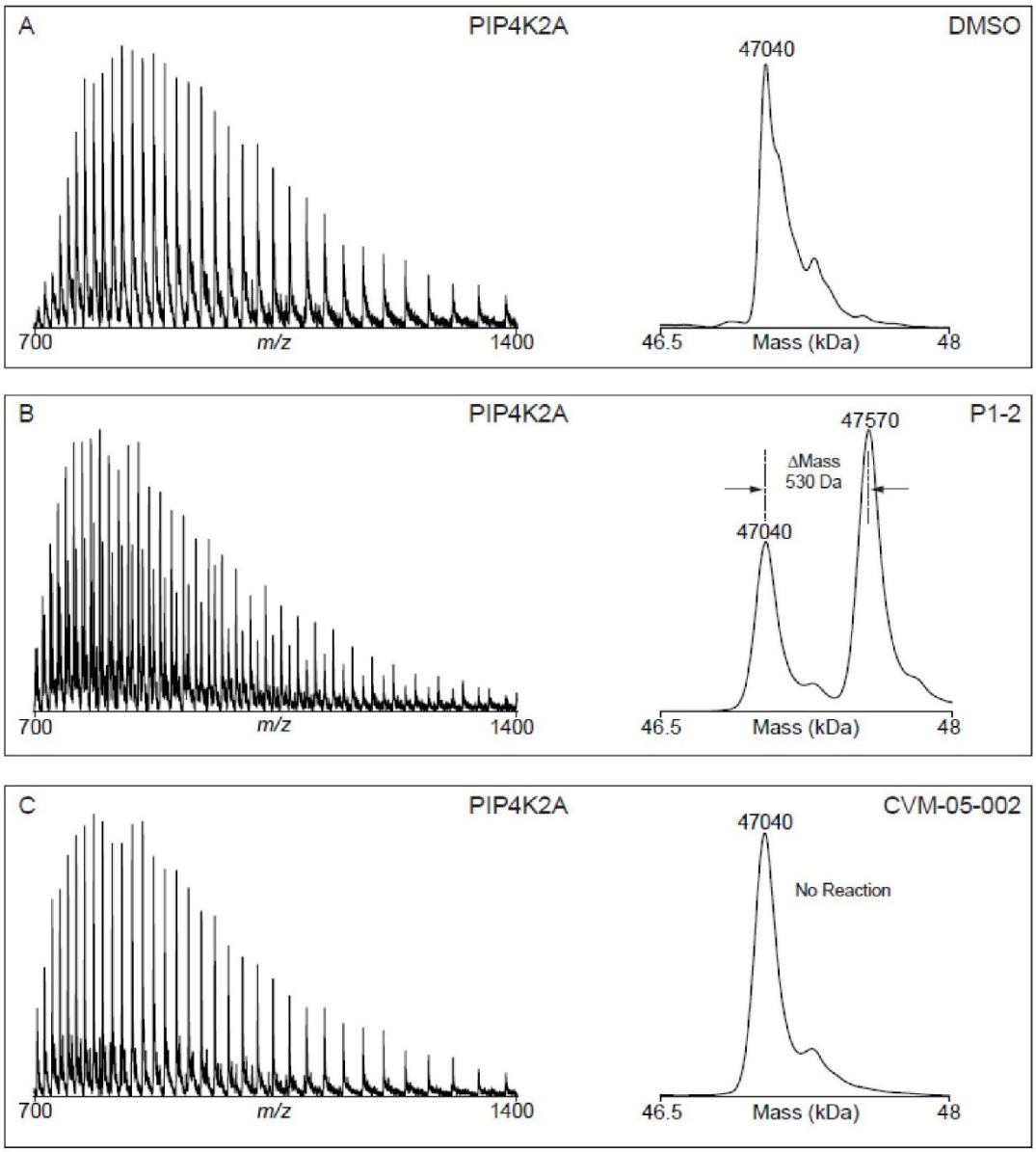


Supporting Figure 3. Plate-well series plot with mean %DMSO control of two replicates normalized to DMSO (100%, upper black circles) for each plate, demonstrating overall assay stability and low plate-to-plate variability. (Lower black circles represent positive controls.)

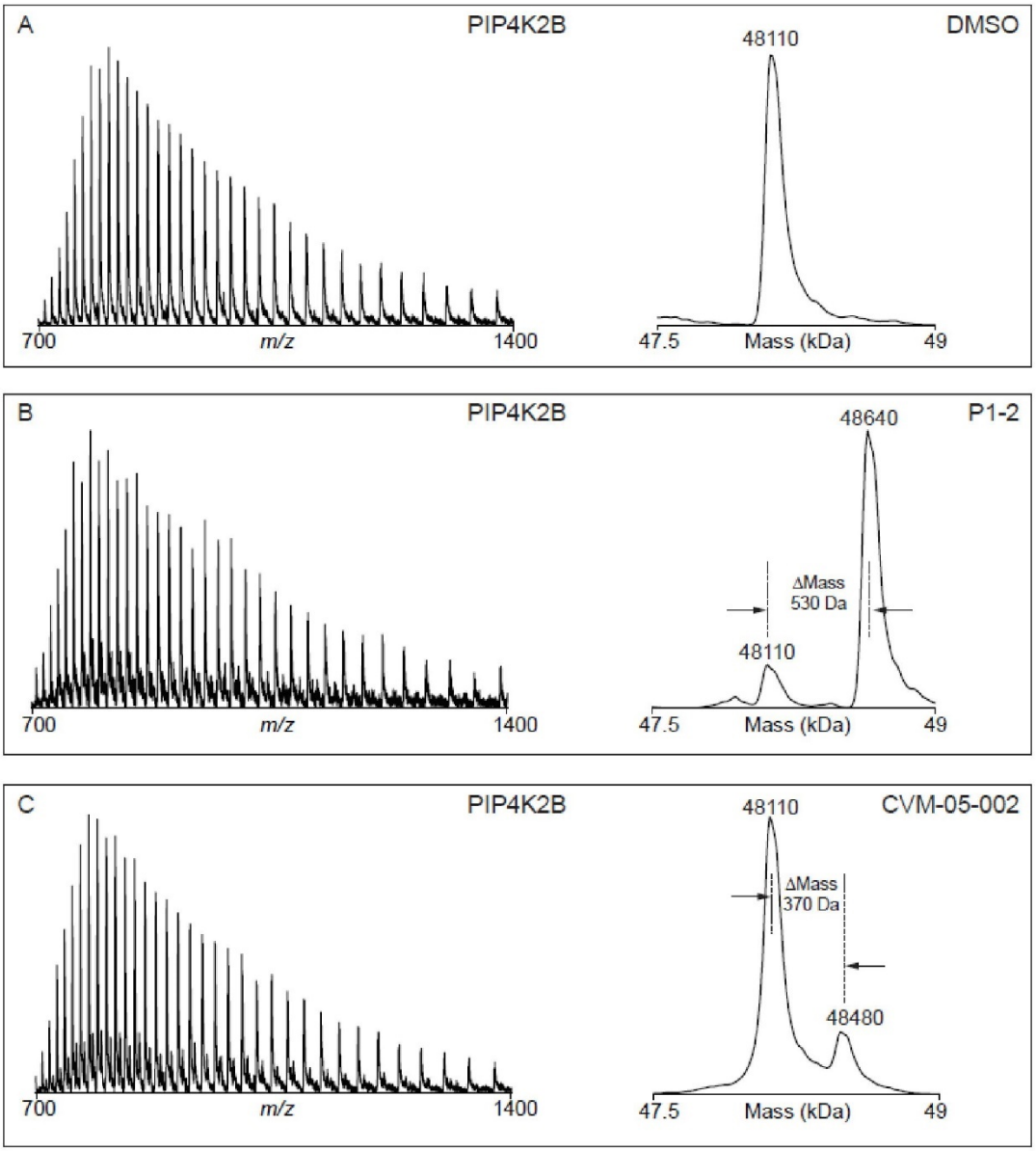


Supporting Figure 4. Complete results of high-throughput screen. CVM-05-002 is depicted in green, DMSO (negative) controls are shown in dark gray. A dotted line marks the diagonal (replicate 1 = replicate 2).

1.2.2 Mass Spectrometry

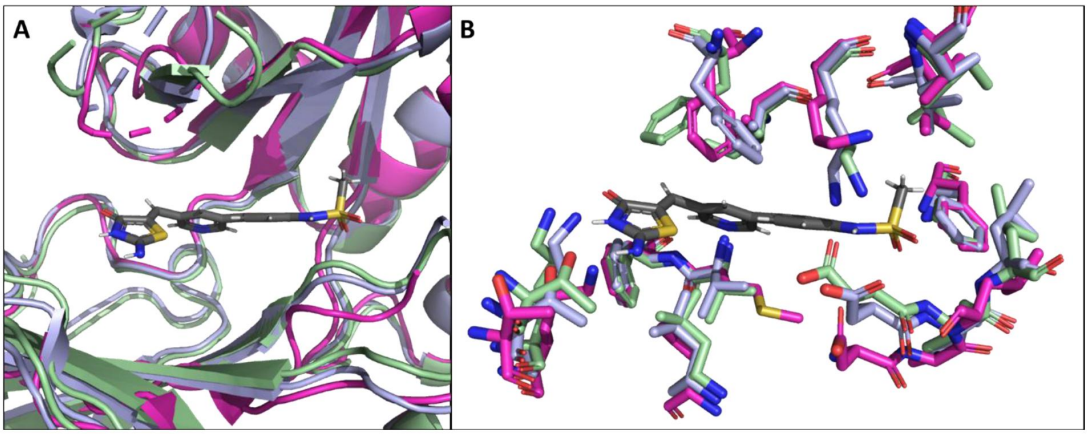


Supporting Figure 5. Mass spectra (left) and zero-charge mass spectra (right) obtained from PI5P4Kα treated with DMSO (**A**), 10 µM THZ-P1-2 (as positive control) (**B**), or 10 µM CVM-05-002 (**C**) for 2 hours at room temperature.

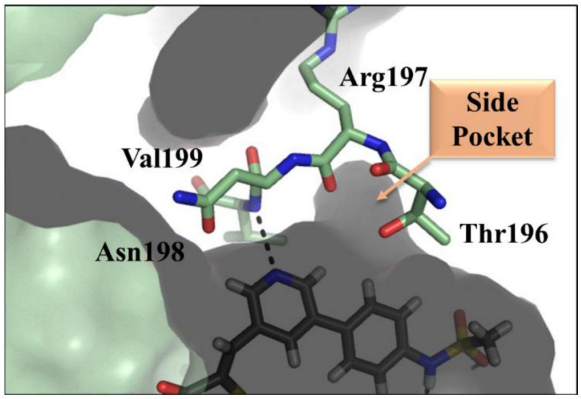


Supporting Figure 6. Mass spectra (left) and zero-charge mass spectra (right) obtained from PI5P4Kβ treated with DMSO (**A**), 10 μM THZ-P1-2 (as positive control) (**B**), or 10 μM CVM-05-002 (**C**) for 2 hours at room temperature.

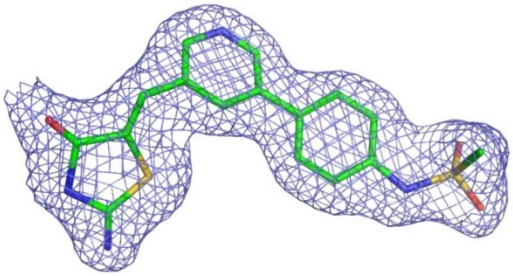
1.2.3 Co-crystal Structure



Supporting Figure 7. Co-crystal structure of PI5P4Kα with bound inhibitor CVM-05-002 was aligned with PI5P4Kβ (PDB ID: 3WZZ) and PI5P4Kγ (PDB ID: 2GK9.A). **A.** Shown is the ATP-binding pocket with inhibitor CMV-05-002 (dark gray). The lipid kinases are depicted as cartoons, showing PI5P4Kα (pale green), -β (pale blue), and -γ (magenta) superimposed. **B.** Amino acids of the PI5P4Kα (pale green) binding pocket, highlighted in Figure 2, and respective amino acids of PI5P4Kβ (pale blue) and -γ (magenta) are depicted as sticks. CVM-05-002 is shown in dark gray.

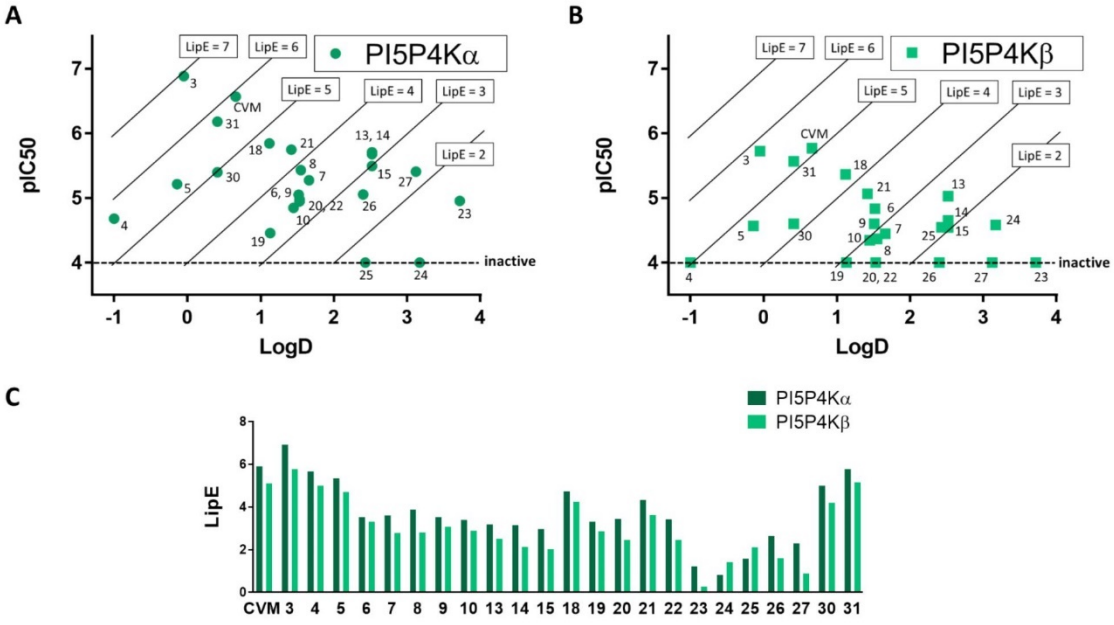


Supporting Figure 8. CVM-05-002 bound to PI5P4Kα. Labeled is a pocket next to the pyridine moiety of the inhibitor at the hinge region of the lipid kinase.



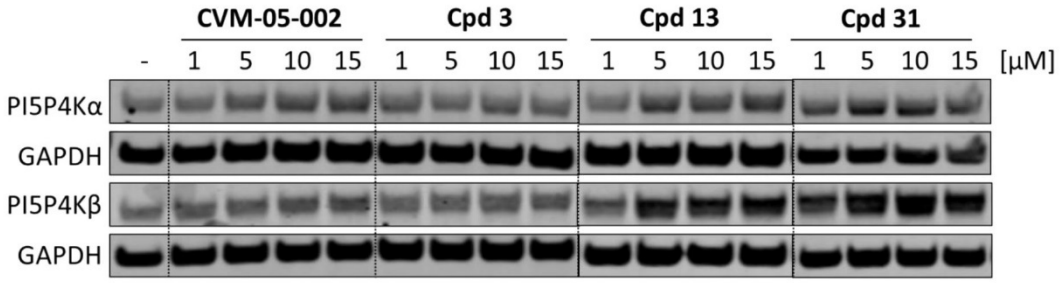
Supporting Figure 9. 2fo-fc electron density map of ligand (CVM-05-002) in co-crystal structure of PI5P4Kα with CVM-05-002.

1.2.4 Lipophilic Efficiency (LipE)



Supporting Figure 10. A. Graphical depiction of lipophilic ligand efficiency (LipE) for PI5P4Kα. Plotted are LogD vs. pIC_{50}. **B.** Graphical depiction of lipophilic ligand efficiency (LipE) for PI5P4Kβ. Plotted are LogD vs. pIC_{50}. **C.** Summary of LipE values for PI5P4Kα and -β, side-by-side.

1.2.5 Isothermal CETSA

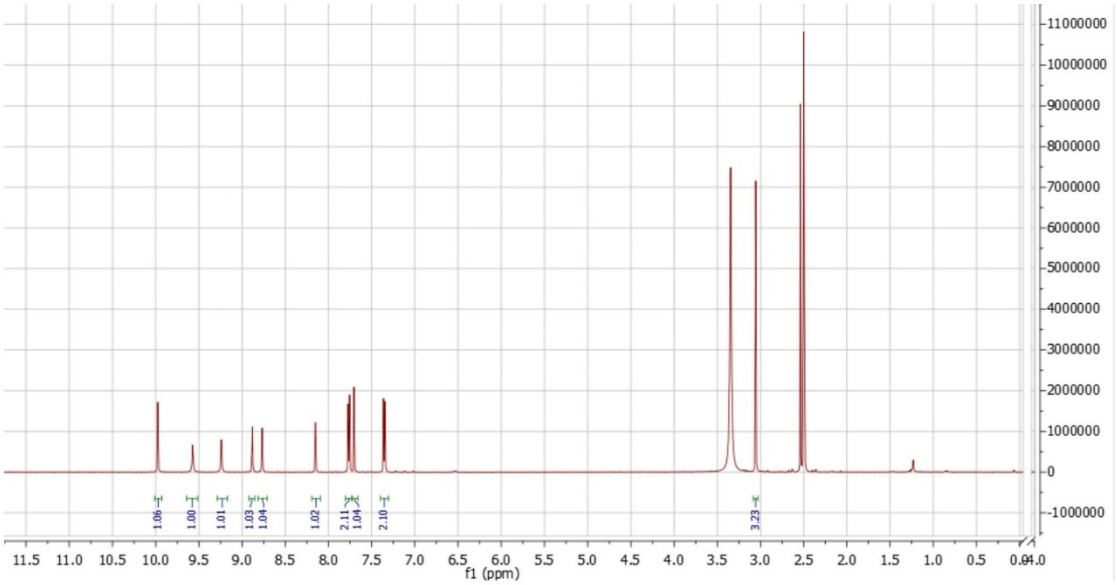


Supporting Figure 11. Representative Western blot of isothermal CETSA experiment. HEK 293T cells were treated with 1, 5, 10, or 15 μM inhibitor or DMSO. After 1 h at 37 °C, the cells were incubated at 49 °C for 3 min, lysed, and the lysates were used for Western blot analysis.

1.3 Compound Characterization

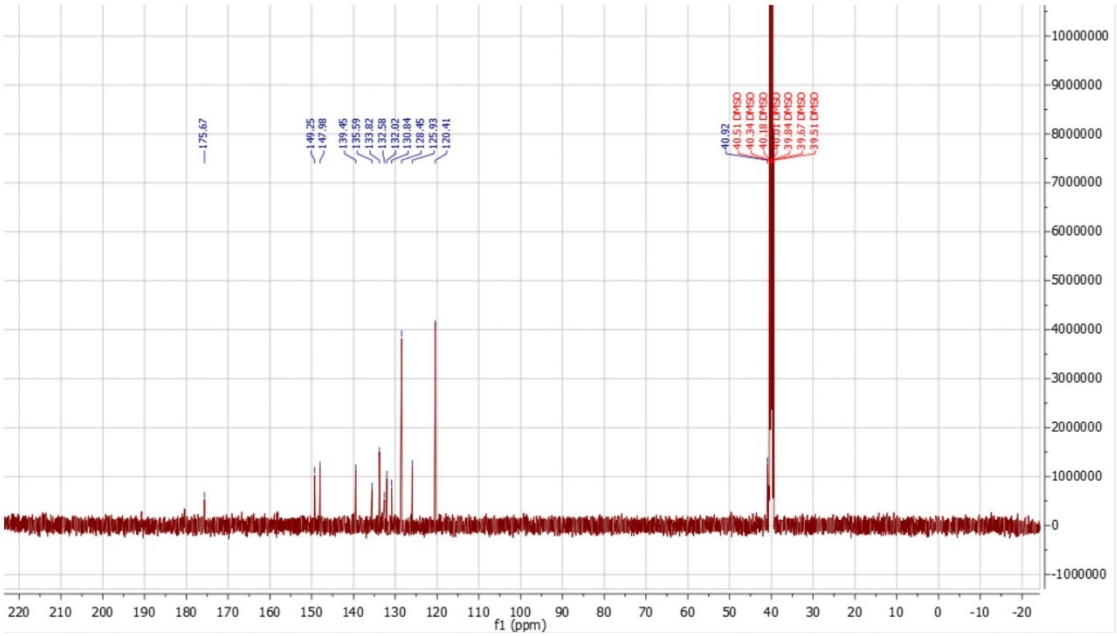
1.3.1 ^{1}H NMR

CVM-05-002



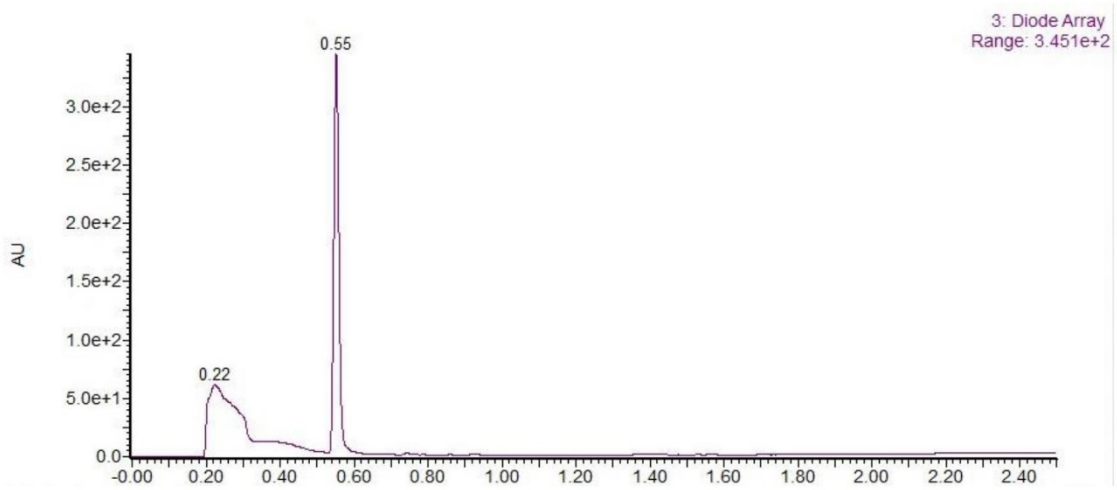
1.3.2 ^{13}C NMR

CVM-05-002

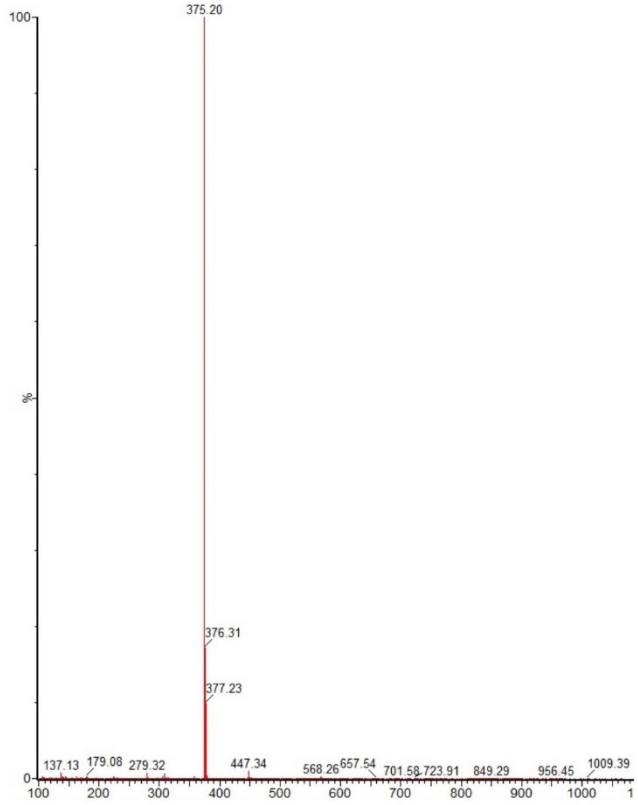


1.3.3 UPLC

CVM-05-002 – UV



CVM-05-002 – Mass (peak, 0.55 min)



2. Abbreviations

CETSA, cellular thermal shift assay; GAPDH, glyceraldehyde 3-phsophate dehydrogenase; PI5P4K, phosphatidylinositol 5-phosphate 4-kinase.

3. References

(1) Manz, T. D.; Sivakumaren, S. C.; Yasgar, A.; Hall, M. D.; Davis, M. I.; Seo, H.-S.; Card, J. D.; Shim, H.; Ficarro, S. B.; Dhe-Paganon, S.; Sasaki, A. T.; Boxer, M. B.; Simeonov, A.; Cantley, L. C.; Shen, M.; Zhang, T.; Ferguson, F. M.; Gray, N. S. Structure-activity relationship study of covalent pan-phosphatidylinositol 5-phosphate 4-kinase inhibitors. *ACS Med. Chem. Lett.* **2019**, *11* (3), 346-352.

Data Summary (including compound SMILES):

Compound ID	SMILES	Biophysical Data (IC50; nM)						KD (nM)
		PI5P4 Kα	PI5P4 Kβ	ACVR 2A	ACVR 2B	TGFBR 2	PI4P5 Kγ	PI5P4 Kγ
CVM-05-002	CS(=O)(=O)NC1=CC=C(C=C1)C1=CN=CC(\C=C2/SC(=N)NC2=O)=C1	0.27	1.7	3.3	5	2.1	93	1.6
3	CS(=O)(=O)NC1=CC=C(C=C1)C1=CN=CC(\C=C2/SC(=O)NC2=O)=C1	0.13	1.9		4.4	1.1	53	0.58
4	[H]/C(C1=CN=CC(C2=CC=C(NS(C)(=O)=O)C=C2)=C1)=C3OC(NC/3=O)=O	21	>50					
5	[H]/C(C1=CN=CC(C2=CC=C(NS(C)(=O)=O)C=C2)=C1)=C3CC(NC/3=O)=O	6.1	27					
6	CS(=O)(=O)NC1=CC=C(C=C1)C1=CN=CC(=C1)C1=CC=CC2=C1CC(=O)N2	8.9	15					
7	CC(=O)NC1=CC=CC(=C1)C1=CN=CC(=C1)C1=CC=C(NS(C)(=O)=O)C=C1	5.3	36					
8	CS(NC(C=C1)=CC=C1C2=CC(C3=CC=C(NC(NC)=O)C=C3)=CN=C2)(=O)=O	3.7	43					
9	CC(=O)NC1=CC=CC(OC2=CN=CC(=C2)C2=CC=C(NS(C)(=O)=O)C=C2)=C1	14	45					
10	CS(NC(C=C1)=CC=C1C2=CC(NC3=CC=CC(NC(C)=O)=C3)=CN=C2)(=O)=O	14	45					

Compound ID	SMILES	Biophysical Data (IC50; nM)						KD (nM)
		PI5P4 Kα	PI5P4 Kβ	ACVR 2A	ACVR 2B	TGFBR 2	PI4P5 Kγ	PI5P4 Kγ
13	O=S(NC(C=C1)=CC=C1C2=CC(C3=C(C=CN4)C4=CC=C3)=CN=C2)(C)=O	2	9.4	1190	344	348	398	3.4
14	O=S(NC(C=C1)=CC=C1C2=CC(C3=CC(NC=C4)=C4C=C3)=CN=C2)(C)=O	2.1	22					
15	O=S(NC(C=C1)=CC=C1C2=CC(C3=CC(C=CN4)=C4C=C3)=CN=C2)(C)=O	3.2	29					
18	[H]/C(C1=CN=CC(C2=CC=C(S(NC)(=O)=O)C=C2)=C1)=C3SC(NC/3=O)=N	0.27	1.7					
19	[H]/C(C1=CN=CC(C2=CC=C(S(C)(=O)=O)C=C2)=C1)=C3SC(NC/3=O)=N	1.4	4.3					
20	CC(=O)NC1=CC=C(C=C1)C1=CN=CC(\C=C2/SC(=N)NC2=O)=C1	35	>50					
21	[H]/C(C1=CN=CC(C2=CC=C(NC(NC)=O)C=C2)=C1)=C3SC(NC/3=O)=N	11	>50					
22	CC(=O)NC1=CC=CC(=C1)C1=CN=CC(\C=C2/SC(=N)NC2=O)=C1	1.8	8.6					
23	[H]/C(C1=CN=CC(C2=CC=C(OC(F)(F)F)C=C2)=C1)=C3SC(NC/3=O)=N	11	>50					

Compound ID	SMILES	Biophysical Data (IC50; nM)						KD (nM)
		PI5P4 Kα	PI5P4 Kβ	ACVR 2A	ACVR 2B	TGFBR 2	PI4P5 Kγ	PI5P4 Kγ
24	FC(F)(F)C1=CC=C(C=C1)C1=CN=CC(\C=C2/SC(=N)NC2=O)=C1	>50	26					
25	FC1=CC=C(C=C1)C1=CN=CC(\C=C2/SC(=N)NC2=O)=C1	>50	28					
26	[H]/C(C1=CN=CC(C2=CC=C(N(C)C)C=C2)=C1)=C3SC(NC/3=O)=N	8.8	>50					
27	[H]/C(C1=CN=CC(C2=CC=C(OC)C3=C2C=CC=C3)=C1)=C4SC(NC/4=O)=N	3.9	>50					
30	CS(=O)(=O)NC1=CC=C(C=C1)C1=CN=C(N)C(\C=C2/SC(=N)NC2=O)=C1	4	25					
31	CS(=O)(=O)NC1=CC=C(C=C1)C1=CC(\C=C2/SC(=N)NC2=O)=CN=C1N	0.66	2.7		2490	4960	1.9	0.023

5.5 References

[1] Xu, J.; Li, X.; Ding, K.; Li, Z. Applications of Activity-Based Protein Profiling (ABPP) and Bioimaging in Drug Discovery. *Chem. Asian. J.* **2020**, *15* (1), 34-41.

[2] Bayer, E. A.; Wilchek, M. Application of avidin-biotin technology to affinity-based separations. *J. Chromatogr.* **1990**, *510*, 3-11.

[3] Dundas, C. M.; Demonte, D.; Park, S. Streptavidin-biotin techology: improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.* **2013**, *97* (21), 9343-9353.

[4] Laitinen, O. H.; Hytönen, V. P.; Nordlund, H. R.; Kulomaa, M. S. Genetically engineered avidins and streptavidins. *Cell Mol. Life Sci.* **2006**, *63* (24), 2992-3017.

[5] Stayton, P. S.; Freitag, S.; Klumb, L. A.; Chilkoti, A.; Chu, V.; Penzotti, J. E.; To, R.; Hyre, D.; Trong, I. L.; Lybrand, T. P.; Stenkamp, R. E. Streptavidin-biotin binding energetics. *Biomol. Eng.* **1999**, *16* (1-4), 39-44.

[6] Sato, S.-i.; Murata, A.; Shirakawa, T.; Uesugi, M. Biochemical Target Isolation for Novices: Affinity-Based Strategies. *Chem. Biol.* **2010**, *17* (6), 616-623.

[7] Wang, S.; Tian, Y.; Wang, M.; Wang, M.; Sun, G.-b.; Sun, X.-b. Advanced Activity-Based Protein Profiling Application Strategies for Drug Development. *Front. Pharmacol.* **2018**, *9*, 353.

[8] Marholz, L. J.; Chang, L.; Old, W. M.; Wang, X. Development of Substrate-Selective Probes for Affinity Pulldown of Histone Demethylases. *ACS Chem. Biol.* **2015**, *10* (1), 129-137.

[9] Melville, D. B.; Dittmer, K.; Brown, G. B.; DU Vigneaud, V. Desthiobiotin. *Science* **1943**, *98* (2553), 497-499.

[10] Hirsch, J. D.; Eslamizar, L.; Filanoski, B. J.; Malekzadeh, N.; Haugland, R. P.; Beechem, J. M.; Haugland, R. P. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.* **2002**, *308* (2), 343-357.

[11] Browne, C. M.; Jiang, B.; Ficarro, S. B.; Doctor, Z. M.; Johnson, J. L.; Card, J. D.; Sivakumaren, S. C.; Alexander, W. M.; Yaron, T. M.; Murphy, C. J.; Kwiatkowski, N. P.; Zhang, T.; Cantley, L. C.; Gray, N. S.; Marto, J. A. A Chemoproteomic Strategy for Direct and Proteome-Wide Covalent Inhibitor Target-Site Identification. *J. Am. Chem. Soc.* **2019**, *141* (1), 191-203.

[12] Redhead, M.; Satchell, R.; McCarthy, C.; Pollack, S.; Unitt, J. Thermal Shift as an Entropy-Driven Effect. *Biochemistry* **2017**, *56* (47), 6187-6199.

[13] Grøftehauge, M. K.; Hajizadeh, N. R.; Swann, M. J.; Pohl, E. Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71*, 36-44.

[14] Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlung, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, *341* (6141), 84-87.

[15] Molina, D. M.; Nordlund, P. The Cellular Thermal Shift Assay: A Novel Biophysical Assay for In Situ Drug Target Engagement and Mechanistic Biomarker Studies. *Annu. Rev. Pharmacol. Toxicol.* **2016**, *56*, 141-161.

[16] Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundbäck, T.; Nordlund, P.; Molina, D. M. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **2014**, *9* (9), 2100-2122.

[17] Dai, L.; Prabhu, N.; Yu, L. Y.; Bacanu, S.; Ramos, A. D.; Nordlund, P. Horizontal Cell Biology: Monitoring Global Changes of Protein Interaction States with the Proteome-Wide Cellular Thermal Shift Assay (CETSA). *Annu. Rev. Biochem.* **2019**, *88*, 383-408.

[18] Prabhu, N.; Dai, L.; Nordlund, P. CETSA in integrated proteomics studies of cellular processes. *Curr. Opin. Chem. Biol.* **2020**, *54*, 54-62.

[19] Kawatkar, A.; Schefter, M.; Hermansson, N.-O.; Snijder, A.; Dekker, N.; Brown, D. G.; Lundbäck, T.; Zhang, A. X.; Castaldi, M. P. CETSA beyond Soluble Targets: a Broad Application to Multipass Transmembrane Proteins. *ACS Chem. Biol.* **2019**, *14* (9), 1913-1920.

[20] Shaw, J.; Dale, I.; Hemsley, P.; Leach, L.; Dekki, N.; Orme, J. P.; Talbot, V.; Narvaez, A. J.; Bista, M.; Molina, D. M.; Dabrowski, M.; Main, M. J.; Gianni, D. Positioning High-Throughput CETSA in Early Drug Discovery through Screening against B-RAF and PARP I. *SLAS Discov.* **2019**, *24* (2), 121-132.

[21] Lundgren, S. Focusing on Relevance: CETSA-Guided Medicinal Chemistry and Lead Generation. *ACS Med. Chem. Lett.* **2019**, *10* (5), 690-693.

[22] Seashore-Ludlow, B.; Axelsson, H.; Almqvist, H.; Dahlgren, B.; Jonsson, M.; Lundbäck, T. Quantitative Interpretation of Intracellular Drug Binding and Kinetics Using the Cellular Thermal Shift Assay. *Biochemistry* **2018**, *57* (48), 6715-6725.

[23] Seashore-Ludlow, B.; Axelsson, H.; Lundbäck, T. Perspective on CETSA Literature: Toward More Quantitative Data Interpretation. *SLAS Discov.* **2020**, *25* (2), 118-126.

[24] Friman, T. Mass spectrometry-based Cellular Thermal Shift Assay (CETSA®) for target deconvolution in phenotypic drug discovery. *Bioorg. Med. Chem.* **2020**, *28* (1), 115174.

[25] Henderson, M. J.; Holbert, M. A.; Simeonov, A.; Kallal, L. A. High-Throughput Cellular Thermal Shift Assays in Research and Drug Discovery. *SLAS Discov.* **2020**, *25* (2), 137-147.

[26] Faguet, G. B. A brief history of cancer: Age-old milestones underlying our current knowledge database. *Int. J. Cancer* **2015**, *136* (9), 2022-2036.

[27] Hajdu, S. I. A Note from History: Landmarks in History of Cancer, Part 1. *Cancer* **2011**, *117* (5), 1097-1102.

[28] Blackadar, C. B. Historical review of the causes of cancer. *World J. Clin. Oncol.* **2016**, *7* (1), 54-86.

[29] World Health Organization. "Cancer; Key facts.", **2018**, https://www.who.int/news-room/fact-sheets/detail/cancer (last checked: 05/06/2020).

[30] Aktipis, C. A.; Boddy, A. M.; Gatenby, R. A.; Brown, J. S.; Maley, C. C. Life history tradeoffs in cancer evolution. *Nat. Rev. Cancer* **2013**, *13* (12), 883-892.

[31] Weinberg, R. A. How Cancer Arises: An explosion of research is uncovering the long-hidden molecular underpinnings of cancer – and suggesting new therapies. *Sci. Am.* **1996**, *275* (3), 62-70.

[32] Hanahan, D.; Weinberg, R. A. The Hallmarks of Cancer. *Cell* **2000**, *100* (1), 57-70.

[33] Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144* (5), 646-674.

[34] Kozlova, N.; Grossman, J. E.; Iwanicki, M. P.; Muranen, T. The Interplay of the Extracellular Matrix and Stromal Cells as a Drug Target in Stroma-Rich Cancers. *Trends Pharmacol. Sci.* **2020**, *41* (3), 183-198.

[35] Aktipis, C. A.; Boddy, A. M.; Gatenby, R. A.; Brown, J. S.; Maley, C. C. Life history tradeoffs in cancer evolution. *Nat. Rev. Cancer* **2013**, *13* (12), 883-892.

[36] Bashraheel, S. S.; Domling, A.; Goda, S. K. Update on targeted cancer therapies, single or in combination, and their fine tuning for precision medicine. *Biomed. Pharmacother.* **2020**, *125*, 110009.

[37] Masood, I.; Kiani, M. H.; Ahmad, M.; Masood, M. I.; Sadaquat, H. Major contributions towards finding a cure for cancer through chemotherapy: a historical review. *Tumori.* **2016**, *102* (1), 6-17.

[38] Bökkerink, J. P.; Stet, E. H.; De Abreu, R. A.; Damen, F. J.; Hulscher, T. W.; Bakker, M. A.; van Baal, J. A. 6-Mercaptopurine: cytotoxicity and biochemical pharmacology in human malignant T-lymphoblasts. *Biochem. Pharmacol.* **1993**, *45* (7), 1455-1463.

[39] Longley, D. B.; Harkin, D. P.; Johnston, P. G. 5-fluorouracil: mechanism of action and clinical strategies. *Nat. Rev. Cancer* **2003**, *3* (5), 330-338.

[40] Dasari, S.; Tchounwou, P. B. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur. J. Pharmacol.* **2014**, *740*, 364-378.

[41] Yoo, H.; Rill, R. L. Actinomycin D Binding to Unstructured, Single-stranded DNA. *J. Mol. Recognit.* **2001**, *14* (3), 145-150.

[42] Jeggo, P. A.; Pearl, L. H.; Carr, A. M. DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer* **2016**, *16* (1), 35-42.

[43] Hengel, S. R.; Spies, M. A.; Spies, M. Small molecule inhibitors targeting DNA repair and DNA repair deficiency in research and cancer therapy. *Cell Chem. Biol.* **2017**, *24* (9), 1101-1119.

[44] Klinakis, A.; Karagiannis, D.; Rampias, T. Targeting DNA repair in cancer: current state and novel approaches. *Cell Mol. Life Sci.* **2020**, *77* (4), 677-703.

[45] Visnes, T.; Grube, M.; Hanna, B. M. F.; Benitez-Buelga, C.; Cázares-Körner, A.; Helleday, T. Targeting BER enzymes in cancer therapy. *DNA Repair* **2018**, *71*, 118-126.

[46] Singh, N.; Pay, S. L.; Bhandare, S. B.; Arimpur, U.; Motea, E. A. Therapeutic Strategies and Biomarkers to Modulate PARP Activity for Targeted Cancer Therapy. *Cancers* **2020**, *12* (4), 972.

[47] Przybycinski, J.; Nalewajska, M.; Marchelek-Mysliwiec, M.; Dziedziejko, V.; Pawlik, A. Poly-ADP-ribose polymerases (PARPs) as a therapeutic target in the treatment of selected cancers. *Expert Opin. Ther. Targets* **2019**, *23* (9), 773-785.

[48] Lecona, E.; Fernandez-Capetillo, O. Targetin ATR in cancer. *Nat. Rev. Cancer* **2018**, *18* (9), 586-595.

[49] Bradbury, A.; Hall, S.; Curtin, N.; Drew, Y. Targeting ATR as Cancer Therapy: A new era for synthetic lethality and synergistic combinations? *Pharmacol. Ther.* **2020**, *207*, 107450.

[50] Pilié, P. G.; Tang, C.; Mills, G. B.; Yap, T. A. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* **2019**, *16* (2), 81-104.

[51] Mohiuddin, I. S.; Kang, M. H. DNA-PK as an Emerging Therapeutic Target in Cancer. *Front. Oncol.* **2019**, *9*, 635.

[52] Forment, J. V.; O'Connor, M. J. Targeting the replication stress response in cancer. *Pharmacol. Ther.* **2018**, *188*, 155-167.

[53] Matheson, C. J.; Backos, D. S.; Reigan, P. Targeting WEE1 Kinase in Cancer. *Trends Pharmacol. Sci.* **2016**, *37* (10), 872-881.

[54] Murai, J.; Thomas, A.; Miettinen, M.; Pommier, Y. Schlafen 11 (SLFN11), a restriction factor for replicative stress induced by DNA-targeting anti-cancer therapies. *Pharmacol. Ther.* **2019**, *201*, 94-102.

[55] Tonzi, P.; Huang, T. T. Role of Y-family translesion DNA polymerases in replication stress: Implications for new cancer therapeutic targets. *DNA Repair* **2019**, *78*, 20-26.

[56] Otto, T.; Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **2017**, *17* (2), 93-115.

[57] Zhao, Y.; Wang, X. PLK4: a promising target for cancer therapy. *J. Cancer Res. Clin. Oncol.* **2019**, *145* (10), 2413-2422.

[58] Tang, A.; Gao, K.; Chu, L.; Zhang, R.; Yang, J.; Zheng, J. Aurora kinases: novel therapy targets in cancers. *Oncotarget* **2017**, *8* (14), 23937-23954.

[59] Liu, X.; Chen, Y.; Li, Y.; Petersen, R. B.; Huang, K. Targeting mitosis exit: A brake for cancer cell proliferation. *Biochim. Biophys. Acta Rev. Cancer* **2019**, *1871* (1), 179-191.

[60] Chen, M.; Xie, S. Therapeutic targeting of cellular stress responses in cancer. *Thorac. Cancer* **2018**, *9* (12), 1575-1582.

[61] Magaway, C.; Kim, E.; Jacinto, E. Targeting mTOR and Metabolism in Cancer: Lessons and Innovations. *Cells* **2019**, *8* (12), 1584.

[62] Tang, W.; Zhao, G. Small molecules targeting HIF-1α pathway for cancer therapy in recent years. *Bioorg. Med. Chem.* **2020**, *28* (2), 115235.

[63] Guang, M. H. Z.; Kavanagh, E. L.; Dunne, L. P.; Dowling, P.; Zhang, L.; Lindsay, S.; Bazou, D.; Goh, C. Y.; Hanley, C.; Bianchi, G.; Anderson, K. C.; O'Gorman, P.; McCann, A. Targeting Proteotoxic Stress in Cancer: A Review of the Role that Protein Quality Control Pathways Play in Oncogenesis. *Cancers* **2019**, *11* (1), 66.

[64] Garg, A. D.; Maes, H.; van Vilet, A. R.; Agostinis, P. Targeting the hallmarks of cancer with therapy-induced endoplasmic reticulum (ER) stress. *Mol. Cell. Oncol.* **2014**, *2* (1), e975089.

[65] Raab-Westphal, S.; Marshall, J. F.; Goodman, S. L. Integrins as Therapeutic Targets: Successes and Cancers. *Cancers* **2017**, *9* (9), 110.

[66] Carneiro, B. A.; El-Deiry, W. S. Targeting apoptosis in cancer therapy. *Nat. Rev. Clin. Oncol.* **2020**, *17* (7), 395-417.

[67] Song, M.; Bode, A. M.; Dong, Z.; Lee, M.-H. AKT as a Therapeutic Target for Cancer. *Cancer Res.* **2019**, *79* (6), 1019-1031.

[68] Roskoski Jr., R. Targeting ERK1/2 protein-serine/threonine kinases in human cancers. *Pharmacol. Res.* **2019**, *142*, 151-168.

[69] Degirmenci, U.; Wang, M.; Hu, J. Targeting Aberrant RAS/RAF/MEK/ERK Signaling for Cancer Therapy. *Cells* **2020**, *9* (1), 198.

[70] de la Cruz López, K. G.; Guzmán, M. E. T.; Sánchez, E. O.; Carrancá, A. G. mTORC1 as a Regulator of Mitochondrial Functions and a Therapeutic Target in Cancer. *Front. Oncol.* **2019**, *9*, 1373.

[71] Boice, A.; Bouchier-Hayes, L. Targeting apoptotic caspases in cancer. *Biochim. Biohys. Acta Mol. Cell Res.* **2020**, *1867* (6), 118688.

[72] Abbas, R.; Larisch, S. Targeting XIAP for Promoting Cancer Cell Death-The Story of ARTS and SMAC. *Cells* **2020**, *9* (3), 663.

[73] Lee, T. I.; Young, R. A. Transcriptional Regulation and its Misregulation in Disease. *Cell* **2013**, *152* (6), 1237-1251.

[74] Darnell Jr., J. E. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2* (10), 740-749.

[75] Bushweller, J. H. Targeting transcription factors in cancer – from undruggable to reality. *Nat. Rev. Cancer* **2019**, *19* (11), 611-624.

[76] Orlova, A.; Wagner, C.; de Araujo, E. D.; Bajusz, D.; Neubauer, H. A.; Herling, M.; Gunning, P. T.; Keserű, G. M.; Moriggl, R. Direct Targeting Options for STAT3 and STAT5 in Cancer. *Cancers* **2019**, *11* (12), 1930.

[77] Lambert, M.; Jambon, S.; Depauw, S.; David-Cordonnier, M.-H. Targeting Transcription Factors for Cancer Treatment. *Molecules* **2018**, *23* (6), 1479.

[78] Ghosh, N.; Hossain, U.; Mandal, A.; Sil, P. C. The Wnt signaling pathway: a potential therapeutic target against cancer. *Ann. N. Y. Acad. Sci.* **2019**, *1443* (1), 54-74.

[79] Zhang, X.; Wang, L.; Qu, Y. Targeting the β-catenin signaling for cancer therapy. *Pharmacol. Res.* **2020**, *160*, 104794.

[80] Smith Jr., T. M.; Tharakan, A.; Martin, R. K. Targeting ADAM10 in Cancer and Autoimmunity. *Front. Immunol.* **2020**, *11*, 499.

[81] Takebe, N.; Miele, L.; Harris, P. J.; Jeong, W.; Bando, H.; Kahn, M.; Yang, S. X.; Ivy, S. P. Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat. Rev. Clin. Oncol.* **2015**, *12* (8), 445-464.

[82] Bykov, V. J. N.; Eriksson, S. E.; Bianchi, J.; Wiman, K. G. Targeting mutant p53 for efficient cancer therapy. *Nat. Rev. Cancer* **2018**, *18* (2), 89-102.

[83] Liu, Y.; Wang, X.; Wang, G.; Yang, Y.; Yuan, Y.; Ouyang, L. The past, present and future of potential small-molecule drugs targeting p53-MDM2-MDMX for cancer therapy. *Eur. J. Med. Chem.* **2019**, *176*, 92-104.

[84] Yao, S.; Fan, L. Y.-N.; Lam, E. W.-F. The FOXO3-FOXM1 axis: A key cancer drug target and a modulator of cancer drug resistance. *Semin. Cancer Biol.* **2018**, *50*, 77-89.

[85] Chen, H.; Liu, H.; Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct. Target. Ther.* **2018**, *3*, 5.

[86] Allen-Petersen, B. L.; Sears, R. C. Mission Possible: Advances in MYC Therapeutic Targeting in Cancer. *BioDrugs* **2019**, *33* (5), 539-553.

[87] Baluapuri, A.; Wolf, E.; Eilers, M. Target gene-independent functions of MYC oncoproteins. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (5), 255-267.

[88] Carrà, G.; Lingua, M. F.; Maffeo, B.; Taulli, R.; Morotti, A. P53 vs NF-kB: the role of nuclear factor-kapa B in the regulation of p53 activity and vice versa. *Cell. Mol. Life Sci.* **2020**, *77* (22), 4449-4458.

[89] Pfister, S. X.; Ashworth, A. Marked for death: targeting epigenetic changes in cancer. *Nat. Rev. Drug Discov.* **2017**, *16* (4), 241-263.

[90] Park, J. W.; Han, J.-W. Targeting epigenetics for cancer therapy. *Arch. Pharm. Res.* **2019**, *42*, 159-170.

[91] Bhat, M.; Robichaud, N.; Hulea, L.; Sonenberg, N.; Pelletier, J.; Topisirovic, I. Targeting the translation machinery in cancer. *Nat. Rev. Drug Discov.* **2015**, *14* (4), 261-278.

[92] Luengo, A.; Gui, D. Y.; Vander Heiden, M. G. Targeting Metabolism for Cancer Therapy. *Cell Chem. Biol.* **2017**, *24* (9), 1161-1180.

[93] Carvalho, T. M. A.; Cardoso, H. J.; Figueira, M. I.; Vaz, C. V.; Socorro, S. The peculiarities of cancer cell metabolism: A route to metastasization and a target for therapy. *Eur. J. Med. Chem.* **2019**, *171*, 343-363.

[94] Deepak, K. G. K.; Vempati, R.; Nagaraju, G. P.; Dasari, V. R.; Nagini, S.; Rao, D. N.; Malla, R. R. Tumor microenvironment: Challenges and opportunities in targeting metastasis of triple negative breast cancer. *Pharmacol. Res.* **2020**, *153*, 104683.

[95] Bian, Y.; Wei, J.; Zhao, C.; Li, G. Natural Polyhenols Targeting Senescence: A Novel Prevention and Therapy Strategy for Cancer. *Int. J. Mol. Sci.* **2020**, *21* (2), 684.

[96] Martinez-Montiel, N.; Rosas-Murrieta, N. H.; Ruiz, M. A.; Monjaraz-Guzman, E.; Martinez-Contreras, R. Alternative Splicing as a Target for Cancer Treatment. *Int. J. Mol. Sci.* **2018**, *19* (2), 545.

[97] Lee, C.; Do, H. T. T.; Her, J.; Kim, Y.; Seo, D.; Rhee, I. Inflammasome as a promising therapeutic target for cancer. *Life Sci.* **2019**, *231*, 116593.

[98] Carotenuto, P.; Pecoraro, A.; Palma, G.; Russo, G.; Russo, A. Therapeutic Approaches Targeting Nucleolus in Cancer. *Cells* **2019**, *8* (9), 1090.

[99] Saravanan, S.; Vimalraj, S.; Pavani, K.; Nikarika, R.; Sumantran, V. N. Intussusceptive angiogenesis as a key therapeutic target for cancer therapy. *Life Sci.* **2020**, *252*, 117670.

[100] Huang, A.; Garraway, L. A.; Ashworth, A.; Weber, B. Synthetic lethality as an engine for cancer drug target discovery. *Nat. Rev. Drug Discov.* **2020**, *19* (1), 23-38.

[101] Wang, Z.; Cole, P. A. Catalytic Mechanisms and Regulation of Protein Kinases. *Methods Enzymol.* **2014**, *548*, 1-21.

[102] Roskoski Jr., R. A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacol. Res.* **2015**, *100*, 1-23.

[103] Beenstock, J.; Mooshayef, N.; Engelberg, D. How Do Protein Kinases Take a Selfie (Autophosphorylate)? *Trends Biochem. Sci.* **2016**, *41* (11), 938-953.

[104] Nasa, I.; Kettenbach, A. N. Coordination of Protein Kinase and Phosphoprotein Phosphatase Activities in Mitosis. *Front. Cell Dev. Biol.* **2018**, *6*, 30.

[105] Lanz, M. C.; Dibitetto, D.; Smolka, M. B. DNA damage kinase signaling: checkpoint and repair at 30 years. *EMBO J.* **2019**, *38* (18), e101801.

[106] Hohmann, U.; Lau, K.; Hothorn, M. The Structural Basis of Ligand Perception and Signal Activation by Receptor Kinases. *Annu. Rev. Plant Biol.* **2017**, *68*, 109-137.

[107] Sreelatha, A.; Kinch, L. N.; Tagliabracci, V. S. The Secretory Pathway Kinases. *Biochim. Biophys. Acta* **2015**, *1854* (10 Pt B), 1687-1693.

[108] Manieri, E.; Sabio, G. Stress kinases in the modulation of metabolism and energy balance. *J. Mol. Endocrinol.* **2015**, *55* (2), R11-R22.

[109] Li, Z.; Zhang, X. Kinases Involved in Both Autophagy and Mitosis. *Int. J. Mol. Sci.* **2017**, *18* (9), 1884.

[110] Akbari-Birgani, S.; Khademy, M.; Mohseni-Dargah, M.; Madadi, Z.; Mokhtar-Ahmadabadi, R.; Davoodi-Monfared, P. Caspases interplay with kinases and phosphatases to determine cell fate. *Eur. J. Pharmacol.* **2019**, *855*, 20-29.

[111] Shahin, R.; Shaheen, O.; El-Dahiyat, F.; Habash, M.; Saffour, S. Research advances in kinase enzymes and inhibitors for cardiovascular disease treatment. *Future Sci. OA.* **2017**, *3* (4), FSO204.

[112] Weiss, A.; Boehm, M.; Egemnazarov, B.; Grimminger, F.; Pullamsetti, S. S.; Kwapiszewska, G.; Schermuly, R. T. Kinases as potential targets for treatment of pulmonary hypertension and right ventricular dysfunction. *Br. J. Pharmacol.* **2020**, *178* (1), 31-53.

[113] Mehdi, S. J.; Rosas-Hernandez, H.; Cuevas, E.; Lantz, S. M.; Barger, S. W.; Sarkar, S.; Paule, M. G.; Ali, S. F.; Imam, S. Z. Protein Kinases and Parkinson's Disease. *Int. J. Mol. Sci.* **2016**, *17* (9), 1585.

[114] Baltussen, L. L.; Rosianu, F.; Ultanir, S. K. Kinases in synaptic development and neurological diseases. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2018**, *84* (Pt B), 343-352.

[115] Fleuren, E. D. G.; Zhang, L.; Wu, J.; Daly, R. J. The kinome 'at large' in cancer. *Nat. Rev. Cancer* **2016**, *16* (2), 83-98.

[116] Kanev, G. K.; de Graaf, C.; de Esch, I. J. P.; Leurs, R.; Würdinger, T.; Westerman, B. a.; Kooistra, A. J. The Landscape of Atypical and Eukaryotic Protein Kinases. *Trends Pharmacol. Sci.* **2019**, *40* (11), 818-832.

[117] Tong, M.; Seeliger, M. A. Targeting Conformational Plasticity of Protein Kinases. *ACS Chem. Biol.* **2015**, *10* (1), 190-200.

[118] Ferguson, F. M.; Gray, N. S. Kinase inhibitors: the road ahead. *Nat. Rev. Drug Discov.* **2018**, *17* (5), 353-377.

[119] Lu, X.; Smaill, J. B.; Ding, K. Medicinal Chemistry Strategies for the Development of Kinase Inhibitors Targeting Point Mutations. *J. Med. Chem.* **2020**, *63* (19), 10726-10741.

[120] Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9* (1), 28-39.

[121] Fabbro, D. 25 Years of Small Molecular Weight Kinase Inhibitors: Potentials and Limitations. *Mol. Pharmacol.* **2015**, *87* (5), 766-775.

[122] Lu, X.; Smaill, J. B.; Ding, K. New Promise and Opportunities for Allosteric Kinase Inhibitors. *Angew. Chem. Int. Ed. Engl.* **2019**, *59* (33), 13764-13776.

[123] Roskoski Jr., R. Classification of small molecule protein kinase inhibitors based upon the structures of their drug enzyme complexes. *Pharmacol. Res.* **2016**, *103*, 26-48.

[124] Jacobsen, A. V.; Murphy, J. M. The secret life of kinases: insights into non-catalytic signaling functions from pseudokinases. *Biochem. Soc. Trans.* **2017**, *45* (3), 665-681.

[125] Zhao, Z.; Bourne, P. E. Progress with covalent small-molecule kinase inhibitors. *Drug Discov. Today* **2018**, *23* (3), 727-735.

[126] Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **2013**, *20* (2), 146-159.

[127] Abdeldayem, A.; Raouf, Y. S.; Constantinescu, S. N.; Moriggl, R.; Gunning, P. T. Advances in covalent kinase inhibitors. *Chem. Soc. Rev.* **2020**, *49* (9), 2617-2687.

[128] Munoz, L. Non-kinase targets of protein kinase inhibitors. *Nat. Rev. Drug Discov.* **2017**, *16* (6), 424-440.

[129] Leonard, T. A.; Hurley, J. H. Regulation of Protein Kinases by Lipids. *Curr. Opin. Struct. Biol.* **2011**, *21* (6), 785-791.

[130] Sasaki, T.; Takasuga, S.; Sasaki, J.; Kofuji, S.; Eguchi, S.; Yamazaki, M.; Suzuki, A. Mammalian phosphoinositide kinases and phosphatases. *Prog. Lipid Res.* **2009**, *48* (6), 307-343.

[131] Michell, R. H. Inositol derivatives: evolution and functions. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (2), 151-161.

[132] Michell, R. H. Inositol lipids in cellular signalling mechanisms. *Trends Biochem. Sci.* **1992**, *17* (8), 274-276.

[133] Irvine, R. F. A short history of inositol lipids. *J. Lipid Res.* **2016**, *57* (11), 1987-1994.

[134] Balla, T. Inositol-lipid binding motifs: signal integrators through protein-lipid and protein-protein interactions. *J. Cell Sci.* **2005**, *118* (Pt 10), 2093-2104.

[135] Hammond, G. R. V.; Balla, T. Polyphosphoinositide binding domains: key to inositol lipid biology. *Biochim. Biophys. Acta* **2015**, *1851* (6), 746-758.

[136] Billcliff, P. G.; Lowe, M. Inositol lipid phosphatases in membrane trafficking and human disease. *Biochem. J.* **2014**, *461* (2), 159-175.

[137] Burke, J. E. Structural Basis for Regulation of Phosphoinositide Kinases and Their Involvement in Human Disease. *Mol. Cell* **2018**, *71* (5), 653-673.

[138] Heck, J. N.; Mellman, D. L.; Ling, K. Sun, Y.; Wagoner, M. P.; Schill, N. J.; Anderson, R. A. A Conspicuous Connection: Structure Defines Function for the Phosphatidylinositol-Phosphate Kinase Family. *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42* (1), 15-39.

[139] Hinchliffe, K. A.; Ciruela, A.; Morris, J. A.; Divecha, N.; Irvine, R. F. The type II PIPkins (PtdIns5*P* 4-kinases): enzymes in search of a function? *Biochem. Soc. Trans.* **1999**, *27* (4), 657-661.

[140] Rameh, L. E.; Tolias, K. F.; Duckworth, B. C.; Cantley, C. L. A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate. *Nature* **1997**, *390* (6656), 192-196.

[141] Rao, V. D.; Misra, S.; Boronenkov, I. V.; Anderson, R. A.; Hurley, J. H. Structure of Type IIβ Phosphatidylinositol Phosphate Kinase: A Protein Kinase Fold Flattened for Interfacial Phosphorylation. *Cell* **1998**, *94* (6), 829-839.

[142] Bultsma, Y.; Keune, W.-J.; Divecha, N. PIP4Kβ interacts with and modulates nuclear localization of the high-activity PtdIns5*P*-4-kinase isoform PIP4Kα. *Biochem. J.* **2010**, *430* (2), 223-235.

[143] Clarke, J. H.; Irvine, R. F. Evolutionarily conserved structural changes in phosphatidylinositol 5-phosphate 4-kinase (PI5P4K) isoforms are responsible for differences in enzyme activity and localization. *Biochem. J.* **2013**, *454* (1), 49-57.

[144] Wang, M.; Bond, N. J.; Letcher, A. J.; Richardson, J. P.; Lilley, K. S.; Irvine, R. F.; Clarke, J. H. Genomic tagging reveals a random association of endogenous PtdIns5*P* 4-kinases IIα and IIβ and a partial nuclear localization of the IIα isoform. *Biochem. J.* **2010**, *430* (2), 215-221.

[145] Clarke, J. H.; Emson, P. C.; Irvine, R. F. Localization of phosphatidylinositol phosphate kinase IIγ in kidney to a membrane trafficking compartment within specialized cells of the nephron. *Am. J. Physiol. Renal Physiol.* **2008**, *295* (5), F1422-F1430.

[146] Bulley, S. J.; Clarke, J. H.; Droubi, A.; Giudici, M.-L.; Irvine, R. F. Exploring phosphatidylinositol 5-phosphate 4-kinase function. *Adv. Biol. Regul.* **2015**, *57*, 193-202.

[147] Ciruela, A.; Hinchliffe, K. A.; Divecha, N.; Irvine, R. F. Nuclear targeting of the β isoform of Type II phosphatidylinositol phosphate kinase (phosphatidylinositol 5-phosphate 4-kinase) by its α-helix 7. *Biochem. J.* **2000**, *346* (Pt 3), 587-591.

[148] Sumita, K.; Lo, Y.-H.; Takeuchi, K.; Senda, M.; Kofuji, S.; Ikeda, Y.; Terakawa, J.; Sasaki, M.; Yoshino, H.; Majd, N.; Zheng, Y.; Kahoud, E. R.; Yokota, T.; Emerling, B. M.; Asara, J. M.; Ishida, T.; Locasale, J. W.; Daikoku, T.; Anastasiou, D.; Senda, T.; Sasaki, A. T. The lipid kinase PI5P4Kβ is an intracellular GTP sensor for metabolism and tumorigenesis. *Mol. Cell* **2016**, *61* (2), 187-198.

[149] Wang, D. G.; Paddock, M. N.; Lundquist, M. R.; Sun, J. Y.; Mashadova, O.; Amadiume, S.; Bumpus, T. W.; Hodakoski, C.; Hopkins, B. D.; Fine, M.; Hill, A.; Yang, T. J.; Baskin, J. M.; Dow, L. E.; Cantley, L. C. PIP4Ks Suppress Insulin Signaling through a Catalytic-Independent Mechanism. *Cell Rep.* **2019**, *27* (7), 1991-2001.e5.

[150] Huang, Z.; Anderson, R. E.; Cao, W.; Wiechmann, A. F.; Rajala, R. V. S. Light-Induced Tyrosine Phosphorylation of Rod Outer Segment Membrane Proteins Regulate the Translocation, Membrane Binding and Activation of Type II α phosphatidylinositol-5-phosphate 4-kinase. *Neurochem. Res.* **2011**, *36* (4), 627-635.

[151] Elouarrat, D.; van der Velden, Y. U.; Jones, D. R.; Moolenaar, W. H.; Divecha, N.; Haramis, A.-P. G. Role of phosphatidylinositol 5-phosphate 4-kinase α in zebrafish development. *Int. J. Biochem. Cell Biol.* **2013**, *45* (7), 1293-1301.

[152] Hu, A.; Zhao, X.-T.; Tu, H.; Xiao, T.; Fu, T.; Wang, Y.; Liu, Y.; Shi, X.-J.; Luo, J.; Song, B.-L. PIP4K2A regulates intracellular cholesterol transport through modulating PI(4,5)P_2 homeostasis. *J. Lipid Res.* **2018**, *59* (3), 507-514.

[153] Lundquist, M. R.; Goncalves, M. D.; Loughran, R. M.; Possik, E.; Vijayaraghavan, T.; Yang, A.; Pauli, C.; Ravi, A.; Verma, A.; Yang, Z.; Johnson, J. L.; Wong, J. C. Y.; Ma, Y.; Hwang, K. S.-K.; Weinkove, D.; Divecha, N.; Asara, J. M.; Elemento, O.; Rubin, M. A.; Kimmelman, A. C.; Pause, A.; Cantley, L. C.; Emerling, B. M. Phosphatidylinositol-5-phosphate 4-kinases regulate cellular lipid metabolism by facilitating autophagy. *Mol. Cell* **2018**, *70* (3), 531-544.

[154] Lamia, K. A.; Peroni, O. D.; Kim, Y.-B.; Rameh, L. E.; Kahn, B. B.; Cantley, L. C. Increased Insulin Sensitivity and Reduced Adiposity in Phosphatidylinositol 5-Phosphate 4-Kinase $\beta^{-/-}$ Mice. *Mol. Cell Biol.* **2004**, *24* (11), 5080-5087.

[155] Carricaburu, V.; Lamia, K. A.; Lo, E.; Favereaux, L.; Payrastre, B.; Cantley, L. C.; Rameh, L. E. The phosphatidylinositol (PI)-5-phosphate 4-kinase type II enzyme controls insulin signaling by regulating PI-3,4,5-trisphosphate degradation. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (17), 9867-9872.

[156] Sharma, S.; Mathre, S.; Ramya, V.; Shinde, D.; Raghu, P. Phosphatidylinositol 5 Phosphate 4-Kinase Regulates Plasma-Membrane PIP_3 Turnover and Insulin Signaling. *Cell Rep.* **2019**, *27* (7), 1979-1990.e7.

[157] Jones, D. R.; Bultsma, Y.; Keune, W.-J.; Halstead, J. R.; Elouarrat, D.; Mohammed, S.; Heck, A. J.; D'Santos, C. S.; Divecha, N. Nuclear PtdIns5P as a Transducer of Stress Signaling: An In Vivo Role for PIP4Kbeta. *Mol. Cell* **2006**, *23* (5), 685-695.

[158] Sano, S.; Okawa, A.; Nakajima, A.; Tahara, M.; Fujita, K.; Wada, Y.; Yamazaki, M.; Moriya, H.; Sasho, T. Identification of *Pip4k2β* as a mechanical stimulus responsive gene and its expression during musculoskeletal tissue healing. *Cell Tissue Res.* **2006**, *323*, 245-252.

[159] Kouchi, Z.; Fujiwara, Y.; Yamaguchi, H.; Nakamura, Y.; Fukami, K. Phosphatidylinositol 5-phosphate 4-kinase type II beta is required for vitamin D receptor-dependent E-cadherin expression in SW480 cells. *Biochem. Biophys. Res. Commun.* **2011**, *408* (4), 523-529.

[160] Clarke, J. H.; Emson, P. C.; Irvine, R. F. Localization of phosphatidylinositol phosphate kinase IIγ in kidney to a membrane trafficking compartment within specialized cells of the nephron. *Am. J. Physiol. Renal Physiol.* **2008**, *295* (5), F1422-F1430.

[161] Yih, L.-H.; Wu, Y.-C.; Hsu, N.-C.; Kuo, H.-H. Arsenic Trioxide Induces Abnormal Mitotic Spindles Through a PIP4KIIγ/Rho Pathway. *Toxicol. Sci.* **2012**, *128* (1), 115-125.

[162] Lin, T.-C.; Kuo, H.-H.; Wu. Y.-C.; Pan, T. S.; Yih, L.-H. Phosphatidylinositol-5-phosphate 4-kinase gamma accumulates at the spindle pole and prevents microtubule depolymerization. *Cell Div.* **2019**, *14*, 9.

[163] Mackey, A. M.; Sarkes, D. A.; Bettencourt, I.; Asara, J. M.; Rameh, L. E. PIP4kγ is a substrate for mTORC1 that maintains basal mTORC1 signaling during starvation. *Sci. Signal.* **2014**, *7* (350), ra104.

[164] Shim, H.; Wu. C.; Ramsamooj, S.; Bosch, K. N.; Chen, Z.; Emerling, B. M.; Yun, J.; Liu, H.; Choo-Wing, R.; Yang, Z.; Wulf, G. M.; Kuchroo, V. K.; Cantley, L. C. Deletion of the gene *Pip4k2c*, a novel phosphatidylinositol kinase, results in hyperactivation of the immune system. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (27), 7596-7601.

[165] Zheng, L.; Conner, S. D. PI5P4Kγ functions in DTX1-mediated Notch signaling. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (9), E1983-E1990.

[166] Thiselton, D. L.; Maher, B. S.; Webb, B. T.; Bigdeli, T. B.; O'Neill, F. A.; Walsh, D.; Kendler, K. S.; Riley, B. P. Association analysis of the *PIP4K2A* gene on chromosome 10p12 and schizophrenia in the Irish study of high density schizophrenia families (ISHDSF) and the Irish case-control study of schizophrenia (ICCSS). *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **2010**, *153B* (1), 323-331.

[167] Kaur, H.; Jajodia, A.; Grover, S.; Baghel, R.; Gupta, M.; Jain, S.; Kukreti, R. Genetic Variations of *PIP4K2A* Confer Vulnerability to Poor Antipsychotic Response in Severely Ill Schizophrenia Patients. *PLoS One* **2014**, *9* (7), e102556.

[168] Al-Ramahi, I.; Giridharan, S. S. P.; Chen, Y.-C.; Patnaik, S.; Safren, N.; Hasegawa, J.; de Haro, M.; Gee, A. K. W.; Titus, S. A.; Jeong, H.; Clarke, J.; Krainc, D.; Zheng, W.; Irvine, R. F.; Barmada, S.; Ferrer, M.; Southall, N.; Weisman, L. S.; Botas, J.; Marugan, J. J. Inhibition of PIP4Kγ ameliorates the pathological effects of mutant huntingtin protein. *eLife* **2017**, *6*, e29123.

[169] Lima, K.; Ribeiro, D. M.; de Melo Campos, P.; Costa, F. F.; Traina, F.; Saad, S. T. O.; de Fátima Sonati, M.; Machado-Neto, J. A. Differential profile of PIP4K2A expression in hematological malignancies. *Blood Cells Mol. Dis.* **2015**, *55* (3), 228-235.

[170] Shin, Y. J.; Sa, J. K.; Lee, Y.; Kim, D.; Chang, N.; Cho, H. J.; Son, M.; Oh, M. Y. T.; Shin, K.; Lee, J.-K.; Park, J.; Jo, Y. K.; Kim, M.; Paddison, P. J.; Tergaonkar, V.; Lee, J.; Nam, D.-H. PIP4K2A as a negative regulator of PI3K in PTEN-deficient glioblastoma. *J. Exp. Med.* **2019**, *216* (5), 1120-1134.

[171] Jude, J. G.; Spencer, G. J.; Huang, X.; Somerville, T. D. D.; Jones, D. R.; Divecha, N.; Somervaille, T. C. P. A targeted knockdown screen of genes coding for phosphoinositide modulators identifies *PIP4K2A* as required for acute myeloid leukemia cell proliferation and survival. *Oncogene* **2015**, *34* (10), 1253-1262.

[172] Liao, F.; Yin, D.; Zhang, Y.; Hou, Q.; Zheng, Z.; Yang, L.; Shu, Y.; Xu, H.; Li, Y. Association Between *PIP4K2A* Polymorphisms and Acute Lymphoblastic Leukemia Susceptibility. *Medicine* **2016**, *95* (18), e3542.

[173] Lima, K.; Coelho-Silva, J. L.; Kinker, G. S.; Pereira-Martins, D. A.; Traina, F.; Fernandes, P. A. C. M.; Markus, R. P.; Lucena-Araujo, A. R.; Machado-Neto, J. A. *PIP4K2A* and *PIP4K2C* transcript levels are associated with cytogenetic risk and survival outcomes in acute myeloid leukemia. *Cancer Genet.* **2019**, *233-234*, 56-66.

[174] Zhang, S.; Li, Z.; Yan, X.; Bao, L.; Deng, Y.; Zeng, F.; Wang, P.; Zhu, J.; Yin, D.; Liao, F.; Zhou, X.; Zhang, D.; Xia, X.; Wang, H.; Yang, X.; Zhang, W.; Gao, H.; Zhang, W.; Yang, L.; Hou, Q.; Xu, H.; Zhang, Y.; Shu, Y.; Wang, Y. Regulatory Network and Prognostic Effect Investigation of PIP4K2A in Leukemia and Solid Cancers. *Front. Genet.* **2019**, *9*, 721.

[175] Vindu, A.; Dandewad, V.; Seshadri, V. Identification of human Phosphatidyl Inositol 5-Phosphate 4-Kinase as an RNA binding protein that is imported into *Plasmodium falciparum*. *Biochem. Biophys. Res. Commun.* **2018**, *498* (3), 529-536.

[176] Luoh, S.-W.; Venkatesan, N.; Tripathi, R. Overexpression of the amplified *Pip4k2β* gene from 17q11-12 in breast cancer cells confers proliferation advantage. *Oncogene* **2004**, *23* (7), 1354-1363.

[177] Fiume, R.; Jones, D. R.; Divecha, N. PIP4K2B: Coupling GTP Sensing to PtdIns5*P* Levels to Regulate Tumorigenesis. *Trends Biochem. Sci.* **2016**, *41* (6), 473-475.

[178] Keune, W.-J.; Sims, A. H.; Jones, D. R.; Bultsma, Y.; Lynch, J. T.; Jirström, K.; Landberg, G.; Divecha, N. Low PIP4K2B expression in human breast tumors correlates with reduced patient survival: a role of PIP4K2B in the regulation of E-cadherin expression. *Cancer Res.* **2013**, *73* (23), 6913-6925.

[179] Yan, J.; Jia, Y.; Chen, H.; Chen, W.; Zhou, X. Long non-coding RNA PXN-AS1 suppresses pancreatic cancer progression by acting as a competing endogenous RNA of miR-3064 to upregulate PIP4K2B expression. *J. Exp. Clin. Cancer Res.* **2019**, *38* (1), 390.

[180] Emerling, B. M.; Hurov, J. B.; Poulogiannis, G.; Tsukazawa, K. S.; Choo-Wing, R.; Wulf, G. M.; Bell, E. L.; Shim, H.-S.; Lamia, K. A.; Rameh, L. E.; Bellinger, G.; Sasaki, A. T.; Asara, J. M.; Yuan, X.; Bullock, A.; DeNicola, G. M.; Song, J.; Brown, V.; Signoretti, S.; Cantley, L. C. Depletion of a Putatively Druggable Class of Phosphatidylinositol Kinases Inhibits Growth of p53-Null Tumors. *Cell* **2013**, *155* (4), 844-857.

[181] Fiume, R.; Stijf-Bultsma, Y.; Shah, Z. H.; Keune, W. J.; Jones, D. R.; Jude, J. G.; Divecha, N. PIP4K and the role of nuclear phosphoinositides in tumour suppression. *Biochim. Biohys. Acta* **2015**, *1851* (6), 898-910.

[182] Davis, M. I.; Sasaki, A. T.; Shen, M.; Emerling, B. M.; Thorne, N.; Michael, S.; Pragani, R.; Boxer, M.; Sumita, K.; Takeuchi, K.; Auld, D. S.; Li, Z.; Cantley, L. C.; Simeonov, A. A Homogeneous, High-Throughput Assay for Phosphatidylinositol 5-Phosphate 4-Kinase with a Novel, Rapid Substrate Preparation. *PLoS One* **2013**, *8* (1), e54127.

[183] Giudici, M.-L.; Clarke, J. H.; Irvine, R. F. Phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ), a lipid signaling enigma. *Adv. Biol. Regul.* **2016**, *61*, 47-50.

[184] Voss, M. D.; Czechtizky, W.; Li, Z.; Rudolph, C.; Petry, S.; Brummerhop, H.; Langer, T.; Schiffer, A.; Schaefer, H.-L. Discovery and pharmacological chararacterization of a novel small molecule inhibitor of phosphatidylinositol-5-phosphate 4-kinase, type II, beta. *Biochem. Biophys. Res. Commun.* **2014**, *449* (3), 327-331.

[185] Clarke, J. H.; Giudici, M.-L.; Burke, J. E.; Williams, R. L.; Maloney, D. J.; Marugan, J.; Irvine, R. F. The function of phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ) explored using a specific inhibitor that targets the PI5P-binding site. *Biochem. J.* **2015**, *466* (2), 359-367.

[186] Kitagawa, M.; Liao, P.-J.; Lee, K. H.; Wong, J.; Shang, S. C.; Minami, N.; Sampetrean, O.; Saya, H.; Lingyun, D.; Prabhu, N.; Diam, G. K.; Sobota, R.; Larsson, A.; Nordlund, P.; McCormick, F.; Ghosh, S.; Epstein, D. M.; Dymock, B. W.; Lee, S. H. Dual blockade of the lipid kinase PIP4Ks and mitotic pathways leads to cancer-selective lethality. *Nat. Commun.* **2017**, *8* (1), 2200.

6 Desthiobiotin Labeling of a Covalent Kinase Inhibitor for the Study of CDK12 and CDK13

6.1 Introduction

6.1.1 The Cyclin-Dependent Kinase Family – An Overview

Cyclin-dependend kinases (CDKs) are serine/threonine-kinases, which were originally found in the 1980s to require cyclins as activating subunits[1,2]. Cyclins on the other hand are named after the cell cycle-dependent oscillations of their activity[1]. While CDK/cyclin complex activity was first found to be an essential driver of cell cycle progression, additional roles for both CDKs and cyclins have been identified since, including CDKs which do not require cyclin activation as well as cyclins exhibiting stable cellular levels throughout the cell cycle[1].

Orchestrating Cell Cycle Progression and Transcription are Main Functions of CDKs
Generally, the CDK family is divided into two main groups based on functional characteristics[2] (Figure 6.1.1). CDK1-6 and CDK14-18 are able to bind a variety of different cyclins, which facilitate the regulation of cell cycle progression initiated by various different signaling pathways[2]. In contrast, CDK7-13 and CDK19/20 selectively bind specific cyclins and are involved in protein transcription[2]. Comparing respective kinase analogs in yeast and humans (Figure 6.1.1) further emphasizes a major evolutionary diversification, especially within the cell cycle-regulating CDKs[2]. Within the two main CDK groups, CDKs are further categorized into subgroups, which are based on their conserved structural features[2] (Figure 6.1.2).

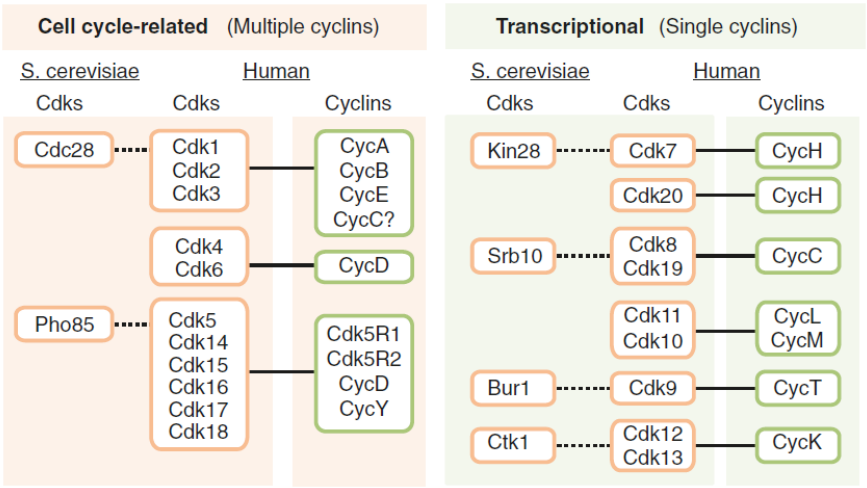


Figure 6.1.1 | Cell Cycle-regulating and Transcriptional CDKs. Overview of CDKs involved in either cell cycle-regulating functions or transcription with subfamilies being indicated by CDKs in the same, orange-framed boxes. Functionally related enzymes in yeast and humans are shown for comparison and highlight the evolutionary development within enzymes of this kinase family. Furthermore, cyclins, which are bound by the CDKs, are shown in green-framed boxes. Taken from Malumbres *et al*., *Genome Biol.* **2014**, *15* (6), 122[2].

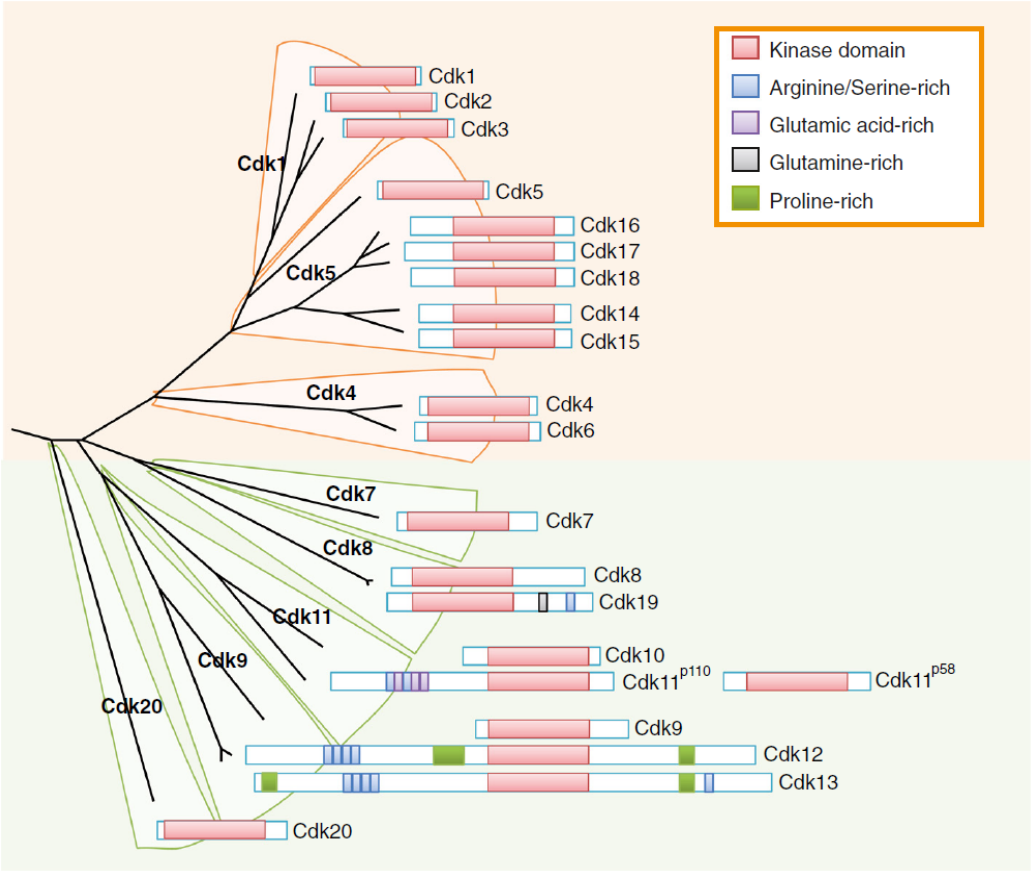


Figure 6.1.2 | Overview of CDK Subfamilies. The evolutionary development of the human CDKs is shown with an emphasis on structural features of the kinases. Cell cycle-regulating CDKs are located in the orange space, whereas transcriptional CDKs are found in the green space. Taken from Malumbres *et al*., *Genome Biol.* **2014**, *15* (6), 122[2].

CDKs exhibit typical structural features of kinases (see Chapter 5) including a certain flexibility within the catalytic site to take on active or inactive kinase conformations[3] (Figure 6.1.3). For example, it was shown that CDK2's active kinase conformation is induced by binding cyclin A[3]. Due to similar structural features among many CDKs, it is thus believed that this mechanism of CDK activation *via* allosteric effects of cyclin-binding is also applicable to other CDKs[3]. Notably, phosphorylation of specific sites in the glycine-rich loop of CDKs is a common mechanism found to cause CDK inactivation[2].

In accordance with cyclin activity throughout the cell cycle, CDKs are main regulators of important cell cycle events[1]. Briefly, before a human cell can enter mitosis to divide into two daughter cells (also referred to as M-phase), the cell has to go through a growth phase in which its genome is replicated (also referred to as S-phase)[1]. In between these two phases of cell growth and division, cells enter gap phases (G1/2-phases)[4] (Figure 6.1.4 A). Consequently, the G1/S transition, as well as the G2/M transition, come with crucial changes to a cell's mode and function and need to be meticulously regulated to ensure optimal cell growth[1].

CDK2, 4, and 6 are essential in regulating G1/S-phase progression[4]. CDK4 and CDK6 are both activated by cyclin D, which is itself activated as the result of several mitotic stimuli[4]. Among other targets, activated CDK4/6 complexes are able to phosphorylate retinoblastoma (Rb) proteins, which normally inhibit DNA replication in their

unphosphorylated state[4]. Another important function of CDK4 and CDK6 is the phosphorylation and subsequent inactivation of CDK-inhibiting proteins, which, amongst other processes, results in CDK2 activation[1]. CDK2/cyclin E complexes act complementary to CDK4 and CDK6 and further inactivate inhibitors of CDKs and DNA replication[1]. After initiation of the S-phase, CDK2 associates mainly with cyclin A (Figure 6.1.4 B), which shifts its activity towards driving DNA replication, including phosphorylation of transcription factors, DNA repair mechanism proteins, histone modifiers, and cell cycle checkpoint factors, such as p53 and MDM2[1,4].

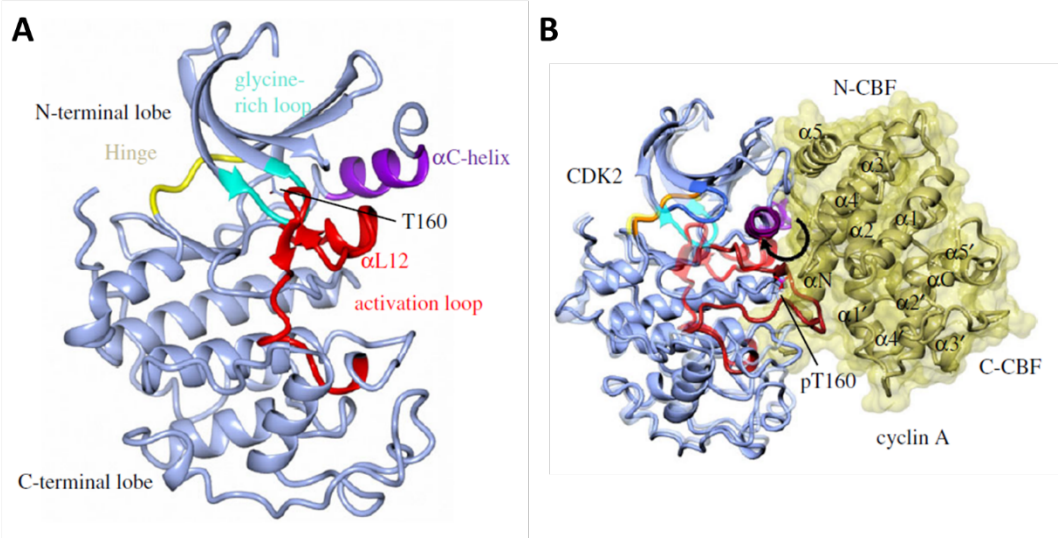


Figure 6.1.3 | Structural Features of CDKs[a]. A. Typical structural features of kinases are highlighted within the crystal structure of CDK2 as a representative example for CDKs. **B.** An overlay of unbound CDK2 and CDK2 in complex with cyclin A is given. The movement of the αC-helix (shown in purple) is highlighted and determines the active or inactive state of CDK2.
[a] Adapted from Wood *et al.*, *Open Biol.* **2018**, *8* (9), 180112[3].

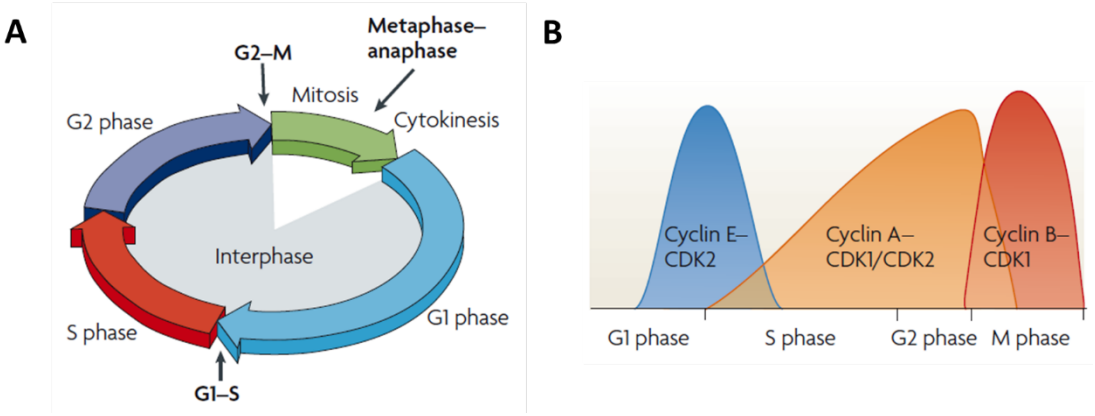


Figure 6.1.4 | Cyclin Activity Within the Cell Cycle[a]. A. Graphical overview of the different phases of the human cell cycle. **B.** The fluctuation of activity of selected cyclins throughout the cell cycle is shown, highlighting how cyclin activity dictates CDK function.
[a] Adapted from Hochegger *et al.*, *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (11), 910-916[5].

In cells experiencing DNA damage, CDK2 activity is generally inhibited to prevent cell cycle progression[6]. Eventually, cyclin A associates with CDK1 (Figure 6.1.4 B), which is activated by the CDK2/cyclin A complex, indicating the end of the S-phase[1,4]. After termination of the S-phase, cyclin B takes the place of cyclin A and associates with CDK1, which has several targets and promotes the G2/M-phase transition[4]. The resulting complex is initially inhibited by the Wee1 kinase but later activated by the CDC25 phosphatase[1]. Subsequently active CDK1/cyclin B complexes form a positive feedback loop in which they phosphorylate Wee1 and CDC25, leading to the inhibition of their inhibiting kinase and activation of their activating phosphatase, respectively, at which point the cell has committed to mitosis[1]. CDK1/cyclin B complex activity further mediates several essential steps of mitosis such as nuclear envelope breakdown, spindle assembly, and chromosomal condensation[1,4]. Notably, CDK1 is the only essential CDK isoform in human cells[2]. Finally, mitotic exit of a cell is facilitated *via* cyclin B degradation[4].

Another important function of CDKs is the regulation of transcription, especially *via* CDK-mediated RNAPII C-terminal domain (CTD) phosphorylation[7]. Similar to the cell cycle-regulating CDKs, transcription is facilitated by the careful interplay of different CDKs to promote various steps during transcription[7] (Figure 6.1.5). For instance, CDK8 and CDK19 form a complex with cyclin C and can inhibit or initiate transcription *via* several mechanisms[7]. CDK8 or CDK19-mediated phosphorylation of the CTD interferes with RNAPII function, while phosphorylation of certain transcription factors can also lead to transcription inhibition[7]. In other cases, CDK8 and CDK19 activate transcription factors through phosphorylation[7]. Furthermore, these CDK complexes have been shown to support RNAPII initiation, phosphorylate histones to support transcription activation, and enable the recruitment of certain transcription factors to trigger transcription elongation[7].

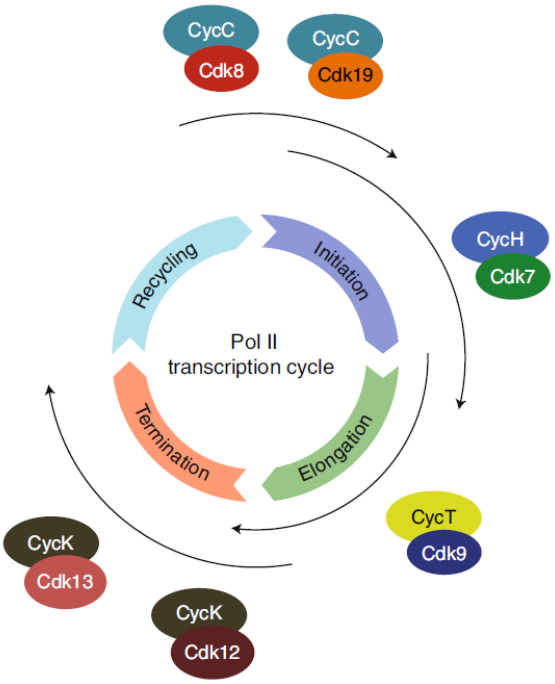


Figure 6.1.5 | CDKs Are Master Regulators of Transcription. Different stages of the transcription cycle are mediated by specific CDK/cyclin complexes. Taken from Parua *et al*., *Nat. Chem. Biol.* **2020**, *16* (7), 716-724[7].

CDK7 teams up with cyclin H and does not only phosphorylate the RNAPII CTD and other factors to promote gene transcription but also plays an important role in phosphorylating and activating cell cycle-regulating CDKs[7]. Interestingly, CDK9 is activated by complex formation with cyclin T and CDK7-mediated phosphorylation[7]. The CDK9/cyclin T complex is involved in transcription elongation processes and checkpoints leading into transcription termination[7]. Furthermore, CDK12 and CDK13 form complexes with cyclin K and have been implicated in transcription elongation and proper termination[7]. Specifically, both CDK12 and CDK13 have been shown to phosphorylate specific motifs within the RNAPII CTD and promote the transcription of a specific pool of genes, including proteins of the DDR[7]. In addition, CDK11/cyclin L complexes have been linked to regulate transcription and RNA splicing events, while also playing a role in receptor signaling pathways and autophagy[2].

Finally, the other known CDKs exhibit rather unusual CDK activity. For instance, CDK5 is a membrane-associated CDK and interacts with non-cyclin proteins[2]. CDK5 has been shown to be involved in several cellular mechanisms including transcription, differentiation, angiogenesis, and senescence[2]. While CDK14-18, in contrast, form complexes with a cyclin (cyclin Y), they were also found to be located at the plasma membrane, where they were shown to be associated with many signaling pathways; however, these functions are still poorly understood[2]. Similarly, CDK10 and CDK20 are rather understudied CDKs and have so far been linked to transcription events and cell cycle progression, respectively[2].

Given the essential role of CDKs in cell cycle progression and transcription, CDK activity has been thus shown to be involved in many cell functions in health and disease. For example, by promoting transcription of proinflammatory transcription factors, including NF-κB and STAT3, CDK activity was shown to affect inflammation[8]. Specifically, CDK2, 5, and 7 have been shown to promote the NF-κB pathway, while CDK9 was shown to be involved in many aspects of TNF-induced signaling cascades[8]. Surprisingly, the cell cycle-regulating CDKs, CDK4 and CDK6, were also associated with effects on proinflammatory gene expression[8]. That is, recent studies have shown that CDK6 is able to regulate transcription of its target proteins *via* kinase-independent mechanisms[9]. Consequently, CDK-mediated transcription regulation was shown to play an important role in infections, including virus-related diseases[10]. Finally, cell cycle-regulating CDKs are involved in the promotion of cell differentiation, especially in the context of immune cells and adaptive immunity[11], the development of pluripotent stem cells[12], ciliogenesis[13], as well as spermatogenesis[14].

Interestingly, CDK5 has been specifically linked to several developmental mechanisms in neurons[15] and has been thus shown to be involved in pathoneurological conditions such as Alzheimer's disease[16]. Another important role of CDKs has been found in the development of cancerous conditions as a result of their involvement in cell proliferation and transcription[17]. Consequently, many approaches have been developed to inhibit CDKs as potential targets for cancer therapy[18]. While first CDK inhibitors lacked in selectivity[18], some CDK family-selective CDK inhibitors have been approved for the treatment of cancer, e.g. CDK4/6 inhibitors for the treatment of hormone receptor-positive breast cancer[19].

6.1.2 CDK12 and CDK13 as Therapeutic Targets

In the early 2000s, CDK12 and CDK13 (back then still referred to as CRKRS and CDC2L5, respectively) were discovered as regulators of the cell cycle[20,21]. However, functional studies showed that CDK12 and CDK13, in contrast to cell cycle-regulating CDKs, act as transcriptional orchestrators by targeting Ser2 at the RNAPII CTD[21]. CDK12 and CDK13 were further initially believed to interact with cyclin L, which turned out to be a false assumption due to biased experimental set-ups[20]. Instead, cyclin K could be identified as the interaction partner of these CDKs[20]. Similarly, cyclin K was first believed to form complexes with CDK9, but later it was demonstrated that CDK9 actually forms complexes with cyclin T[20].

Structural and Functional Characteristics of CDK12 and CDK13

Co-crystal structures of CDK12 and CDK13 in complex with cyclin K have clarified the mechanism by which these CDKs regulate transcription[22-24] (Figure 6.1.6). As is common for transcriptional CDKs, CDK12 and CDK13 exhibit C-terminal extensions, which are associated with their ability to phosphorylate the CTD of RNAPII thus promoting transcriptional elongation[22,24]. Notably, the pattern formed by the electrostatic surface in this region is in perfect contrast to the negatively charged RNAPII CTD tail, allowing for optimal substrate binding as well as determining substrate specificity[22,24]. It was further shown that the flexibility of this extended loop is also responsible for the opening and closing of the ATP-binding pocket to ensure ATP binding and ADP release[23]. Interestingly, CDK12 and CDK13 show conserved loops forming additional cyclin interactions, which is in contrast to other transcriptional CDKs[23]. Cell cycle-regulating CDKs on the other hand show even stronger cyclin interactions, while also exhibiting a strong connection between the cyclin-binding site and their kinase active site[23]. This is commonly not the case for transcriptional CDKs, such as CDK9, which instead interact with other factors to regulate their kinase activity and is also not true for CDK12[23].

In 2011, Blazek and coworkers could link CDK12/cyclin K and CDK13/cyclin K activity with the transcription of genes involved in DDR, including BRCA1 and ATR[25]. Further studies have indicated that CDK12 activity regulates transcription, as well as splicing events and polyadenylation or cleavage sites[26], while CDK12 targets have been shown to be involved in cellular stress responses, such as nuclear factor erythroid 2-related factor 2 (Nrf2) target genes[27], cell proliferation and immune response *via* the NF-κB pathway[28], mitotic genome stability through promotion of mTORC1 activity[29], and G1/S-phase transition[30]. In contrast, little is reported about further targets of CDK13. Notably, some studies have indicated that inhibition of individual CDK12 or CDK13 activity resulted in changes in DDR gene transcription, while only dual inhibition of CDK12 and CDK13 also led to antiproliferative effects[31]. Interestingly, it was shown that CDK12 and CDK13 are involved in CDK5 expression and therefore neuronal development[32].

CDK12 and CDK13 have consequently been shown to be involved in several diseases, especially cancer, including breast[33] and ovarian cancers[34]. In the past couple of years, CDK12 activity has been studied in various other oncological conditions, e.g. thyroid[35], prostate[36], and hepatocellular cancer[37], Ewing sarcoma[38], and others. Notably, low CDK12 levels were associated with late stage cancers and poor clinical outcomes in gastric cancer[39]. However, CDK12 has also been shown to have tumorigenic effects[40], which is thought to be due to its effects on both tumor suppressors and oncogenes, which leads to

an imbalance in case of loss in CDK12 activity as well as in case of its overactivation or overexpression[41]. The potential role for CDK12 in cancer development could also vary throughout the cancer cycle, in which loss of CDK12 increases genomic instability and could promote tumorigenesis in early stages of a cancer cell, while an increase in CDK12 expression or gain-of-function mutations could lead to overexpression of oncogenes throughout the cancer cell development[42]. Consequently, several studies report CDK12 as an attractive target to treat cancers due to its effects on transcriptional and post-translational effects[43], especially regarding DDR genes[44].

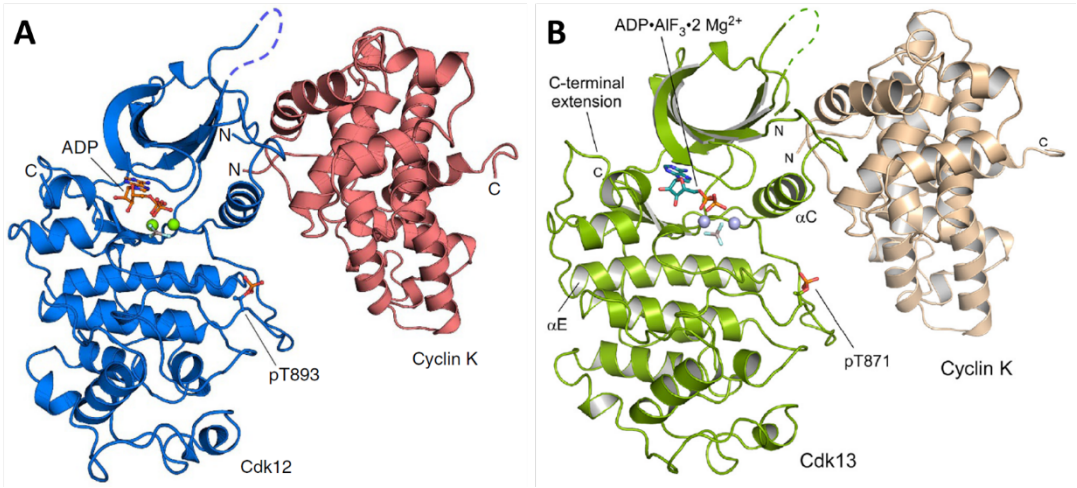


Figure 6.1.6 | CDK12 and CDK13 in Complex with Cyclin K. A. Shown is the co-crystal structure of CDK12 (blue) in complex with cyclin K (red). The kinase is in its active conformation with bound ADP and including the phosphorylated T-loop moiety T893. Adapted from Bösken *et al.*, *Nat. Commun.* **2014**, *5*, 3505[22]. **B.** Shown is the co-crystal structure of CDK13 (green) in complex with cyclin K (puder). The kinase is in its active conformation with bound ADP·AlF_3·$2Mg^{2+}$ and including the phosphorylated T-loop moiety T871. Adapted from Greifenberg *et al.*, *Cell Rep.* **2016**, *14* (2), 320-331[24].

While CDK12 and CDK13 are thus attractive targets for the treatment of diseases such as cancer, at the start of this project, potent and selective CDK12 and CDK13 inhibitors were still needed to further investigate their potential as targets in tumor therapy as well as to further study their modes of action e.g., their effects on disease biology.

6.2 Publication D: *Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors*

6.2.1 Specific Aims

Since potent and selective CDK12/13 inhibitors were still missing at the start of this project, the aim was to develop a CDK12/13-selective inhibitor from the in-house CDK7/12/13-targeting inhibitor THZ1. When joining these efforts, thorough medicinal chemistry efforts had yielded a CDK12/13-selective compound, THZ531. In order to further analyze this covalent CDK inhibitor, the goal was to develop and synthesize a biotin-labeled analog of the compound (bioTHZ531), which could be applied to investigate the cellular target engagement of the probe. In addition, first cellular characterizations of THZ531 target engagement were to be performed.

6.2.2 Main Outcomes and Context

In this study, the first CDK12/13-selective, covalent CDK inhibitor THZ531 is presented. Starting from the CDK7/12/13 inhibitor THZ1, rational design efforts were able to out-dial CDK7 binding of the compound, resulting in the CDK12/13-selective derivative THZ531. The covalent binding mode of the compound was demonstrated through co-crystallization with the CDK12/cyclin K complex. With the synthesis of a biotinylated THZ531 analog (bioTHZ531), it was possible to confirm covalent target engagement and selectivity towards CDK12/13 in the cellular context.

After thorough characterization of the biochemical target inhibition effects and their kinetics, THZ531 was shown to cause apoptosis in Jurkat cells. It could further be confirmed that this effect was dependent on the covalent binding of CDK12 and CDK13 by THZ531. Moreover, effects on transcription after THZ531 treatment were analyzed and could verify that CDK12/13 inhibition led to an increase in Ser2 phosphorylation in RNAPII CTD as well as an inhibition of transcription elongation. These effects negatively regulated several super enhancer-associated oncogenes and genes involved in the DDR.

Notably, while the biotin-labeled THZ531 probe served as an essential tool to validate THZ531 as a selective and potent CDK12/13 inhibitor in cells, the synthetic and experimental work was performed as part of my “Diplom” thesis[45] (degree: Dipl.-Pharm.; comparable to a Master’s degree), while the majority of the analysis of the developed probe, including the publication of the work, were still ongoing at my graduation and finalized during the beginnings of my time as a Ph.D. student. Experimental descriptions of the synthetic route for the bioTHZ531 probe and cellular assays are given in the Methods section and the Supplementary Note.

6.2.3 Published Article

The following peer-reviewed publication was originally published as:

Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors

Tinghu Zhang, Nicholas Kwiatkowski, Calla M. Olson, Sarah E. Dixon-Clarke, Brian J. Abraham, Ann K. Greifenberg, Scott B. Ficarro, Jonathan M. Elkins, Yanke Liang, Nancy M. Hannett, Theresa Manz, Mingfeng Hao, Bartlomiej Bartkowiak, Arno L. Greenleaf, Jarrod A. Marto, Matthias Geyer, Alex N. Bullock, Richard A. Young, Nathanael S. Gray.

Nat. Chem. Biol. **2016**, *12*, 876-884.

DOI: https://doi.org/10.1038/nchembio.2166

This article, including Supporting Information, was reproduced with kind permission from the journal *Nature Chemical Biology* © 2016, Springer Nature.

ARTICLE
PUBLISHED ONLINE: 29 AUGUST 2016 | DOI: 10.1038/NCHEMBIO.2166
nature chemical biology

Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors

Tinghu Zhang[1,2,11], Nicholas Kwiatkowski[1–3,11], Calla M Olson[1,2,11], Sarah E Dixon-Clarke[4], Brian J Abraham[3], Ann K Greifenberg[5,6], Scott B Ficarro[1,2,7], Jonathan M Elkins[4], Yanke Liang[1,2], Nancy M Hannett[3], Theresa Manz[1,8], Mingfeng Hao[1,2], Bartlomiej Bartkowiak[9], Arno L Greenleaf[9], Jarrod A Marto[1,2,7], Matthias Geyer[5,6], Alex N Bullock[4], Richard A Young[3,10]* & Nathanael S Gray[1,2]*

Cyclin-dependent kinases 12 and 13 (CDK12 and CDK13) play critical roles in the regulation of gene transcription. However, the absence of CDK12 and CDK13 inhibitors has hindered the ability to investigate the consequences of their inhibition in healthy cells and cancer cells. Here we describe the rational design of a first-in-class CDK12 and CDK13 covalent inhibitor, THZ531. Co-crystallization of THZ531 with CDK12–cyclin K indicates that THZ531 irreversibly targets a cysteine located outside the kinase domain. THZ531 causes a loss of gene expression with concurrent loss of elongating and hyperphosphorylated RNA polymerase II. In particular, THZ531 substantially decreases the expression of DNA damage response genes and key super-enhancer-associated transcription factor genes. Coincident with transcriptional perturbation, THZ531 dramatically induced apoptotic cell death. Small molecules capable of specifically targeting CDK12 and CDK13 may thus help identify cancer subtypes that are particularly dependent on their kinase activities.

© 2016 Nature America, Inc. All rights reserved.

RNA polymerase II (Pol II) is responsible for the transcription of protein-coding genes in eukaryotic cells[1]. The process of transcription itself can be divided into discrete stages, including initiation, elongation, and termination. Pol II activity throughout the transcription cycle is controlled by coordinated, reversible post-translational modification of residues in the heptad (YSPTSPS) amino acid repeats found in its C-terminal domain (CTD)[2–4]. Phosphorylation of serine at position 5 (Ser5) of the CTD is required for proper transcriptional initiation from gene promoters, while phosphorylation of Ser2 promotes elongation of Pol II through the gene body and the production of mature mRNA transcript[5]. In mammalian cells, Ser2 phosphorylation has, until recently, been attributed solely to the activity of cyclin-dependent kinase 9 (CDK9), the kinase component of the positive transcription elongation factor b (P-TEFb)[6]. Research in both yeast and metazoans demonstrates that CDK12 and CDK13 may also play important roles in Ser2 phosphorylation and gene transcription, particularly elongation, though their exact roles in these processes remain unclear[7,8].

Complexes containing CDK12 and CDK13 regulate transcriptional elongation as well as processes occurring co-transcriptionally, including mRNA splicing and 3′-end RNA processing[9–11]. CDK12 and CDK13 aid in regulating RNA processing both directly, by physical interaction with RNA processing factors, and indirectly, by phosphorylation of the CTD, which recruits these processing factors[11–15]. Because of their roles in regulating these processes, loss of CDK12 and CDK13, or of their associated cofactor cyclin K, impedes both Pol II processivity and RNA processing. For example, CDK12 binds in exon junction complexes with other arginine–serine (RS) domain-containing splicing factors, including SRSF1, and its loss leads to mRNA splicing defects[11,14]. Factors involved in 3′-end cleavage and polyadenylation of RNA transcripts, including CstF-64 and CstF-77, are recruited to 3′ ends coincident with CTD Ser2 phosphorylation, which is dependent on CDK12 function. Depletion of CDK12 leads to simultaneous loss of Ser2 phosphorylation, reduced recruitment of the aforementioned factors, and subsequent 3′ processing defects[12,13,15]. Lastly, CDK12 lacking N-terminal RS domains also exhibit 3′-end processing defects, suggesting that dominant negative mutant forms of CDK12 that disrupt structure and physical interactions may also impact transcription[12].

CDK12–cyclin K and CDK13–cyclin K complexes exhibit both distinct and overlapping regulation of Pol II–mediated gene expression. Genetic depletion of CDK12 or CDK13 demonstrated that both complexes similarly regulate the expression of roughly 1,000 genes, including RNA processing genes[11], while separately regulating distinct classes of genes[11,16]. In particular, loss of CDK13, but not CDK12, decreases the expression of genes encoding proteins that regulate protein translation[11]. Conversely, depletion of CDK12, but not CDK13, reduces the expression of core members of the DNA damage response (DDR), leading to a marginal increase in unrepaired double-strand breaks and an increased susceptibility to DNA damaging agents[11,16–19]. Interestingly, breast and ovarian cancers harboring inactivating mutations in *BRCA1* and *BRCA2*, two members of the homologous recombination DDR pathway, demonstrate synthetic lethality with pharmacological inhibition of PARP1, which repairs single-strand breaks[20]. Consistent with these findings, genetic screens have indicated that loss of CDK12 potentiates a response to PARP1 inhibition[21,22]. These data suggest that CDK12 and CDK13 inhibitors alone or in combination treatment may

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [4]Structural Genomics Consortium, University of Oxford, Oxford, UK. [5]Department of Structural Immunology, Institute of Innate Immunity, University of Bonn, Bonn, Germany. [6]Center of Advanced European Studies and Research, Bonn, Germany. [7]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [8]Pharmaceutical and Medicinal Chemistry, Department of Pharmacy, Saarland University, Saarbrücken, Germany. [9]Department of Biochemistry, Duke University Medical Center, Durham, North Carolina, USA. [10]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [11]These authors contributed equally to this work. *e-mail: young@wi.mit.edu or Nathanael_Gray@dfci.harvard.edu

represent an effective strategy for targeting cancers that demonstrate extreme dependencies on the expression of CDK12- and CDK13-regulated genes.

While genetic depletion of CDK12 and CDK13 indicates that the proteins themselves are required for gene transcription, how CDK12 and CDK13 kinase activities regulate these processes remains unclear. Mutational studies using CDK12 variants that either are devoid of N-terminal RS domains or cannot complex with cyclin K indicate that loss-of-function mutants may elicit phenotypes comparable to deletion[12,19]. Recently, CRISPR–Cas9-mediated gene editing was used to generate cells expressing a CDK12 mutant that is selectively inhibitable by a cell-permeable adenine analog[23]. While this approach permits investigation of CDK12 kinase activity, its utility is limited to a particular genetic background. Therefore, identification of inhibitors of wild-type CDK12 and CDK13 function is necessary to understand the effects of pharmacologically inhibiting CDK12 and CDK13 in normal and disease states.

Inspection of the recently published CDK12–cyclin K and CDK13–cyclin K co-crystal structures reveals that CDK12 Cys1039 and CDK13 Cys1017 occupy similar spatial positions to Cys312 in CDK7 (refs. 24–26). Therefore, we surmise that we could use CDK7 inhibitor THZ1, which covalently targets Cys312 of CDK7 and CDK12 and CDK13 at higher concentrations, as starting material for developing selective CDK12 and CDK13 agents[27]. Here we report the development of THZ531 as a new covalent inhibitor of CDK12 and CDK13 and present a 2.7-Å co-crystal structure with CDK12–cyclin K that shows THZ531 bound to the ATP pocket and connected to Cys1039 via its flexible linker. Treatment of Jurkat cells with THZ531 resulted in diminished global Ser2 phosphorylation (pSer2). Closer inspection indicated that loss of pSer2, and resulting loss of elongating Pol II, occur specifically at THZ531-responsive genes, indicating that loss of transcription elongation following CDK12 and CDK13 inhibition contributes to loss of gene expression. Analysis of the genes whose expression was particularly sensitive to treatment with THZ531 showed two classes of genes: at low doses of THZ531, genes encoding DDR factors were exquisitely sensitive, whereas complete inhibition of CDK12 and CDK13 with a high dose of THZ531 led to significant downregulation of key Jurkat transcription factors. Concomitant with this extreme transcriptional stress, potent induction of apoptosis was observed. Thus, inhibition of CDK12- and CDK13-mediated transcription in tumors via targeting of CDK12 and CDK13 kinase activity may be a potent means of affecting oncogenic cell survival.

RESULTS

THZ531 covalently inhibits CDK12 and CDK13

We previously described the characterization of THZ1, a covalent inhibitor of CDK7, which has the unprecedented ability to covalently modify Cys312 located on a C-terminal extension of the kinase domain that traverses the ATP-binding site. Affinity chromatography with a biotin-modified version of THZ1 revealed that it is additionally capable of covalently binding to CDK12 and CDK13, both of which were also predicted by sequence alignment to have C-terminal cysteines[27]. Comparison of the CDK7 and CDK12 crystal structures revealed that indeed both kinases have similarly positioned, yet spatially distinct, cysteines in their C-terminal domains (**Fig. 1a**)[24,26]. We reasoned that modification of the substituent bearing the acrylamide could serve to alter the orientation of the acrylamide warhead to selectively target Cys1039 of CDK12. To discriminate amongst CDK7, CDK12, and CDK13 binders, we developed a 'target-engagement' assay, where the ability of new compounds to protect these kinases from affinity purification by biotinylated THZ1 (bioTHZ1) was assessed by western blot (see detailed protocol in **Supplementary Results**, **Supplementary Fig. 1a**). Following the iterative synthesis and testing of a variety of analogs (to be reported elsewhere), we discovered THZ531 (**1**), which possesses a piperidine ring in the acrylamide 'arm' (**Fig. 1b**). We discovered that this configuration resulted in more selective binding of THZ531 to CDK12–cyclin K and CDK13–cyclin K relative to CDK7–cyclin H and CDK9–cyclin T (**Fig. 1c** and **Supplementary Fig. 1b**). Mass spectrometry of recombinant CDK12–cyclin K complex incubated with the inhibitor demonstrated covalent modification of Cys1039 of CDK12 (**Supplementary Fig. 1c**,**d**). To corroborate covalent bond formation in cell lysates, we synthesized a biotinylated analog of THZ531 (bioTHZ531, **2**) that could affinity purify covalently bound targets with streptavidin. Inspection of precipitated proteins confirmed CDK12–cyclin K and CDK13–cyclin K complexes as the main targets of THZ531 (**Fig. 1d**, **Supplementary Data Set 1**).

To confirm that THZ531 inhibits the enzymatic activity of CDK12 and CDK13, we performed a radiometric kinase assay measuring the ability of recombinant CDK12 and CDK13 to phosphorylate a Pol II CTD-peptide substrate[24]. In fixed end-point kinase assays, THZ531 potently inhibited CDK12 and CDK13 with half-maximal inhibitory concentrations (IC_{50}s) of 158 nM and 69 nM, respectively, whereas inhibition of CDK7 and CDK9 was more than 50-fold weaker, with IC_{50}s of 8.5 μM and 10.5 μM, respectively (**Fig. 1e**). In contrast, THZ531R (**3**), where the electrophilic acrylamide is replaced with a propyl amide incapable of covalent reaction, and THZ532 (**4**), the enantiomer of THZ531 (**Supplementary Fig. 1e**), were both 50- to 100-fold less active on CDK12 and CDK13, but exhibited similarly low inhibition activity on CDK7 and CDK9, implying that THZ531 inhibits the enzymatic functions of CDK12 and CDK13 and that covalent binding was important for this activity (**Supplementary Fig. 2a–c**). As an additional control, THZ531 showed no appreciable inhibitory effect on ERK1, an example used for the MAP kinase family of the CMGC group (**Supplementary Fig. 2d**). As with THZ1, increasing incubation time augmented THZ531 potency, further supporting the notion that covalent binding of THZ531, which occurs with slower kinetics than reversible binding, is critical for activity (**Fig. 1f** and **Supplementary Fig. 2e**,**f**).

To broadly survey kinase selectivity, we performed *in vitro* profiling using the Ambit binding technology and intracellular profiling using KiNativ profiling, which measures the ability of a compound to block nucleotide-dependent enzymes from biotinylation with a reactive ATP–desthiobiotin probe[28]. Ambit profiling indicated that CDK13 was among the top scoring kinases (**Supplementary Data Set 2**). KiNativ profiling confirmed CDK12 and CDK13 as the primary targets of THZ531 and demonstrated the excellent selectivity of the inhibitors; none of the other 211 kinases were labeled to more than 55% percent (**Supplementary Table 1** and **Supplementary Fig. 2g**).

THZ531 co-crystal structure with CDK12–cyclin K

To understand how THZ531 selectively targets Cys1039 in CDK12, we determined its co-crystal structure with CDK12–cyclin K (see **Supplementary Table 2** for diffraction data collection and refinement statistics). Two CDK12–cyclin K complexes were found in the asymmetric unit, each binding to THZ531 in a different rotamer (**Fig. 2a–c**). The 2.7-Å structure revealed that the labile αK helix can be displaced from CDK12, allowing Cys1039 to reorient toward the ATP-binding pocket for cross-linking. The aminopyrimidine bound to the kinase hinge region, forming two hydrogen bonds to the backbone of M816, while the pendant 3-indolyl and piperazine groups mediated further hydrophobic interactions. Different conformations of the secondary amide were observed in the two THZ531 complexes, causing the solvent-exposed groups to pack either against the N-lobe β1 strand or C-lobe αD helix. These exposed sites were poorly defined, except for the positions of Cys1039, which were clearly visible in the electron density maps (**Supplementary Fig. 3**). Notably, the equivalent sulfhydryl in CDK7, Cys312, was at a more distant site (over 12 Å from CDK12 Cys1039), suggesting a basis for the weaker binding of THZ531 to CDK7.

© 2016 Nature America, Inc. All rights reserved.

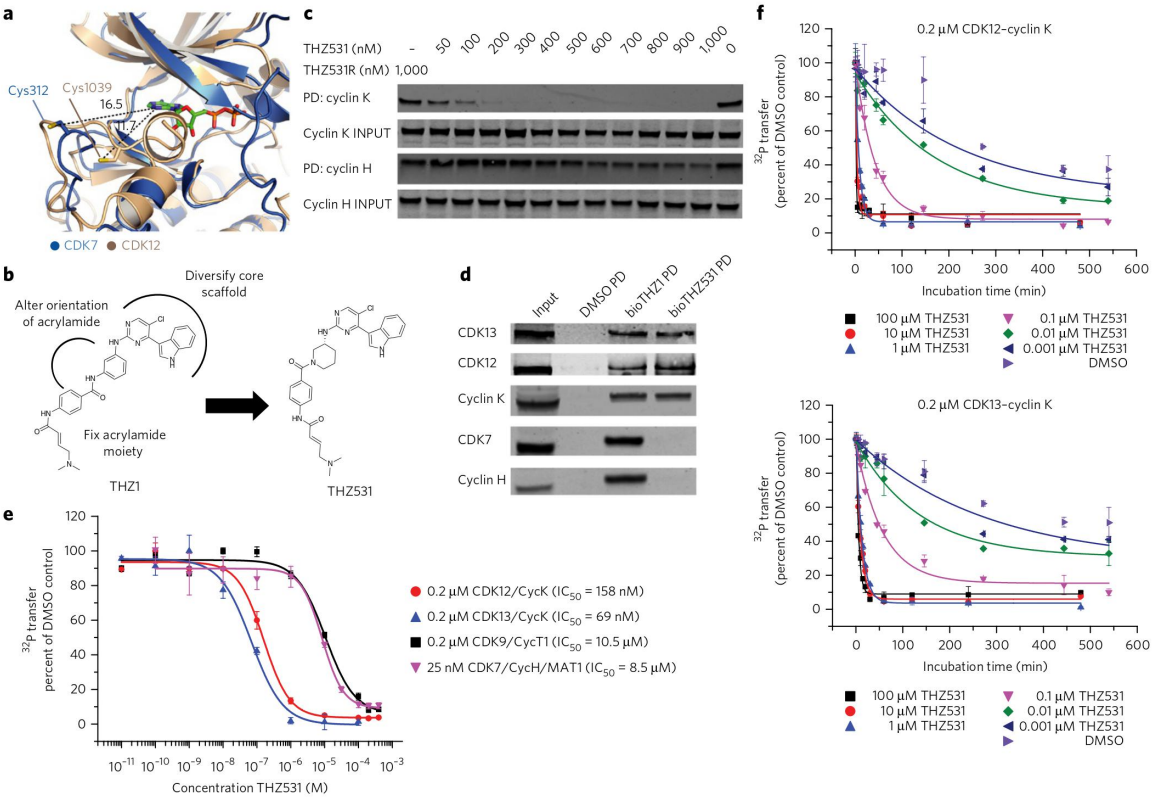


Figure 1 | THZ531 covalently inhibits CDK12 and CDK13. (**a**) Overlay of CDK7 and CDK12 crystal structures reveals similarly positioned cysteines in the C-terminal extension segments of the two kinase domains. Cys312 was modeled into the crystal structure of CDK7 as previously described[27]. PDB codes, 1UA2 (CDK7) (ref. 26) and 4NST (CDK12) (ref. 24). (**b**) Structures of THZ1 and THZ531 and rational design strategy to synthesize THZ531 from THZ1. (**c**) THZ531 binds intracellular CDK12–cyclin K and CDK13–cyclin K complexes with higher affinity than CDK7–cyclin H complexes out competing lysate-introduced bioTHZ1 in target engagement assay. Cells were treated with DMSO or THZ531, lysed, and treated with bioTHZ1 to pull down available CDK12, CDK13 and CDK7 complexes. PD, pulldown. (**d**) Biotinylated THZ531 (bioTHZ531, 1 μM) pulls down CDK12–cyclin K and CDK13–cyclin K complexes from cellular lysates, but not CDK7–cyclin H complexes, whereas bioTHZ1 (1 μM) pulls down CDK7, CDK12, and CDK13 complexes from cellular lysates. PD, pulldown. (**e**) THZ531 potently inhibits CDK12 and CDK13. IC_{50} values for CDK12 and CDK13 were 158 and 69 nM, respectively, compared to 10.5 μM for CDK9 and 8.5 μM for CDK7. Measurements were performed in triplicate and data represent mean values ± s.d. (**f**) THZ531 inhibition of CDK12 and CDK13 is time dependent. *In vitro* kinase activity assay of CDK12–cyclin K (top) and CDK13–cyclin K (bottom) with different concentrations of THZ531 and varying preincubation times. For all incubation time series, the counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Measurements were performed in triplicate and data represent the mean values ± s.d. Uncut western blots are in **Supplementary Figure 10**.

THZ531 induces apoptosis in Jurkat cells

THZ1, the predecessor to THZ531, displays strong anti-proliferative activity in models of various cancer types[27,29,30]. To evaluate if THZ531 exhibits similar anti-cancer activity, we treated Jurkat cells with THZ531 in escalating doses for 72 h. THZ531 treatment led to a dramatic and irreversible decrease in Jurkat cell proliferation with an IC_{50} of 50 nM (**Fig. 3a**,**b**), which correlates with the IC_{50} concentration required to achieve 50% covalent CDK12 and CDK13 binding (**Fig. 1c**). In agreement with THZ531R and THZ532's reduced anti-CDK12 and CDK13 activity, both compounds displayed minimal antiproliferative activity in Jurkat cells (**Fig. 3a**,**b**). FACS cell cycle analysis following treatment with escalating doses of THZ531 displayed a dose- and time-dependent increase in the number of cells exhibiting sub-G1 content, which is suggestive of cell death (**Fig. 3c**). To investigate this further, we used flow cytometry to assess annexin V (AV) and propidium iodide (PI) staining to identify apoptotic and dead cells, respectively. At 50 nM THZ531, no increase in the percentage of apoptotic cells was observed over DMSO control for the time course of the experiment (**Fig. 3d**). However, higher doses of THZ531 led to a pronounced annexin V signal, with 30–40% annexin-V-stained cells by 72 h (**Fig. 3d** and **Supplementary Fig. 4a**). Similarly, PARP (PARP1) cleavage further indicated that cell death as a result of apoptosis occurred following THZ531 administration (**Supplementary Fig. 4b**). Thus, THZ531 induces apoptosis in a dose- and time-dependent manner, with low doses (<350 nM) leading to a slow onset of apoptosis over several cell doublings, while doses exceeding 350 nM induce rapid apoptosis that may result from a combination of on- and off-target effects.

To assess whether mutation of CDK12 Cys1039 to a less nucleophilic serine residue could rescue THZ531-mediated effects, we used CRISPR–Cas9 to mutate the CDK12 gene locus (**Supplementary Fig. 5a**,**b**). This substitution was sufficient to make CDK12 refractory to covalent affinity pulldown and led to partial restoration of cellular proliferation and reduced apoptosis (**Fig. 3e**,**f** and **Supplementary Fig. 5c**). For these experiments, the chronic myeloid leukemia cell line HAP1 was used, as technical problems precluded the use of

© 2016 Nature America, Inc. All rights reserved.

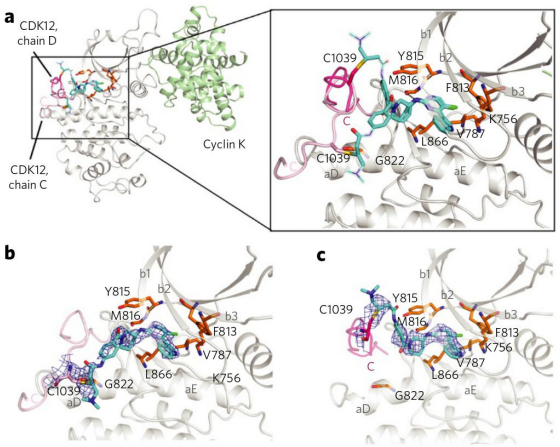


Figure 2 | THZ531 co-crystal structure with CDK12–cyclin K. (**a**) THZ531 binds to M816 in the kinase hinge region and connects to Cys1039 in two conformations via the compound's flexible linker. Solvent-exposed regions of THZ531 with poor electron density are represented by thin sticks. (**b**) Omit map contoured at 2.5σ for THZ531 bound to CDK12 chain C. (**c**) Omit map contoured at 2.5σ for THZ531 bound to CDK12 chain D.

CRISPR–Cas9 in Jurkat cells. Additionally, we believe that the partial rescue may result, in part, from the continued presence of WT CDK13 or of an unknown off-target.

CDK12 binds regulatory and coding regions of active genes

Given the role that CDK12 and CDK13 play in transcriptional regulation, we next sought to explore the potential transcriptional underpinnings of this anti-proliferative effect. CDK12 is believed to regulate Pol II–mediated transcription. However, with the exception of *Drosophila*, in which CDK12 was shown to colocalize with hyperphosphorylated Pol II (II0) on larval polytene chromosomes and to ChIP mainly to gene interiors on cultured cell chromosomes, little is known about how CDK12 interacts with the genome[7]. To address this question, we performed chromatin immunoprecipitation coupled to massively parallel DNA sequencing (ChIP–seq) using an antibody against CDK12 (ref. 14). Genome-wide analysis of the CDK12 ChIP–seq signal indicated that CDK12 bound to both protein-coding genes and enhancer regions (**Fig. 4a**). At enhancers, the CDK12 ChIP–seq signal coincided with Pol II and histone H3K27Ac, marks associated with active enhancers and the production of enhancer RNAs, suggesting that CDK12 binds to active enhancers (**Fig. 4a**). In protein-coding genes, CDK12 ChIP–seq signal was found at promoters and gene bodies, where it largely overlapped with Pol II signal, indicating that it may travel with elongating Pol II (**Fig. 4a**,**b**). Furthermore, the intensity of CDK12 ChIP–seq signal at both enhancers and promoters increased commensurate with Pol II and H3K27Ac signals, implying that CDK12 levels correlate with transcriptional activity (**Fig. 4c**,**d**). Therefore, these data suggest that CDK12 is present at actively transcribed regions of the genome.

THZ531 inhibits transcriptional elongation

CDK12 is implicated in the control of elongating Pol II, in part through phosphorylation of the Pol II CTD. To assess whether CDK12 and CDK13 kinase activity is required for Pol II CTD phosphorylation, we treated Jurkat cells with THZ531 at doses across the effective concentration range of the inhibitor (**Fig. 1c**). In contrast to THZ1, which causes complete loss of Pol II CTD phosphorylation[27], THZ531 selectively reduced Ser2 phosphorylation levels without appreciable effect on CTD pSer5 or pSer7 levels (**Fig. 5a**). To understand how these findings extend to gene-associated pSer2 signal, we conducted a genome-wide analysis of pSer2 ChIP–seq signal. In comparison to total Pol II ChIP–seq signal that displays enrichment primarily at the transcription start sites (TSS), pSer2 ChIP–seq signal shows additional pronounced signal in the gene body, particularly near the 3′ ends of genes, consistent with previous data indicating that pSer2 is an established mark of elongating Pol II (**Fig. 5b**). While 50 nM THZ531 treatment did not appreciably change pSer2 ChIP–seq signal, 500 nM THZ531 led to a dramatic decrease in 3′-end-associated pSer2 ChIP–seq signal with a reciprocal accumulation of 5′-end-associated pSer2 signal (**Fig. 5b**).

Reduction of Pol II CTD Ser2 phosphorylation, a mark of active transcriptional elongation, led us to investigate whether THZ531 affects gene expression. To address this question, we treated Jurkat cells with increasing doses of THZ531 for 6 h and examined the effect on steady-state mRNA levels. Treatment with 50 nM THZ531, a concentration that leads to minimal loss of CDK12 activity and pSer2 levels, resulted in the loss of expression of a small subset of genes (**Fig. 5c**, **Supplementary Data Set 3**). Further dose escalation of THZ531, using concentrations that resulted in the complete targeting of CDK12 and greater reduction of pSer2 levels, led to gradual erosion of mRNA output, with 25% and 46% of transcripts displaying at least 2- and 1.5-$\log_2$ fold reduction (4 and 2.25 linear fold change) in expression, respectively, following 500 nM THZ531 treatment (**Fig. 5c**). CRISPR–Cas9-mediated replacement of CDK12 WT with CDK12 C1039S was sufficient to partly restore Ser2 CTD phosphorylation and gene expression (**Fig. 5d**,**e** and **Supplementary Fig. 6a**). We believe that the partial restoration of gene expression may result from continued CDK13 activity or from an unknown off-target. Consistent with a lack of anti-CDK12 activity, THZ531R and THZ532 decreased neither Ser2 phosphorylation nor steady-state mRNA levels (**Fig. 5a** and **Supplementary Fig. 6b**). Gene expression analysis following THZ1 treatment revealed a similar, but not identical, transcriptional downregulation, consistent with the general role of CDK7 in mediating the transcription initiation-to-elongation step[31] (**Supplementary Fig. 6c**).

THZ531 treatment did not completely abolish Ser2 phosphorylation, suggesting that loss of Ser2 phosphorylation may occur in a gene-specific manner. To understand if THZ531-responsive genes preferentially suffer loss of Ser2 phosphorylation upon THZ531 treatment, thereby contributing to their sensitivity, we first analyzed pSer2 Pol II ChIP–seq signal at genes whose expression is reduced with THZ531 treatment. Coincident with loss of gene expression, we observed reduced pSer2 ChIP–seq signal near the 3′ ends of THZ531-responsive genes (**Fig. 5f** and **Supplementary Fig. 7a**). Global analysis of pSer2 ChIP–seq density at all THZ531-responsive genes (>2 $\log_2$ fold reduction in gene expression) demonstrated a significant decrease in 3′-end-associated and an increase in 5′-end-associated pSer2 ChIP–seq signal density compared to non-responsive genes, suggestive of a loss of Pol II processivity at THZ531-responsive genes (**Fig. 5g**).

In agreement with a marked decrease in gene-body-associated pSer2 ChIP–seq signal, THZ531 led to a dramatic depletion of elongating Pol II at these same genes (**Fig. 5f**,**h** and **Supplementary Fig. 7a**). Importantly, the Pol II elongation defect observed following THZ531 was significantly different from that obtained following inhibition of CDK9 by flavopiridol treatment. Whereas flavopiridol causes an arrest proximal to TSS, THZ531 treatments results in reduced TSS-associated Pol II and reduced Pol II processivity across the gene body (**Supplementary Fig. 7b**,**c**). These results suggest that THZ531 downregulates gene expression, in part through the loss of Pol II processivity.

© 2016 Nature America, Inc. All rights reserved.

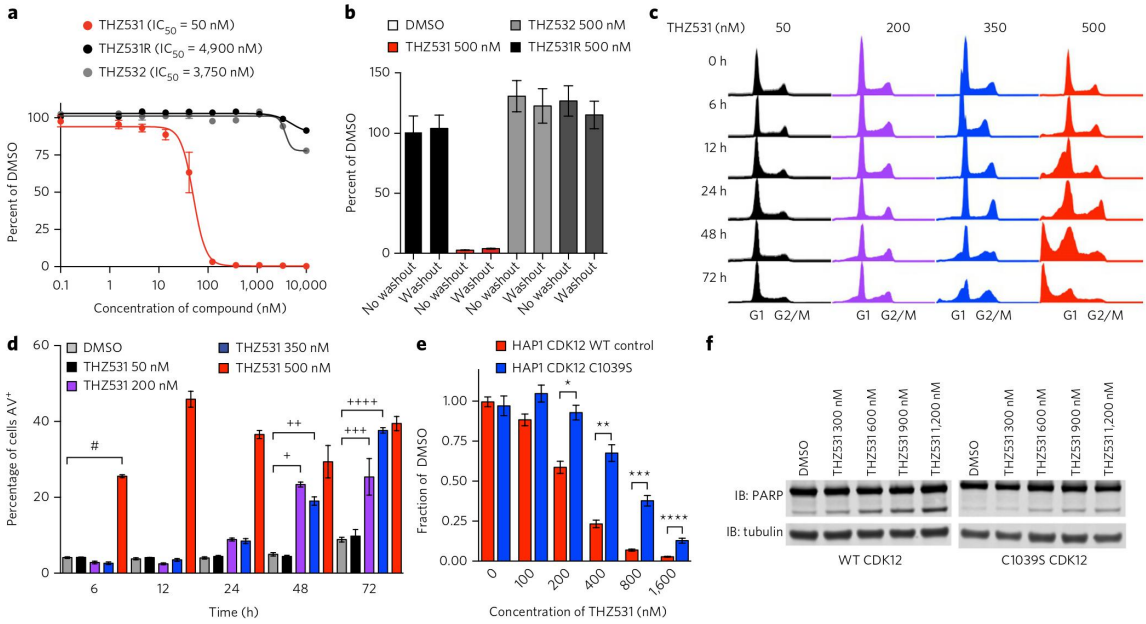


Figure 3 | THZ531 induces apoptosis in Jurkat cells. (**a**) THZ531, but not THZ531R or THZ532, exhibits strong antiproliferative effects in Jurkat T-cell acute lymphoblastic leukemia cells. Cells were treated with the indicated compounds for 72 h and assessed for antiproliferative effect (as compared to cells given only the DMSO vehicle) using CellTiter Glo. (**b**) THZ531 demonstrates irreversible suppression of Jurkat cell proliferation. Jurkat cells were treated with the indicated compounds for 6 h. Cell growth medium containing inhibitors was then removed (washout) and cells were allowed to grow for the remainder of the 72 h proliferation assay without inhibitors. CellTiter Glo results were compared to cellular treatments whereby cells were treated with inhibitors for the full 72 h (no washout). (**c**) THZ531 increases sub-G1 population. Jurkat T-ALL cells were treated with THZ531 for the indicated time periods. Cell cycle progression was assessed using FACS cell cycle analysis. (**d**) Treatment with THZ531 induces apoptosis. Jurkat cells were treated with THZ531 for the indicated times. Cells were stained with annexin V and propidium iodide. All proliferation and apoptosis assays were performed in biological triplicate and error bars are ± s.d. $^{\#}P$ = 3.33e-07, ^{+}P = 1.04e-04, ^{++}P = 1.35e-04, ^{+++}P = 9.94e-04, ^{++++}P = 1.53e-07. (**e**) 72-h antiproliferation assay using WT and C1039S HAP1 cells. Cells were analyzed using Cell Titer Glo. The experiment was performed in biological triplicate and error bars are ± s.d. $^{*}P$ = 5.84e-04, $^{**}P$ = 1.76e-04, $^{***}P$ = 9.09e-05, $^{****}P$ = 4.25e-04. (**f**) Left, expression of C1039S CDK12 partially reduces PARP cleavage in HAP1 cells. Right, western blot analysis of indicated proteins in HAP1 CDK12 WT and C1039S cells. Uncut western blots are in **Supplementary Figure 10**. *P* values were determined with a two-tailed Student's *t*-test.

© 2016 Nature America, Inc. All rights reserved.

THZ531 reduces DDR and super-enhancer gene expression

We then asked whether the loss of expression of certain genes or gene sets could contribute to the sensitivity of Jurkat cells to THZ531. To address this, we subjected the set of gene transcripts sensitive to THZ531 to gene ontology (GO) term enrichment analysis and identified two distinct gene signatures. At 50 nM THZ531 we observed enrichment for DDR-pathway genes among those genes most sensitive to THZ531 (**Fig. 6a** and **Supplementary Fig. 8a**). Core DDR genes including *BRCA1*, *FANCF*, and *ERCC4* all showed substantial loss of expression with 50 nM THZ531, an effect that was further exacerbated at higher doses (**Fig. 6b** and **Supplementary Fig. 8b**). Loss of DDR gene expression following CDK12–cyclin K knockdown is believed to result, in part, from reduced active transcription across these genes[16]. Indeed, we also observed reduced Pol II occupancy at these genes commensurate with the loss of gene expression (**Fig. 5** and **Supplementary Fig. 7**).

Interestingly, while the transcripts most sensitive to 50 nM THZ531 were enriched for regulators of DDR pathway, the transcripts that were most sensitive to higher doses of THZ531 (200 nM or greater) were enriched for transcripts that encode transcription factors, including those known to be drivers of Jurkat T-cell acute lymphoblastic leukemia (**Fig. 6c** and **Supplementary Fig. 8c**). Among these transcription factors are *RUNX1, MYB, TAL1*, and *GATA3*, which comprise the Jurkat transcriptional autoregulatory loop (**Fig. 6d** and **Supplementary Fig. 8d**)[32,33]. All of these transcription factor genes are associated with super-enhancers that were previously shown to contain disproportionate amounts of transcriptional apparatus, including CDK7 (**Fig. 6e**, **Supplementary Fig. 8e** and **Supplementary Data Set 4**) (ref. 27). Indeed, super-enhancer-associated genes were collectively more downregulated upon THZ531 treatment than those associated with typical enhancers (**Supplementary Fig. 8f**). Analyzing genome-wide CDK12 ChIP–seq signal revealed that these genes, in particular *RUNX1*, *MYB*, *TAL1*, and *GATA3*, as well as their associated super-enhancers, contained large amounts of CDK12 ChIP–seq signal as well as Pol II (**Fig. 6f** and **Supplementary Fig. 9a**,**b**). Interestingly, we observed a dramatic reduction in elongating Pol II following THZ531 treatment, in contrast to a reduction in overall Pol II occupancy following THZ1 treatment (**Supplementary Fig. 9a**). As CDK12 has been implicated in the phosphorylation of Ser2, a mark of transcriptional elongation, these results are consistent with CDK12 inhibition. Therefore, we asked whether the profound loss of gene expression following 200 nM THZ531 treatment correlated with the levels of CDK12 (i.e. THZ531 target). Gene set enrichment analysis of the top 500 transcripts downregulated by 200 nM THZ531 revealed that the most THZ531-responsive genes tended to be particularly enriched in promoter-proximal and enhancer-bound CDK12 signal (**Fig. 6g** and **Supplementary Fig. 9c**). This indicated that at

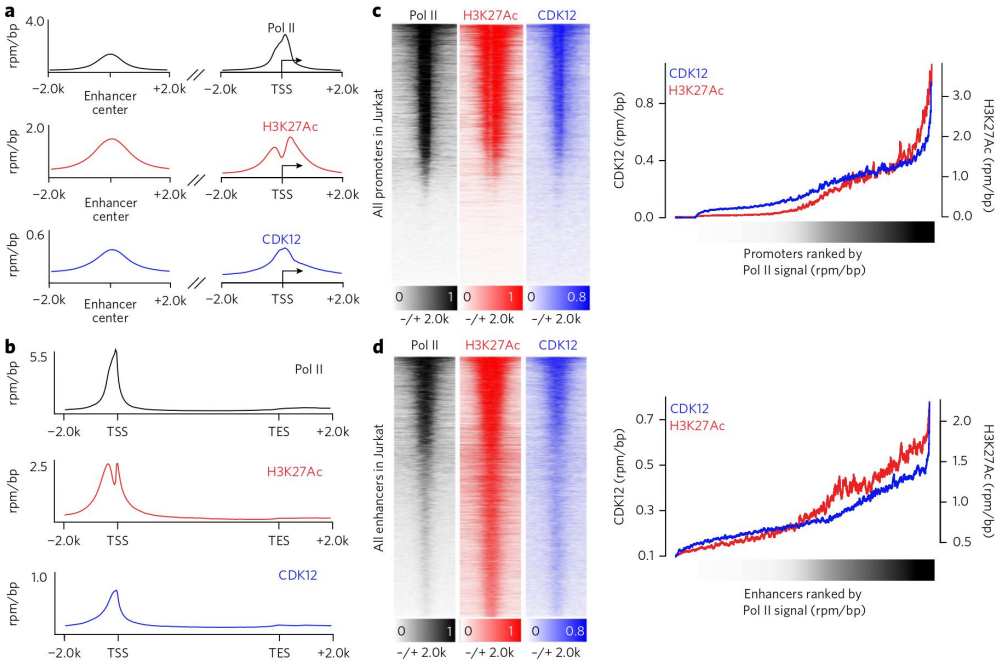


Figure 4 | CDK12 binds regulatory and coding regions of active genes. (**a**) Pol II, H3K27Ac, and CDK12 co-occupy active enhancers and promoters. Metagene representation of average Pol II, H3K27Ac, and CDK12 occupancy at H3K27ac-defined enhancers (left) and 37974 RefSeq promoters (right). (**b**) Pol II, H3K27Ac, and CDK12 bind at promoters and across gene bodies. Metagene representation of global Pol II, H3K27Ac, and CDK12 occupancy at promoters and gene bodies. Average ChIP-seq signal in 13906 RefSeq genes expressed in 6h DMSO conditions. (**c**) CDK12, Pol II, and H3K27Ac co-occupy active gene promoters. Left, heatmaps show the density of Pol II (left), H3K27Ac (center), and CDK12 (right) ChIP-Seq reads relative to RefSeq promoters and comprising the region ±2 kb from RefSeq transcription start sites. Right, ChIP-seq read densities for CDK12 and H3K27ac at RefSeq promoters of 4 kb centered on transcription start sites. (**d**) CDK12, Pol II, and H3K27Ac co-occupy active enhancers. Left, heatmaps show the density of Pol II- (left), H3K27Ac- (center), and CDK12- (right) ChIP-Seq reads relative to 32855 H3K27ac-enriched enhancer sites. For each H3K27ac-bound site (*y* axis), the densities of H3K27Ac (red) and CDK12 (blue) are displayed within a 2-kb window centered on the center of the H3K27ac-bound site. Right, ChIP-seq read densities for CDK12 and H3K27ac at enhancers of 4 kb centered on the middle of H3K27ac peaks. For all heatmaps, promoters and enhancers were ranked by Pol II density (black) and a smoothing function was applied. All metagenes are presented in units of rpm/bp. k, kilobases of DNA.

© 2016 Nature America, Inc. All rights reserved.

saturating doses of THZ531, CDK12 target occupancy may contribute to gene expression sensitivity to THZ531.

DISCUSSION

Here we have reported the rational design and characterization of the first covalent inhibitor of CDK12 and CDK13 enzymatic activities. The CDK12 crystal structure revealed that it contains an extended C-terminal domain that traverses the kinase domain, positioning a cysteine residue proximal to the ATP-binding site in a manner analogous to CDK7. Therefore, we reasoned that THZ1, a recently reported CDK7 inhibitor that targets the CDK7 C-terminal cysteine, would make a suitable synthetic starting point. Indeed, we show that simple chemical modification to alter the orientation of the electrophilic acrylamide permitted us to selectively target CDK12 and CDK13. We demonstrated that covalent binding to CDK12 and CDK13 is critical for the potency and selectivity of THZ531, as its reversible analog is more than 100-fold less active. Therefore, while high sequence and structural similarity within ATP-binding pockets poses a significant challenge to developing selective ATP-competitive inhibitors of CDKs, targeting unique cysteine nucleophilic hotspots, particularly those lying outside the conserved kinase domain, may circumvent these confounding factors. It remains to be seen whether this strategy can be readily deployed to target other cysteine-targetable members of the kinome (or proteome). The identification of starting chemical matter that can be modified to deliver an electrophilic warhead proximal to these cysteines presents the first hurdle.

We found that pharmacological inhibition of CDK12–cyclin K and CDK13–cyclin K complexes with THZ531 led to diminished pSer2 and elongating Pol II at THZ531-target genes. This post-translation modification has previously been ascribed to CDK9 kinase activity, largely from studies using pharmacological agents. However, recent work has shown that flavopiridol inhibits CDK12 at concentrations that decrease Ser2 phosphorylation, suggesting that CDK12 may influence the pharmacology of this inhibitor[14,24]. Furthermore, using several unbiased and orthogonal approaches, we found no evidence that THZ531 targets CDK9, which lacks a C-terminal cysteine. Lastly, Pol II occupancy at THZ531-responsive genes following drug treatment is distinct from that of flavopiridol. Whereas Pol II undergoes a pronounced pause proximal to gene promoters following flavopiridol treatment, following THZ531 treatment, Pol II entered the gene body but experienced reduced processivity, suggesting a defect in elongation proficiency or termination steps. These data would seem to support existing models of Pol II regulation in other organisms, including *Saccharomyces cerevisiae*, in which regulation of transcriptional elongation is sequentially and differentially regulated by *BUR1* and *CTK1* (ref. 34).

Coincident with THZ531 decreasing the elongating Pol II population, THZ531 downregulated the expression of certain genes in Jurkat T-ALL. We found two general signatures of genes that were

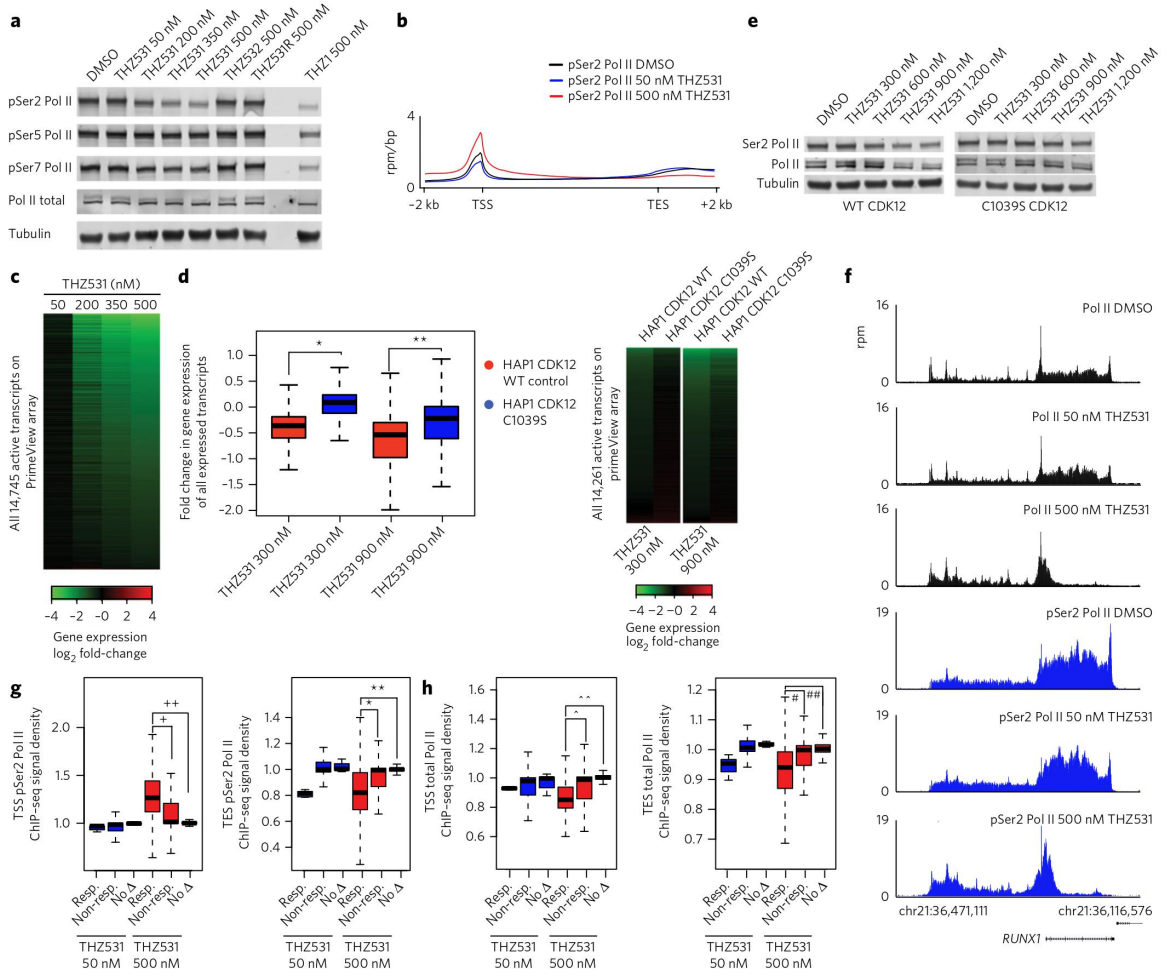


Figure 5 | THZ531 inhibits transcriptional elongation. (**a**) THZ531 reduces Pol II CTD Ser2 phosphorylation (pSer2) in a concentration-dependent manner. (**b**) THZ531 reduces 3′ pSer2 Pol II signal density. Metagene representation of global pSer2 Pol II occupancy at gene bodies. Average ChIP-seq signal (rpm/bp) across 13,906 genes expressed in DMSO. TSS, transcriptional start sites; TES, transcriptional end sites. kb, kilobases of DNA. (**c**) THZ531 downregulates mRNA levels in a concentration-dependent manner. Heatmaps display the $\log_2$ fold change in gene expression versus DMSO for 14,745 transcripts expressed in DMSO. Heatmap is ranked by 500 nM THZ531. (**d**) C1039S CDK12 expression partially restores gene expression in HAP1 cells. Box plots (left) and heatmaps (right) displaying expression of 14,261 expressed transcripts . $^{*}P < 10^{-15}$, $^{**}P < 10^{-15}$. Heatmaps are ranked by WT HAP1 fold-change. (**e**) C1039S CDK12 expression partially restores pSer2 in HAP1 cells. (**f**) THZ531 reduces elongating Pol II at THZ531-response genes. (**g**,**h**) THZ531 reduces 3′ pSer2 Pol II (**g**) and Pol II (**h**) ChIP-seq signal at THZ531-responsive genes. Box plots of pSer2 Pol II and Pol II signal density at TSS (top) and TES (bottom) at THZ531-responsive genes (Resp.) and equal numbers of randomly selected non-responsive genes (Non-resp.) or genes exhibiting no change (no Δ). 7 and 3801 genes were sensitive to THZ531 at 50 and 500 nM THZ531, respectively. ^{+}P = 1.64e-252, ^{++}P = 0, $^{*}P$ = 1.31e-136, $^{**}P$ = 9.00e-180. $^{\wedge}P$ = 1.17e-231, $^{\wedge\wedge}P$ = 0, $^{\#}P$ = 4.43e-101, $^{\#\#}P$ = 1.21e-155. Sensitive genes are defined by >2 $\log_2$ fold change in gene expression. Uncut western blots are in **Supplementary Figure 10**. *P* values were determined with a two-tailed Student's *t*-test.

particularly sensitive to THZ531 treatment. At concentrations leading to complete CDK12 and CDK13 inhibition, genes encoding factors that regulate transcription were exceptionally sensitive to THZ531. At these concentrations, the loss of expression of the most sensitive genes positively correlated with the amount of promoter- and enhancer-bound CDK12, with super-enhancer-associated genes being especially sensitive. These data suggest that inhibition of CDK12, like that of CDK7, may provide another means of targeting super-enhancer-associated oncogene expression in various cancers. Additionally, we found that genes involved in DDR were exquisitely sensitive to low doses of THZ531 and continue to lose expression with escalating doses of THZ531. A similar set of DDR genes was recently shown to be sensitive to loss of CDK12– and CDK13–cyclin K complexes[11,16]. In contrast to THZ531 treatment, neither depletion of CDK12 and CDK13, nor mutations that disrupt complex integrity, lead to gross disruption of the larger gene expression program. This discrepancy could be explained by numerous factors. We favor the explanation that THZ531 treatment effectively creates inactive or dominant negative CDK12 and CDK13 complexes, eliciting fundamentally different effects on transcription. However, the modest recovery of wild-type function following expression of CDK12 C1039S suggests that remaining CDK13 activity,

© 2016 Nature America, Inc. All rights reserved.

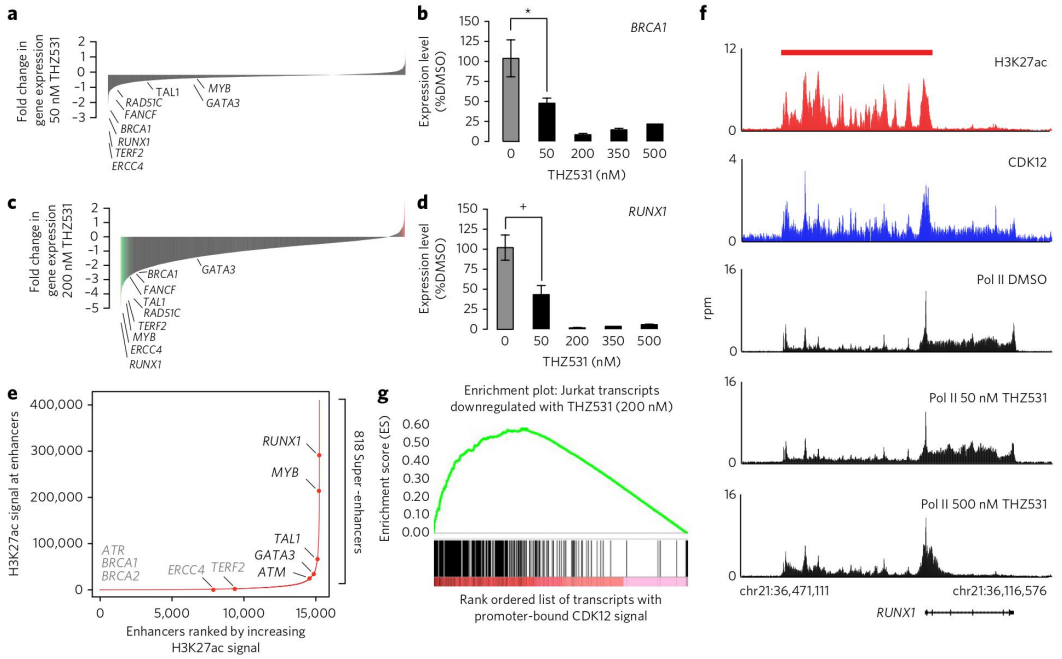


Figure 6 | THZ531 reduces DDR and super-enhancer gene expression. (**a**) DDR genes are sensitive to 50 nM THZ531. (**b**) RT-qPCR of *BRCA1* gene expression following THZ531 treatment. *P = 9.42e-05, (**c**) Transcription factor genes are sensitive to 200 nM THZ531. (**d**) RT-qPCR of *RUNX1* gene expression following THZ531 treatment. *P = 5.33e-06. (**e**) Total H3K27Ac ChIP-seq signal in enhancer regions for all stitched enhancers ranked by increasing H3K27Ac ChIP-seq signal. Gene names in gray either have no associated enhancer or associated typical enhancers. (**f**) *RUNX1* gene contains exceptional CDK12 and H3K27Ac signal and suffers profound loss of elongating Pol II following THZ531 treatment. CDK12 (blue), H3K27Ac (red) and Pol II (black) ChIP-seq occupancy at the promoter and across the gene body of *RUNX1* locus following THZ531 treatment. The red bar indicates the genomic coordinates of a super -enhancer. (**g**) Transcripts downregulated by 200 nM THZ531 are enriched for transcripts whose gene promoters contain high levels of CDK12. Gene set enrichment analysis of top 500 genes downregulated following THZ531 treatment in comparison to CDK12 promoter-bound ChIP-seq signal of these transcripts, GSEA-supplied $P < 0.001$. Heatmaps in **a** and **c** display the $\log_2$ fold change in gene expression versus DMSO for the 14,745 transcripts expressed in Jurkat cells. RT-qPCRs were performed in biological triplicate and error bars are ±s.d. *P* values were determined with a two-tailed Student's *t*-test.

© 2016 Nature America, Inc. All rights reserved.

or off-target activity on other transcription-associated kinases such as JNK1, JNK2 and JNK3 (identified in **Supplementary Table 1** and **Supplementary Data Set 2**) or other unknown off-targets may also account for this discrepancy.

Recent research in both yeast and mammalian systems has identified genetic and physical interactions between the CDK12–cyclin K complex and the DDR pathway. Genetic depletion of the complex reduces the expression of a core set of DDR genes[11,16–19]. Therefore, while the CDK12–cyclin K complex exerts control over Pol II–mediated transcription, its loss appears to have pleiotropic consequences for both transcription and the DDR pathway. THZ531 does indeed downregulate core DDR pathway genes, which may have very important mechanistic and therapeutic implications for CDK12 and CDK13 inhibition. Mechanistically, this suggests a link between the integrity of DNA damage signaling and transcription that is specific to CDK12, as other CDKs that were probed in previous research did not affect DDR gene expression[11,16].

From a therapeutic standpoint, it is attractive to think that an inhibitor, such as THZ531, could target both aberrant transcription and genomic instability, two enabling characteristics that facilitate the acquisition of cancer hallmark capabilities. While both of these characteristics are common features of all cancers, certain cancers may be particularly susceptible to CDK12 and CDK13 inhibition. T-cell acute lymphoblastic leukemia (T-ALL) exploits a variety of genetic perturbations, including chromosomal translocations, deletions, and insertions, to create super-enhancers that drive aberrant expression of oncogenic transcription factors, including TAL1 (refs. 32,35,36). As a single agent, THZ531 would appear to induce significant transcriptional stress that primes cells for apoptotic cell death. In Jurkat T-ALL, the loss of transcription factors that form the transcriptional regulatory core is sufficient to initiate apoptotic signaling[27,32,33]. Breast and ovarian cancers harboring *BRCA1* and *BRCA2* mutations are deficient in homologous recombination (HR), yet they show excellent response to PARP1 inhibitors, resulting from a synthetic lethal relationship between the combined loss of HR and loss of single-strand repair. The loss of *BRCA* expression following CDK12 inhibition would suggest that a similar relationship may hold for CDK12 and PARP1, allowing treatment with potentially low doses of THZ531 in combination with recently clinically approved PARP1 inhibitors. Indeed, recent work in ovarian cancers that harbor CDK12 inactivating mutations shows that attenuation of CDK12 signaling is sufficient to sensitize cells to PARP1 inhibition[21,22]. CDK12 and CDK13 inhibition therefore may hold promise for the treatment of cancers that utilize aberrant transcription or genomic instability to gain a selective growth advantage.

Received 18 August 2015; accepted 13 June 2016;
published online 29 August 2016

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession number 5ACB. Sequencing and expression data were deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE72031 (**Supplementary Table 3**). CDK12 RefSeq Accession NM_016507 was used as repair template in CRISPR studies.

References

1. Young, R.A. RNA polymerase II. *Annu. Rev. Biochem.* **60**, 689–715 (1991).
2. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
3. Fuda, N.J., Ardehali, M.B. & Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature* **461**, 186–192 (2009).
4. Muñoz, M.J., de la Mata, M. & Kornblihtt, A.R. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem. Sci.* **35**, 497–504 (2010).
5. Sansó, M. & Fisher, R.P. Pause, play, repeat: CDKs push RNAP II's buttons. *Transcription* **4**, 146–152 (2013).
6. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
7. Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316 (2010).
8. Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15**, 3319–3329 (2001).
9. Chen, H.H., Wang, Y.C. & Fann, M.J. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol. Cell. Biol.* **26**, 2736–2745 (2006).
10. Chen, H.H., Wong, Y.H., Geneviere, A.M. & Fann, M.J. CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochem. Biophys. Res. Commun.* **354**, 735–740 (2007).
11. Liang, K. *et al.* Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol. Cell. Biol.* **35**, 928–938 (2015).
12. Eifler, T.T. *et al.* Cyclin-dependent kinase 12 increases 3′ end processing of growth factor-induced c-FOS transcripts. *Mol. Cell. Biol.* **35**, 468–478 (2015).
13. Davidson, L., Muniz, L. & West, S. 3′ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev.* **28**, 342–356 (2014).
14. Bartkowiak, B. & Greenleaf, A.L. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J. Biol. Chem.* **290**, 1786–1795 (2015).
15. Ahn, S.H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* **13**, 67–76 (2004).
16. Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **25**, 2158–2172 (2011).
17. Winsor, T.S., Bartkowiak, B., Bennett, C.B. & Greenleaf, A.L. A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II. *PLoS One* **8**, e60909 (2013).
18. Bennett, C.B. *et al.* Yeast screens identify the RNA polymerase II CTD and SPT5 as relevant targets of BRCA1 interaction. *PLoS One* **3**, e1448 (2008).
19. Ekumi, K.M. *et al.* Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res.* **43**, 2575–2589 (2015).
20. Rottenberg, S. *et al.* High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. *Proc. Natl. Acad. Sci. USA* **105**, 17079–17084 (2008).
21. Bajrami, I. *et al.* Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res.* **74**, 287–297 (2014).
22. Joshi, P.M., Sutor, S.L., Huntoon, C.J. & Karnitz, L.M. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. *J. Biol. Chem.* **289**, 9247–9253 (2014).
23. Bartkowiak, B., Yan, C. & Greenleaf, A.L. Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. *Biochim. Biophys. Acta* **1849**, 1179–1187 (2015).
24. Bösken, C.A. *et al.* The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* **5**, 3505 (2014).
25. Greifenberg, A.K. *et al.* Structural and functional analysis of the Cdk13/Cyclin K complex. *Cell Rep.* **14**, 320–331 (2016).
26. Lolli, G., Lowe, E.D., Brown, N.R. & Johnson, L.N. The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12**, 2067–2079 (2004).
27. Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).
28. Patricelli, M.P. *et al.* In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem. Biol.* **18**, 699–710 (2011).
29. Christensen, C.L. *et al.* Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. *Cancer Cell* **26**, 909–922 (2014).
30. Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126–1139 (2014).
31. Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat. Struct. Mol. Biol.* **19**, 1108–1115 (2012).
32. Mansour, M.R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
33. Sanda, T. *et al.* Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
34. Bowman, E.A. & Kelly, W.G. RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus* **5**, 224–236 (2014).
35. Ferrando, A.A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).
36. Brown, L. *et al.* Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. *EMBO J.* **9**, 3343–3351 (1990).

Acknowledgments

We thank R. George, K. Wong, P. Hammerman, J. Bradner and members of the Gray and Young laboratories for helpful discussions. We thank P. Wisniewski and C. Zollo from the Whitehead FACS facility for help with FACS instruments. We thank J. Love, T. Volkert, and S. Gupta from the Whitehead Genome Core for help with genomics experiments. The authors would like to thank Diamond Light Source for beamtime (proposal mx8421), and the staff of Beamline I03 for assistance with data collection. We thank W. Massefski from the Dana–Farber NMR lab for help in collecting NMR data for the manuscript. M.G. is a member of the DFG excellence cluster ImmunoSensation. This work was supported by the National Institutes of Health (HG002668 and CA109901 to R.A.Y. and 5 R01 CA179483-02 C to N.S.G.), the Koch Institute and Dana–Farber/Harvard Cancer Center Bridge Grant (N.K., N.S.G., R.A.Y.), the DFG (GE 976/9-1 to M.G.), the Hope Funds for Cancer Research Grillo-Marxuach Family Fellowship (B.J.A.), and a NDM Prize Studentship, funded in part by the Medical Research Council (S.E.D.-C.). The SGC is a registered charity (no. 1097737) that receives funds from AbbVie, Bayer, Boehringer Ingelheim, Genome Canada through Ontario Genomics Institute Grant OGI-055, GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and Wellcome Trust Grant 092809/Z/10/Z.

Author contributions

N.S.G., R.A.Y., N.K., and T.Z. conceived the project. N.S.G. and T.Z. conceived and directed chemistry effort with input from Y.L. Chemical synthesis and small molecule structure determination was performed by T.Z., M.H., and T.M. R.A.Y., N.S.G., B.J.A. and N.K. conceived genomics effort. N.K. and C.M.O. designed and executed cellular biological experimental research with input from N.S.G. and R.A.Y.S.E.D.-C. solved co-crystal structure of CDK12–cyclin K with THZ531 with guidance from J.M.E. and A.N.B. A.K.G. designed and executed CDK *in vitro* kinase assays with input from M.G. S.B.F. designed and performed protein mass spectrometry on THZ531–CDK12 adducts with input from J.A.M. N.M.H. provided cloning expertise. B.B. and A.L.G. developed and provided CDK12 and CDK13 antibodies. B.J.A. designed and performed genomics data analyses. N.K., T.Z., N.S.G. and R.A.Y. co-wrote the paper. All authors edited the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to N.S.G. or R.A.Y.

© 2016 Nature America, Inc. All rights reserved.

ONLINE METHODS

Chemistry. Chemical synthesis and characterization is provided as a **Supplementary Note**.

Cell culture conditions. Jurkat cell lines (A.T. Look Laboratory, DFCI) were grown in RPMI 1640, Glutamax (Life Technologies) supplemented with 10% FBS (Life Technologies). HAP1 cells (Horizon Discovery Group, cat # C859) were grown in IMDM medium supplemented (Life Technologies) with 10% FBS and 100 U/ml penicillin, 100 mg/ml (Life Technologies). All cell lines were cultured at 37 °C in a humidified chamber in the presence of 5% CO_2, unless otherwise noted. All cells were mycoplasma tested upon initial receipt.

Reagents and antibodies. The following antibodies were used for immunoblots: Pol II CTD Ser2 (cat# 04-1571, 1:1,000), Ser5 (cat# 04-1572, 1:5,000), and Ser7 (cat# 04-1570, 1:1,000) phospho-antibodies (Millipore); Total Pol II (Santa Cruz cat# sc-899, 1:500); CDK7 (Santa Cruz cat# sc-723, 1:500); CDK9 (Bethyl cat# A303-493A, 1:1,000); CDK12 and CDK13 (Arno Greenleaf, 1:1,000); cyclin K (Bethyl cat# A301-939A, 1:500); cyclin H (Bethyl cat# A301-674A, 1:1,000); PARP (Cell Signaling cat# 9542, 1:1,000); and α-Tubulin DM1α (Sigma cat# T9026, 1:5,000). The following ChIP-grade antibodies were used for ChIP–seq (10 μg each): Pol II (Santa Cruz cat# sc-899); CDK12 (gift of Arno Greenleaf); Pol II CTD Ser-2 (Millipore cat# 04-1571); and H3K27Ac (Abcam cat# AB4729).

Constructs. For structural studies, human CDK12 residues 715–1052 (Uniprot Q9NYV4) and human cyclin K residues 11–267 (Uniprot O75909) were cloned into the baculoviral transfer vector pFB-LIC-Bse by ligation-independent cloning. The vector encodes an N-terminal hexahistidine tag and a Tobacco Etch Virus Protease A (TEV) cleavage site. A pFBHTb expression vector containing Cdk-activating kinase (CAK1) from *Saccharomyces cerevisiae* (Uniprot P43568) was purchased from the MRC Protein and Phosphorylation Unit (University of Dundee, UK). For CRISPR studies, guide RNAs (sgRNA) targeting CDK12 were cloned into pX330 (Addgene, cat# 42230). pUC57–AMP was used as storage vector for CDK12 genome reference sequence.

Genome editing. The CRISPR–Cas9 system was used to mutate the endogenous CDK12 WT locus to encode for CDK12 C1039S, a THZ531 refractory mutant of CDK12. Target-specific oligonucleotides were cloned into pX330 (Addgene, cat# 42230), which carries a codon-optimized version of Cas9 and was further modified to express GFP for identifying transfectants. Cells were co-transfected (X-tremeGENE 9 (Roche) with (1) pX330 expressing Cas9 and CDK12 targeting sgRNA, and (2) a pUC57-AMP construct bearing 1,500 bp of modified CDK12 reference genome (RefSeq Accession NM_016507) that is centered around the CRISPR targeting site in CDK12. Two days after transfection, cells were sorted using GFP as a marker of transfected cells and cells were re-plated for five days. Cells were then re-plated at low density to facilitate the isolation of individual clones. Individual clones were isolated, expanded, and PCR genotyped using mutant specific PCR primers. Following initial PCR screening, individual clones were Sanger sequenced to confirm the presence of the desired mutation. Western blot analysis confirmed the presence of intact CDK12 kinase. Subsequent experiments were conducted using a CDK12 C1039S clone and a WT control clone that was carried through the entirety of the CRISPR protocol but that was verified by Sanger sequencing to be WT CDK12.

The genomic sequence complementary to the CDK12-directed guide RNA that was cloned into pX330 and used in the genome editing experiments is: GGCAGGATTGCCATGAGTTG.

The modified genomic sequence that was cloned into pUC57–AMP by Genewiz and used as the repair template for genome editing is found in **Supplementary Figure 11**.

The primers used for PCR based diagnostic sequencing were the following:

1. GGATTGCCATGAGTTGTGGAGTA (specific to WT CDK12 allele)
2. AGACTCCCACGAGCTCTGGAGTA (specific to C1039S mutant allele)
3. TGCTTTCGCTTAAGAAATACAGG (anchored in the genome and amplifies both WT and mutant alleles)

The primers used for PCR amplification and Sanger sequencing were the following:

1. GCTTTAATGTCAAATTTTGGAGGTC
2. CTGGCACTGACCTATTGCAT
3. GGCTTGTTAATGTATACAGAAGGTCG
4. GATGCAATAGGTCAGTGCCAGAATGG
5. TGGCAGATGGTAGTCTGAGGA

***In vitro* kinase assays.** Recombinant kinases CDK9–cyclin T1, CDK12–cyclin K and CDK13–cyclin K were prepared from baculovirus-infected insect cells similarly as described[24,37]. Radioactive kinase activity measurements were performed at a concentration of 0.2 μM CDK–cyclin complex. Recombinant CDK7–cyclin H-MAT1 was purchased from EMD Millipore (cat#14-476). Typically, 35 μl reaction volumes of 0.2 μM active kinase were equilibrated in kinase buffer (40 mM Hepes (pH 7.6)), 34 mM NaCl, 34 mM KCl, 10 mM $MgCl_2$, 5% glycerol, and 1× PhosSTOP (Roche). Cold ATP to a final concentration of 200 μM and 3 μCi [γ-^{32}P]ATP (PerkinElmer) and 50 μM of a substrate peptide containing five phosphorylation sites were added and the reaction mixture incubated for 30 min at 30 °C at 350 rpm in a thermomixer (Eppendorf). Reactions were stopped by adding EDTA to a final concentration of 50 mM. Aliquots of 15 μl each were spotted onto paper squares using the Optitran BA-S85 reinforced membrane. Paper squares were washed three times for 5 min with 0.75% (v/v) phosphoric acid, with at least 5 ml washing solution per paper. Radioactivity was counted in a Beckman Scintillation Counter (Beckman Coulter) for 1 min. Measurements were performed in triplicate and are represented as mean with s.d. For experiments studying the effect of an upstream pre-incubation time, compounds THZ531 or THZ531R were added at varying concentrations to 0.2 μM CDK–cyclin complex and incubated for varying times from 1 to 540 min at 30 °C and 350 rpm before ATP and substrate were added to the reaction mix.

Inhibitor treatment experiments. Cells were treated with THZ531, THZ531R, or THZ532 for 6 h at the indicated concentrations unless otherwise noted. For inhibitor washout experiments (**Fig. 3b**) cells were treated with compounds for 6 h. Medium containing inhibitors was subsequently removed to effectively 'wash out' the compound and the cells were allowed to grow in the absence of inhibitor for an additional 66 h (total time = 72 h).

Immunoblotting. Whole-cell lysates for immunoblotting were prepared by pelleting cells from each cell line at 4 °C (1,200 rpm) for 5 min using a Sorvall Legend centrifuge (ThermoFisher). The resulting cell pellets were washed 1× with ice-cold 1× PBS and then resuspended in lysis buffer containing 50 mM TrisHCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.1% SDS, 5 mM EDTA with protease (Roche) and phosphatase inhibitors (Roche). Whole-cell lysates were collected and snap-frozen in liquid nitrogen before being stored at –80 °C. Protein concentrations were determined by using the Bio-Rad DC protein assay kit (Bio-Rad) or Pierce BCA protein assay kit (Life Technologies). Whole-cell lysates were loaded into Bolt 4–12% or 8% bis-tris gels (Invitrogen) and separated by electrophoresis at 90 V for 2 h. The gels were then transferred onto nitrocellulose membrane (Bio-rad) and blocked by incubation with 5% dry milk or BSA in TBST (TBS with 0.2% Tween-20). Membranes were probed using antibodies raised against the indicated proteins. Appropriate IR-labeled secondary antibodies were used and Licor Odyssey CLx was used for detection.

Pull down/Immunoprecipitation (IP) experiments. Cells were treated with THZ531, THZ531R, or DMSO for 6 h. Following treatment, cells were washed twofold with cold PBS and then lysed in the following lysis buffer: 50 mM Hepes pH 7.4, 150 mM NaCl, 1% Nonidet P40 substitute, 5 mM EDTA, 1mM DTT, and protease/phosphatase cocktails. Following clearance, lysates were treated with bio-THZ1 or bio-TH531 for pulldown overnight at 4 °C. Lysates were further incubated at room temperature for 3 h to increase the efficiency of covalent bond formation. Lysates were then incubated with streptavidin agarose (Thermo scientific, cat. #20349) for pulldown for an additional 2–3 h at 4 °C. Agarose beads were washed 5 times with lysis buffer and then boiled in 2× LDS for 10 min at 95 °C. SDS–PAGE resolved precipitated proteins were probed for the indicated proteins.

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/nchembio.2166

Proliferation assays. Proliferation assays were conducted using Cell Titer Glo assay kit (Promega cat# G7571). Jurkat cells were plated in 96-well plates at 20,000 cells/well in fresh media and treated with THZ531, THZ531R, THZ532, or DMSO at the indicated concentrations for 72 h. HAP1 cells were seeded in 96-well plates at 12,000 cells/well in fresh media and 24 h later were treated with THZ531 at the indicated concentrations for 72 h. Anti-proliferative effects of compounds were assessed using Cell Titer Glo as described in product manual by luminescence measurements on a Tecan Safire plate reader. To assess the effect of inhibitor washout on anti-proliferation of Jurkat cells, cells were treated with THZ531, THZ531R, THZ532, or DMSO for 6 h. Inhibitor-containing medium was then removed and incubated with or without inhibitors for 66 h. Anti-proliferative effects of compounds were assessed using Cell Titer Glo as described in product manual by luminescence measurements on an Envision 2104 Multilabel Reader. All proliferation assays were performed in biological triplicate and error bars are +/– s.d. IC_{50}s were determined using GraphPad Prism 6 nonlinear regression curve fit.

Fluorescence-activated cell sorting analysis (FACS). For cell cycle analysis, cells were treated with inhibitors for indicated time periods. Cells were collected by centrifugation, washed once in ice-cold phosphate buffered saline (PBS), and fixed overnight at –20 °C with 80% ethanol in PBS. Cells were washed three times with PBS. Finally, cells were resuspended in PBS containing 0.1% Triton X-100, 25 μg ml^{-1} propidium iodide (PI, Molecular Probes), and 0.2 mg ml^{-1} RNase A (Sigma) and incubated for 45 min at 37 °C. For discrimination of apoptotic vs. non-apoptotic cells by annexin V/PI staining, cells were treated with compound for indicated periods of time. Cells were collected by centrifugation, washed once with PBS, and processed according to the manufacturer's protocol (Invitrogen cat# V13242). All FACS samples (cell cycle distribution or PI/annexin V double staining) were analyzed on a FACS Canto II (BD Biosciences) instrument and processed on FlowJo (Treestar). All PI/annexin V experiments were performed in biological triplicate. Error bars are +/– s.d.

HPLC/mass spectrometry. Recombinant CDK12–cyclin K complex was incubated with a fivefold excess of THZ531 or DMSO for 1 h at room temperature and analyzed by LC-ESI-MS essentially as described[38]. In each analysis, 5 μg protein was injected onto a self-packed reversed phase column (1/32" O.D. × 500 um I.D., 5 cm of POROS 10R2 resin). After desalting for four minutes, protein was eluted with an HPLC gradient (0–100% B in 4 min, A = 0.2M acetic acid in water, B = 0.2 M acetic acid in acetonitrile, flow rate = 10 μL/min) into an LTQ ion trap mass spectrometer (ThermoFisher). Mass spectra were deconvoluted using MagTran1.03b2 software[39]. To determine the site of modification, proteins were reduced with TCEP (10 mM final concentration), alkylated with iodoacetamide (22.5 mM final concentration) and digested with glu-C (37 °C, overnight) and analyzed by nanoLC-MS[40]. Digested peptides were injected onto the precolumn (4 cm POROS 10R2, Applied Biosystems) and eluted with an HPLC gradient (NanoAcquity UPLC system, Waters, Milford, MA; 10–70% B in 60 min; A = 0.1 M acetic acid in water, B = 0.1M acetic acid in acetonitrile). Peptides were resolved on a self-packed analytical column (50 cm Monitor C18, Column Engineering, Ontario, CA) and introduced to the mass spectrometer (LTQ Orbitrap XL) at a flow rate of ~30 nL/min (ESI spray voltage = 3.2 kV). The mass spectrometer was programmed to perform data-dependent MS/MS on the five most abundant precursors (35% collision energy) in each MS1 scan (image current detection, 30K resolution, m/z 300–2000). MS/MS spectra were matched to peptide sequences using Mascot (version 2.2.1) after conversion of raw data to .mgf using multiplierz scripts[41]. Search parameters specified glu-C digestion with up to 2 missed cleavages, fixed carbamidomethylation of cysteine residues, as well as variable oxidation of methionine and THZ531 modification of cysteine (i.e. cysteine residues are either carbamidomethylated or THZ531 modified). Precursor and peptide ion mass tolerances were 10 ppm and 0.6 Da, respectively.

Protein expression and purification. Bacmid DNA was prepared in *Escherichia coli* strain DH10Bac and used to generate baculovirus in Sf9 insect cells. Baculoviruses for CDK12, cyclin K and yeast CAK1 were used to co-infect Sf9 cells grown in suspension to a density of 2 × 10^6 cells/mL. Cells were lysed by sonication in binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM Imidazole) supplemented with protease inhibitor cocktail set III (Calbiochem). Polyethylenimine (PEI) was then added to a final concentration of 0.5% to precipitate DNA. The lysate was clarified by centrifugation at 21,000 rpm for 1 h at 4 °C. Recombinant proteins were purified using nickel sepharose resin (GE Healthcare) and eluted stepwise with imidazole. Fractions containing the phosphorylated CDK12–cyclin K complex were treated with TEV protease for cleavage of the N-terminal hexahistidine tags. The cleaved proteins were then exchanged into a buffer containing 50 mM MES pH 6.5, 0.5 mM TCEP and loaded onto a 5 mL HiTrap SP column (GE Healthcare) equilibrated in the same buffer. A linear elution gradient was run from 0 to 1 M NaCl. Fractions containing the CDK12 complex were pooled and the buffer adjusted to 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP during concentration in a 10 kD MWCO Amicon Ultra concentrator.

Crystallization. The purified CDK12–cyclin K complex at 17 μM concentration was incubated overnight at 4 °C with 25 μM THZ531. Near-complete covalent binding was shown by a mass shift of 559 Da using mass spectrometry. Excess free compound was removed by size-exclusion chromatography using a HiLoad Superdex S75 26/60 column (GE Healthcare). The protein–inhibitor complex was concentrated to 5.5 mg/mL buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP. Crystals were grown at 20 °C in 195 nL sitting drops mixing 75 nL protein solution and 20 nL crystal seed stock with 100 nL of a reservoir solution comprising 0.1 M HEPES pH 7.0, 10% PEG8000, 0.2 M magnesium chloride. Crystals were cryoprotected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen after mounting.

Structure determination. Diffraction data were collected at a wavelength of 0.97625 Å at 100 K from a single crystal on Diamond Light Source beamline I03. Data were indexed and integrated using XDS[42] and scaled using AIMLESS[43] in the CCP4 suite of programs[44]. Phases were found using molecular replacement in PHASER[45]. The structure of the $CDK12^{715-1038}$/cyclin K^{11-267} complex (PDB: 4UN0) was used as a search model, with input of CDK12 and cyclin K done separately. A PDB file for the inhibitor, THZ531, was created using PRODRG in the CCP4 suite. Models were built initially using COOT[46] and then refined and modified using alternate rounds of REFMAC5 (ref. 47) and COOT. The refined structures were validated using MolProbity[48] and the atomic coordinate files deposited in the Protein Data Bank. A Ramachandran plot indicated good stereochemistry of the three-dimensional structure with all residues falling within the favored (96%) and allowed (4%) regions. Structure figures were prepared with PyMOL[49].

Mass spectrometry for crystallography. Protein masses were determined using an Agilent LC/MSD TOF system with reversed-phase high-performance liquid chromatography coupled to electrospray ionization and an orthogonal time-of-flight mass analyzer. Proteins were desalted before mass spectrometry by rapid elution off a C3 column with a gradient of 5–95% isopropanol in water with 0.1% formic acid. Spectra were analyzed using the MassHunter software (Agilent).

RNA extraction and synthetic RNA spike-in. Total RNA and sample preparation was performed as previously described[50]. Briefly, Jurkat cells were incubated in media containing THZ531, THZ531R, or THZ532 at the indicated concentrations or with DMSO for the specified duration of time. Cell numbers were determined by manually counting cells using C-Chip disposable hemocytometers (Digital Bio, DHC-N01) before lysis and RNA extraction. Total RNA from biological replicates (equivalent to 5 million cells per replicate) was subsequently isolated using RNeasy Plus Mini kit (Qiagen) following the manufacturer's instructions and re-suspended in 50 μL nuclease-free water (Ambion). Total RNA was spiked in with ERCC RNA Spike-In Mix (Ambion, cat# 4456740) and analyzed on Agilent 2100 Bioanalyzer for integrity. RNA with the RNA Integrity Number (RIN) above 9.8 was hybridized to GeneChip PrimeView Human Gene Expression Arrays (Affymetrix).

cDNA preparation and TaqMan expression analysis. RNA used for RT-PCR was extracted as outlined above. One microgram of purified RNA was reverse

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/nchembio.2166

transcribed using Superscript III First-Strand (Invitrogen) with oligo dT primers to prime first-strand synthesis according to the manufacturer's protocol. qPCR was carried out on the 7000 ABI Detection System using the following Taqman probes according to the manufacturer's protocol (Applied Biosystems). All experiments shown were performed in biological triplicate. Each individual biological sample was qPCR-amplified in technical duplicate. Error bars are ± s.d. Expression was normalized to GAPDH, and fold change in expression was calculated relative to the indicated conditions. The following Taqman probes from Life Technologies were used for qPCR-based gene expression analysis: *RUNX1* - Hs00231079_m1; *TAL1* - Hs01097987_m1; *GATA3* - Hs00231122_m1; *MYB* - Hs00920556_m1; *GAPDH* - Hs02758991_g1; *BRCA1* - Hs01556193_m1; *ERCC4* - Hs00193342_m1; *TERF2* - Hs00194619_m1; FANCF - Hs00256030_s1; *PARP9* - Hs00967084_m1; *MDC1* - Hs00206182_m1; *RAG2* - Hs01851142_s1; *RAD51C* - Hs00427442_m1; and *ATR* - Hs00992123_m1.

Microarray sample preparation and analysis. For microarray analysis, 100 ng of total RNA containing ERCC RNA Spike-In Mix (see 'RNA extraction and synthetic RNA spike-in' above) was used to prepare biotinylated aRNA (or cRNA) according to the manufacturer's protocol (30 IVT Express Kit, Affymetrix 901228). Briefly, total RNA undergoes T7 oligo(dT)-primed reverse transcription to synthesize first-strand cDNA containing a T7 promoter sequence. This cDNA is then converted into a double-stranded DNA template for transcription using DNA Polymerase and RNase H to simultaneously degrade the RNA and synthesize second strand cDNA. *In vitro* transcription synthesizes aRNA and incorporates a biotin-conjugated nucleotide. The aRNA is then purified to remove unincorporated NTPs, salts, enzymes, and inorganic phosphate. Fragmentation of the biotin-labeled aRNA prepares the sample for hybridization onto GeneChip 3′ expression arrays. Samples were prepared for hybridization using 10 µg of biotinylated aRNA in a 1× hybridization cocktail according the Affymetrix hybridization manual. Additional hybridization cocktail components were provided in the Affymetrix GeneChip Hybridization, Wash and Stain Kit. GeneChip arrays (Human PrimeView, Affymetrix 901837) were hybridized in a GeneChip Hybridization Oven at 45 °C for 16 h at 60 rpm. Washing was done using a GeneChip Fluidics Station 450 according to the manufacturer's instructions, using the buffers provided in the Affymetrix GeneChip Hybridization, Wash and Stain Kit. Images were extracted with Affymetrix GeneChip Command Console (AGCC) and analyzed using GeneChip Expression Console. A PrimeView CDF that included probe information for the ERCC controls (GPL16043), provided by Affymetrix, was used to generate .CEL files.

Batches of microarray experiments were normalized within themselves to ERCC spike-in probes as described in refs. 27,50. Briefly, probes were collapsed into probesets using *expresso* with mas5 normalization. Probeset expression profiles were normalized using *loess* from the *affy* R package to equilibrate ERCC spike-in probes across microarrays in a given batch. Where possible, log2 values of biological replicates were averaged. Fold changes were taken by subtracting average log2 DMSO signal from average log2 treatment signal. Expressed genes were those with log2(expression) > log2(100) in the corresponding DMSO sample. Sensitive genes were defined as those that decreased > = 2 log2-fold (4 linear fold-change). Expression heatmaps (**Fig. 5c**,**d** and **Supplementary Fig. 6b**) were created using the fold change of expressed genes and plotted using heatmap.2 ranked by the values noted in the legend. Expression barplots (a.k.a. waterfall plots; **Fig. 6a**,**c**) were created by ranking expressed transcripts by their fold change and plotting the fold change of each as a bar plot.

PrimeView microarrays were submitted in the following batches used for normalization:

- HAP1 –WT DMSO, WT THZ531 300 nM, WT THZ531 900 nM, C1039S CDK12 - DMSO, C1039S CDK12 - THZ531 300 nM, C1039S CDK12 - THZ531 900 nM
- Jurkat – 6 h DMSO (2 replicates), 6 h THZ531 50 nM (2 replicates), 6 h THZ531 200 nM (2 replicates), 6 h THZ531 350 nM (2 replicates), 6 h THZ531 500 nM (2 replicates), THZ531R 500 nM (2 replicates), THZ532 500 nM (2 replicates)

Gene set enrichment analysis (***Fig. 6g*** *and* ***Supplementary Fig. 9c***). Gene Set Enrichment Analysis[51] was performed using a gene list pre-ranked by CDK12 ChIP–seq signal at promoters or enhancers. The top 500 genes expressed in 6h DMSO from the same microarray batch as 6h 200 nM THZ531 and most downregulated by 6h 200 nM THZ531 treatment were used as the gene set.

Functional enrichment analysis using DAVID (**Supplementary Fig. 8a**,**c**). Gene transcripts that were sensitive to THZ531 (50 or 200 nM) were used as input in the functional analysis tool at http://david.abcc.ncifcrf.gov/[52]. GO terms enrichment was performed using the default settings of the DAVID tool V6.7.

Chromatin immunoprecipitation. Cells were crosslinked for 20 min at room temperature by the addition of one-tenth of the volume of 11% formaldehyde solution (11% formaldehyde, 50 mM HEPES pH 7.3, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0) to the growth media followed by 5 min quenching with 100 mM glycine. Cells were washed twice with PBS, then the supernatant was aspirated and the cell pellet was flash frozen in liquid nitrogen. Frozen crosslinked cells were stored at –80 °C. 50 µL of Dyna magnetic beads (Sigma) were blocked with 0.5% BSA (w/v) in PBS. Magnetic beads were bound with 10 µg of the indicated antibody. For RNA polymerase II-occupied genomic regions, we performed ChIP–Seq experiments using a SantaCruz Biotechnology (sc-899, lot C1413) antibody. The affinity-purified antibody was raised in rabbit against an epitope mapping to the N terminus of murine RBP1, the largest subunit of RNA Pol II. For CDK12 occupied genomic regions, we performed ChIP–Seq experiments using an antibody shared with us from Arno Greenleaf[14]. For pSer2 Pol II occupied genomic regions, we performed ChIP–seq experiments using a Millipore (04-1571) antibody. For H3K27Ac-occupied genomic regions, we performed ChIP–Seq experiments using an Abcam (AB4729A) antibody. The affinity-purified antibody was raised in rabbit against an epitope corresponding to amino acids 1–100 of human Histone H3 that is acetylated at K27. For Jurkat, crosslinked cells were lysed with lysis buffer 1 (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100), pelleted and resuspended in lysis buffer 2 (10 mM TrisHCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA). The subsequent pellet was resuspended in and sonicated in sonication buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA, 0.1% Na-deoxycholate, 0.1% SDS, and 1% Triton X-100). Cells were sonicated for 10 cycles at 30 s each on ice (18–24 W) with 60 s on ice between cycles. Sonicated lysates were cleared and incubated overnight at 4 °C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed two times with sonication buffer, one time with sonication buffer with 500 mM NaCl, one time with LiCl wash buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and one time with TE. DNA was eluted in elution buffer (50 mM TrisHCl pH 8.0, 10 mM EDTA, 1% SDS). Cross-links were reversed overnight. RNA and protein were digested using RNase A and Proteinase K, respectively and DNA was purified with phenol chloroform extraction and ethanol precipitation.

ChIP–seq analysis. Illumina sequencing libraries were generated and data was processed according to ref. 53. In brief, libraries were generated for ChIP samples following the Illumina TruSeq™ DNA Sample Preparation v2 kit protocol with minor changes. All ChIP–Seq data sets were aligned using bowtie[54] to build NCBI37/hg19 of the human genome with parameters -p 4–best -k 2 -m 2–sam -l 40. Wiggle files for gene tracks were created using Macs 1.4.2 (ref. 55) with options –w –S –space = 50 –shiftsize = 200 –nomodel to count extended reads in 50-bp bins, were divided by the number of considered reads to normalize to mapped reads per million, and were displayed in the UCSC genome browser (**Figs. 5f**, **6f** and **Supplementary Figs. 5a**, **7a**,**b**, and **9a**).

Enhancer identification. Enhancers (**Figs. 4a**,**d** and **6e**,**f** and **Supplementary Figs. 8e**,**f** and **9a**–**c**) were identified as having a significant enrichment in ChIP–Seq reads for H3K27ac as determined by MACS 1.4.2 (ref. 55) with input control and parameters –keep-dup = auto –p 1e-9.

Metagene analysis (***Figs. 4a**,**b** and **5b** and* ***Supplementary Fig. 7b***). Metagenes were created to summarize the binding activities of proteins at particular genomic region types. For promoter and enhancer metagenes, the reads per million (RPM)-normalized (-r) ChIP–Seq read counts in 150 equally sized bins (-m 150) were calculated using bamToGFF

© 2016 Nature America, Inc. All rights reserved.

(https://github.com/BradnerLab/pipeline), and the mean for each bin across all regions was plotted. Enhancer metagenes (**Fig. 4a**) used a 4 kb window centered on the center of MACS-defined H3K27ac peaks. Promoter metagenes (**Fig. 4a**) used a 4 kb window centered on the transcription start sites defined by RefSeq[56]. Gene body metagenes (**Figs. 4b** and **5b**, **Supplementary Fig. 7b**) were created in three parts: 2 kb upstream of the transcription start site (TSS) to TSS; TSS to transcription end site (TES); TES to 2 kb downstream of the TES. 50 bins divided the gene-flanking regions; 150 bins divided the genic regions. Strandedness was preserved. Only genes whose transcripts were considered expressed in the DMSO condition calculated in the THZ531 treatment batch were considered.

ChIP–Seq heatmap analysis (***Fig. 4c**,**d***). Heatmaps were created to show per-region occupancy of proteins. The RPM-normalized (-r) ChIP–Seq read counts (-d) in 150 equally sized bins (-m 150) dividing 4 kb regions (±2 kb from transcription start sites or ±2 kb from the center of an H3K27ac peak) was calculated using bamToGFF (https://github.com/BradnerLab/pipeline) and plotted using heatmap.2 in R.

Correlation lineplots (***Fig. 4c**,**d***). To correlate the presence of CDK12, H3K27ac, and RNA Pol II at enhancers and promoters, the RPM-normalized read counts from ChIP–Seq experiments targeting these factors in these regions was calculated using bamToGFF with one window for the entire region (-m 1). Promoters and enhancers were separately ranked by RNA Pol II read counts and CDK12 and H3K27ac read counts were displayed using rollmean from the zoo package in R with a width of 200 values. The Pol II heatmap was constructed using heatmap.2 on the Pol II read counts.

Region boxplots (***Fig. 5g**,**h***). Boxplots were created to show the effects of different transcriptional inhibitors on Pol II distributions across gene bodies. Promoter boxplots refer to a 4 kb window centered on the transcription start site; TES boxplots refer to a 4 kb window centered on the transcription end site, both defined by RefSeq. The RPM-normalized (-r) read counts from ChIP–Seq experiments targeting RNA Pol II in these regions was calculated using bamToGFF with one window for the entire region (-m 1). Boxplots display these normalized read counts across all regions.

Super-enhancer identification (***Fig. 6e**,**f**, **Supplementary Figs. 8e**,**f** and **9a**,**b***). Super enhancers were identified as described in refs. 32,57,58 using the ROSE algorithm with a stitching distance of 12.5 kb. Briefly, Macs peaks of H3K27ac were used to identify constituent enhancers. H3K27ac signal (less input control) was used to rank enhancers by their enrichment. In Jurkat cells, 818 super-enhancers were separated from typical enhancers using ROSE. Super-enhancers were assigned to the expressed gene whose TSS falls nearest to the center of the super-enhancer. Expressed genes for enhancer assignment were determined by interrogating a 1 kb region centered on RefSeq transcription start sites for RPM-normalized (-r) H3K27ac read density (-d) using bamToGFF with one bin (-m 1). The top two-thirds of genes were considered expressed.

Super-enhancer analysis (***Supplementary Fig. 9b***). To determine if super-enhancers and their associated genes are particularly enriched in CDK12, we interrogated the set of stitched enhancer regions for CDK12 ChIP–Seq read counts and 37973 4 kb promoters centered on RefSeq transcription start sites. bamToGFF was used with the –t option to count reads per region.

Data presentation and statistics. All the results presented in graphics are reported as the mean ± s.d. unless otherwise noted. All statistical analyses herein represent comparison of continuous variables performed using an unpaired (two-tailed) *t*-test unless otherwise noted. All graphics and statistics were generated using GraphPad Prism, Microsoft Excel, or R.

37. Czudnochowski, N., Bösken, C.A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, 842 (2012).
38. Zhang, T. *et al.* Discovery of potent and selective covalent inhibitors of JNK. *Chem. Biol.* **19**, 140–154 (2012).
39. Zhang, Z. & Marshall, A.G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225–233 (1998).
40. Ficarro, S.B. *et al.* Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal. Chem.* **81**, 3440–3447 (2009).
41. Askenazi, M., Parikh, J.R. & Marto, J.A. mzAPI: a new strategy for efficiently sharing mass spectrometry data. *Nat. Methods* **6**, 240–241 (2009).
42. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
43. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
44. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
45. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
47. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
48. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
49. The PyMOL Molecular Graphics System (Version 1.2r3pre, Schrödinger, LLC).
50. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
51. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
52. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
53. Lin, C.Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
55. Zhang, Y. *et al.* Model-based analysis of ChIP–Seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Pruitt, K.D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756–D763 (2014).
57. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
58. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/nchembio.2166

Supplementary Information

Covalent targeting of remote cysteine residues to develop CDK12 and 13 inhibitors

Authors: Tinghu Zhang[1,2,10], Nicholas Kwiatkowski[1,2,3,10], Calla M Olson[1,2,10], Sarah E Dixon-Clarke[5], Brian J Abraham[3], Ann K Greifenberg[6], Scott B Ficarro[1,2,7], Jonathan M Elkins[5], Yanke Liang[1,2], Nancy M Hannett[3], Theresa Manz[1,8], Mingfeng Hao[1,2], Bartlomiej Bartkowiak[9], Arno L Greenleaf[9], Jarrod A Marto[1,2,7], Matthias Geyer[6], Alex N Bullock[5], Richard A Young[3,4*], Nathanael S Gray[1,2*]

Affiliations: [1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [3]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Structural Genomics Consortium, University of Oxford, Old Road Campus, Oxford OX3 7DQ, UK. [6]Department of Structural Immunology, Institute of Innate Immunity, University of Bonn, 53127 Bonn, and Center of Advanced European Studies and Research, 53175 Bonn, Germany. [7]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [8]Pharmaceutical and Medicinal Chemistry, Department of Pharmacy, Saarland University, 66123 Saarbrücken, Germany. [9]Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA. [10]These authors contributed equally to this work.

*To whom correspondence may be addressed: Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Ave., Boston, MA 02115. Tel.: 617-582-8590; Fax: 617-582-8615; E-mail: Nathanael_Gray@ dfci.harvard.edu or Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Tel.: 617-258-5218; E-mail: young@wi.mit.edu.

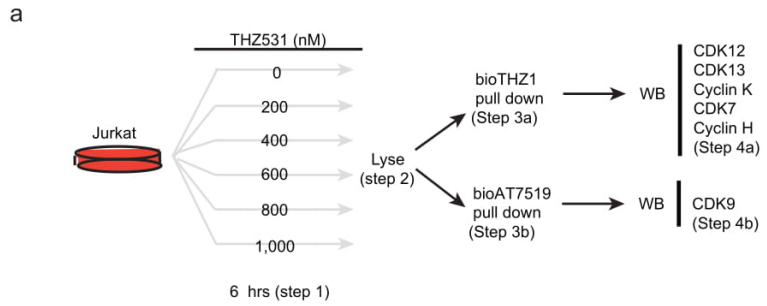
Nature Chemical Biology: doi:10.1038/nchembio.2166

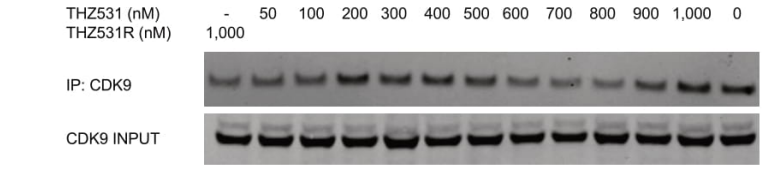
Supplementary Results

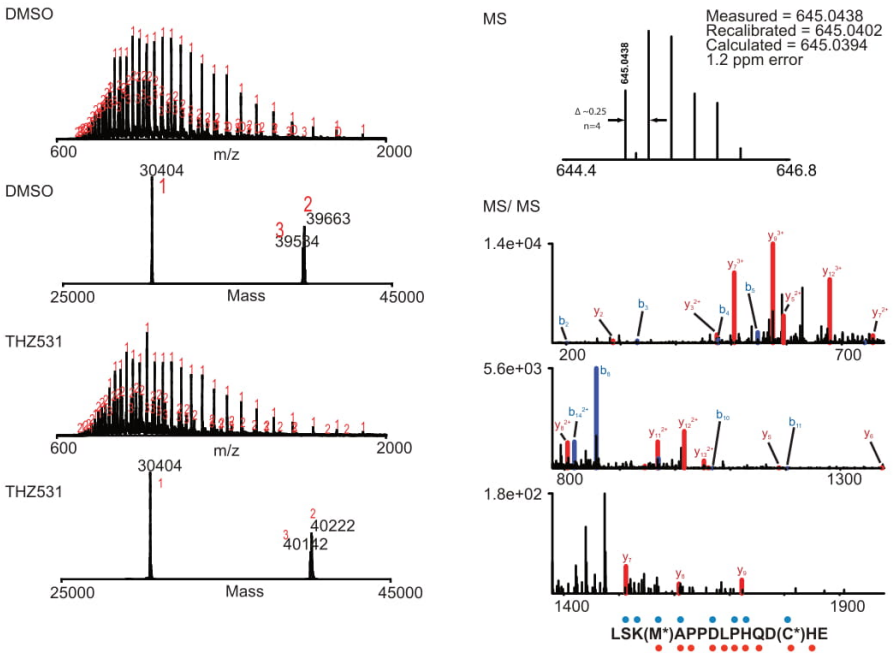
Supplementary Figures

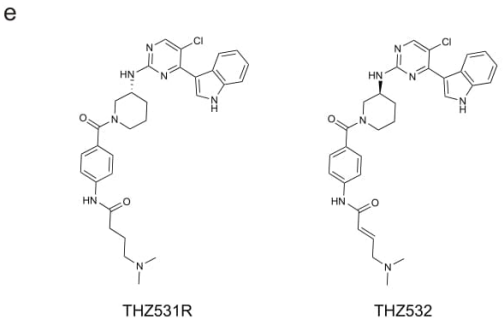
Supplementary Figure 1 | THZ531 targets CDK12 and 13 by covalent modification of distal C-terminal cysteine residues. **a.** Schematic of target engagement experiment. Jurkat cells were treated with increasing doses of THZ531 or DMSO for 6 hrs (Step 1). Cellular lysates were made from cells from each DMSO or THZ531 concentration point (Step 2). Clarified lysates from each treatment condition were then incubated with either 1μM bioTHZ1 (Step 3a), a concentration that binds CDK7-cyclin H, CDK12-cyclin K, and CDK13-cyclin K complexes or 1 μM bioAT7519 (Step 3b) a concentration known to bind CDK9. Subsequent addition of streptavidin -coated beads permits the immunoprecipitation of the indicated protein complexes. Following washing of beads with lysis buffer, the immunoprecipitated proteins were eluted from the beads by boiling in SDS buffer. Western blotting of precipitated proteins for CDK12, CDK13, or cyclin K was used to identify precipitated CDK12-cyclin K or CDK13-cyclin K complexes (Step 4a). Western blotting of precipitated proteins for CDK7 or cyclin H was used to identify precipitated CDK7-cyclin H complexes (Step 4a). And finally, CDK9 western blotting was used to identify precipitated CDK9 complexes (Step 4b). As THZ531 binds to its intended targets covalently, pretreatment of cells with THZ531 parent compound would be expected to block subsequent capture and immunoprecipitation of CDK12, 13, and 7 complexes with bioTHZ1 (or CDK9 with bioAT7519). Therefore, treating cells with THZ531 in dose titration permits us to ascertain at what concentration THZ531 is binding to each of these kinase complexes in cells, giving us a readout of intracellular selectivity. Uncut western blots are in Supplementary Fig. 10. **b.** THZ531 does not bind intracellular CDK9 complexes. THZ531 does not compete with lysate – introduced bioAT7519 for binding to CDK9. **c,** Mass spectra (top, middle bottom) and zero-charge mass spectra (middle top, bottom) of CDK12- cyclin K complex that were treated with DMSO (top, middle top) or THZ531 (middle bottom, bottom) for 1 hr at room temperature. After covalent bond formation, the masses of CDK12 (3) and phosphorylated CDK12 (2) increase by ~558 Da (THZ531). Although cyclin K (1) contains 7 cysteine residues it does not exhibit a mass shift indicating THZ531 does not form a covalent bond with this protein. **d,** Mass spectrum (top) and MS/MS spectrum (bottom) recorded during nanoLC-MS analysis of glu-c digested CDK12-cyclin K complex after treatment with THZ531 for 1 hr at room temperature illustrate detection of precursor (top) and product (bottom) ions of the peptide LSK(M*)APPDLPHQD(C*)HE (CDK12 residues 1026-1041). Ions y2 and y3 indicate C1039 forms a covalent bond with THZ531. Blue and red dots next to the sequence highlight detected ions of type b and y, respectively. (C*), THZ531 labeled cysteine; (M*), oxidized methionine. **e,** Structures of THZ531R and THZ532.

Nature Chemical Biology: doi:10.1038/nchembio.2166







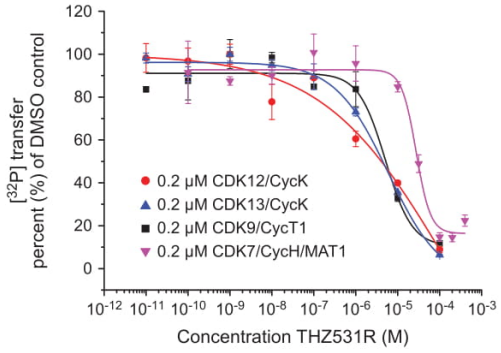


Nature Chemical Biology: doi:10.1038/nchembio.2166

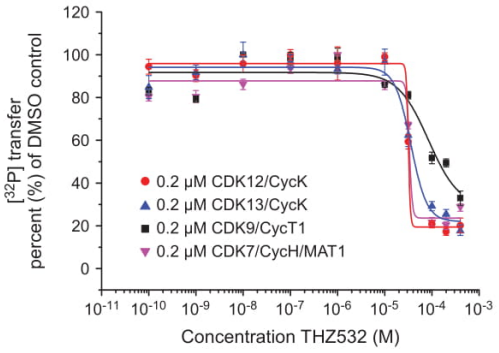
Supplementary Figure 2 | Time series of transcription kinase inhibition at different pre-incubation times and varying inhibitor concentrations. **a,** Kinase inhibition assay using THZ531R, the reversible analog of THZ531, resulted in similarly high IC_{50} values against CDK12, CDK13, CDK9, and CDK7. Measurements were made in triplicate and data represent the mean ± S.D. **b,** Kinase inhibition assays using THZ532, the inactive enantiomer of THZ531, produced similarly high IC_{50} values against CDK12, CDK13, CDK9, and CDK7. Measurements were made in triplicate and data represent the mean values ± S.D. **c,** THZ531 exhibits reduced activity against CDK7. Measurements were made in triplicate and data represent the mean values ± S.D. **d,** THZ531 and THZ531R do not inhibit Erk1. *In vitro* kinase activity assays using recombinant protein were applied to analyze the effect of the THZ531 compound on a member of the MAP kinase family. Measurements were made in triplicate and data represent the mean values ± S.D. **e,** Assay schematic: To a concentration of 0.2 μM CDK- cyclin complex different concentrations of THZ531 were added, ranging from 0.001 μM to 100 μM. Pre-incubation times of 1 min to 9 hrs were followed before the kinase reaction was started by addition radioactively labeled ATP and substrate peptide. A kinase reaction time course of 30 min was applied before the reaction was stopped and the kinase activity measured. The incubation time and the kinase activity time course were performed at 30°C at 350 r.p.m. Measurements were made in triplicate and data represent the mean values ± S.D. **f,** *In vitro* kinase activity assay of 0.2 μM CDK9- cyclin T1 after different preincubation times with varying concentrations of THZ531. Increasing concentrations of THZ531 do not result in significantly reduced kinase activity at longer pre-incubation times, supporting that the CDK9–THZ531 interaction is reversible. As control, the decrease of kinase activity in the absence of THZ531 was measured to monitor the loss of enzymatic activity over time. The counts per minute of the kinase activity measurements were normalized to the relative [^{32}P] transfer. Measurements were made in triplicate and data represent the mean values ± S.D. **g,** THZ531 has lower affinity for PLK1, Aurora A and Aurora B, compared to CDK12/13. *In vitro* kinase assays were performed by Life Technologies in duplicate at an ATP concentration = K_m for each kinase. Data represent the mean of values ± S.D.

Nature Chemical Biology: doi:10.1038/nchembio.2166

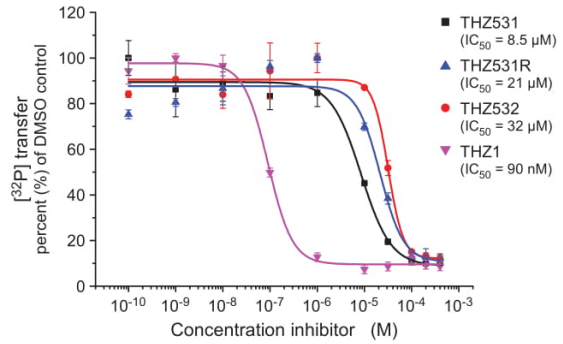
a



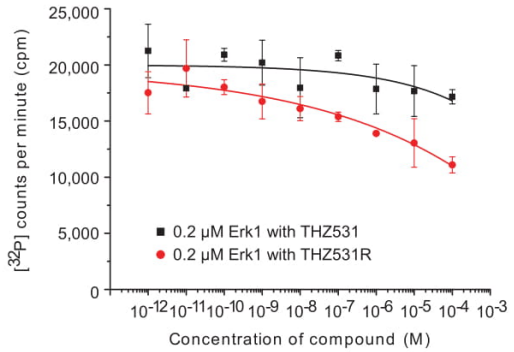
b



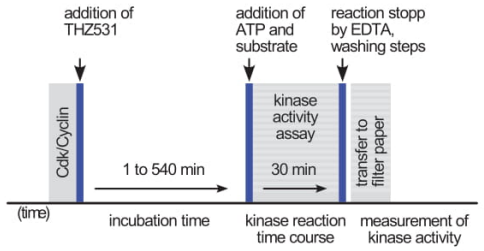
c



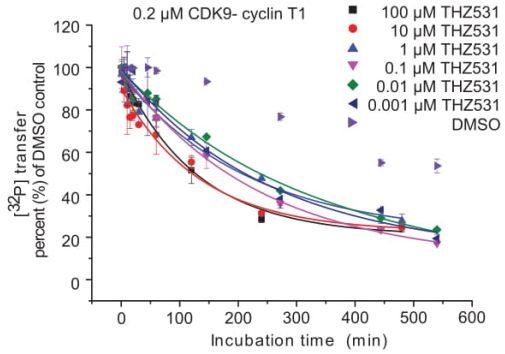
d



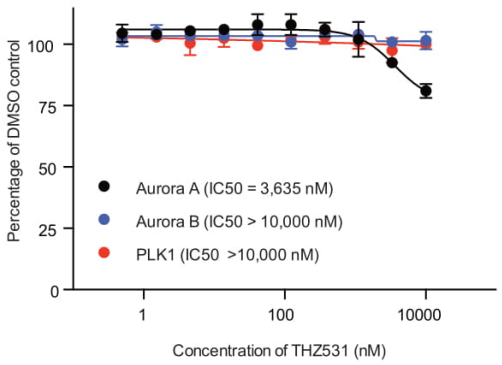
e



f



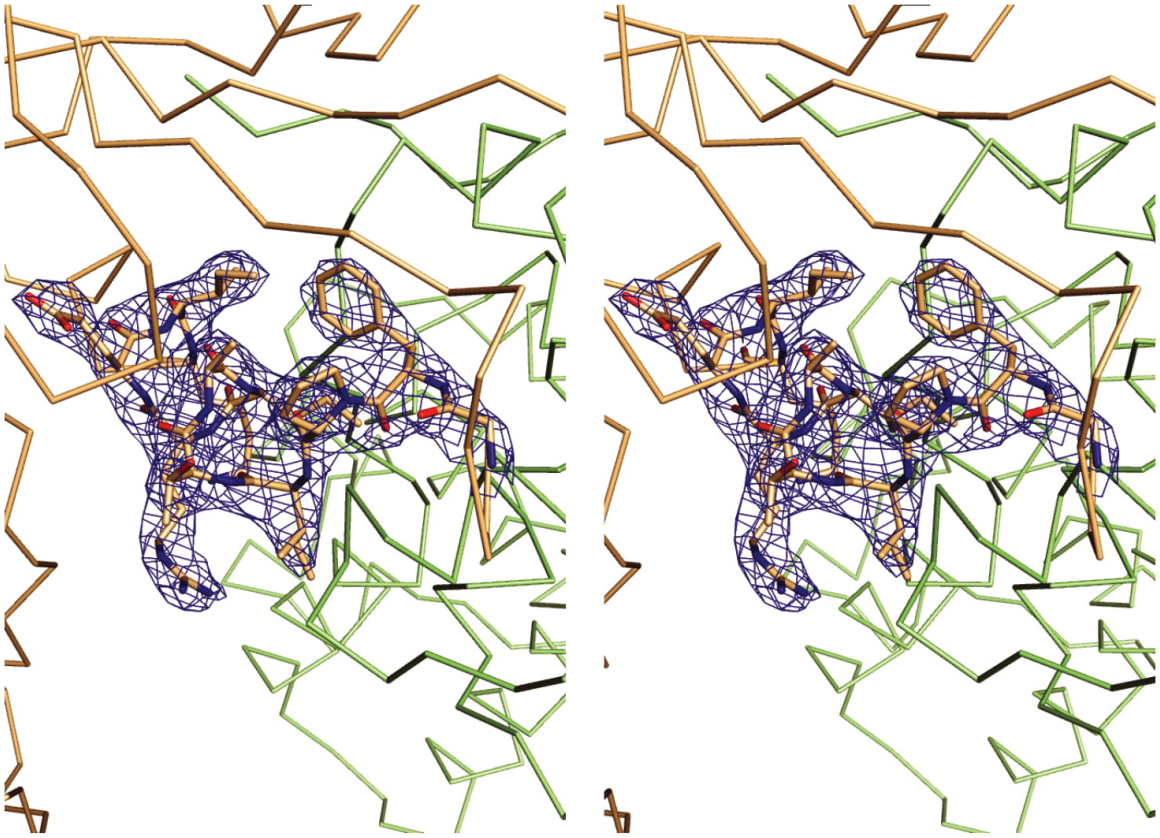
g



Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 3 | Electron density for THZ531 and the PITAIRE helix of CDK12. Stereo view showing electron density (2Fo-Fc contoured at 1.0σ) of the CDK12 PITAIRE helix (αC).

Nature Chemical Biology: doi:10.1038/nchembio.2166

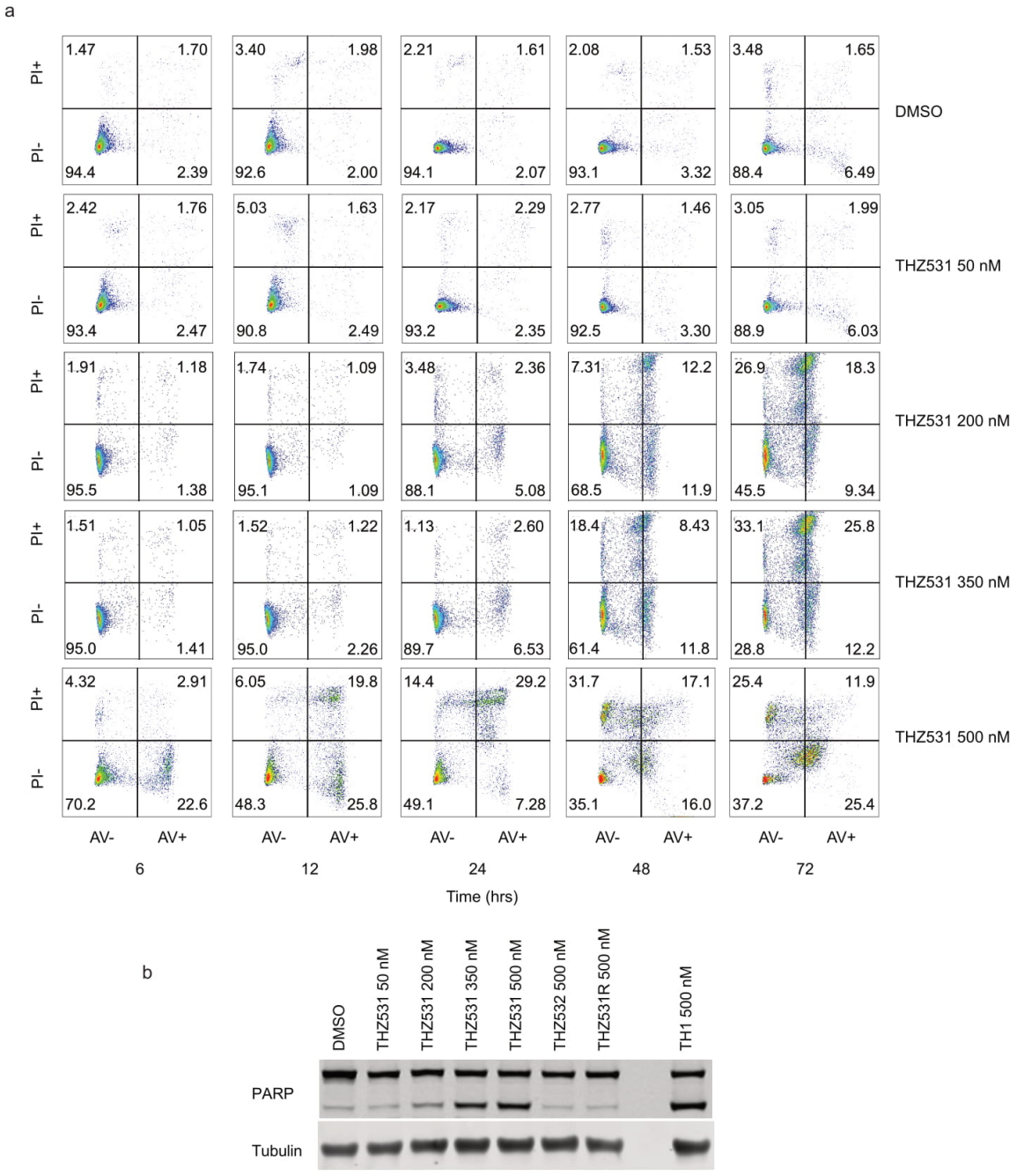


Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 4 | THZ531 induces apoptosis to Jurkat cells

a. Representative Annexin V and propidium iodide stainings for Jurkat cells incubated with THZ531. Jurkat cells were treated with 50, 200, 350, and 500 nM THZ531 for the indicated times. Cells were stained with Annexin V and propidium iodide. Experiments were performed in biological triplicate. **b.** THZ531 induces PARP cleavage. Jurkat cells were treated with 50, 200, 350, and 500 nM THZ531 for 24 hours. Lysates were probed with PARP and tubulin antibodies. Uncut western blots are in Supplementary Fig. 10.

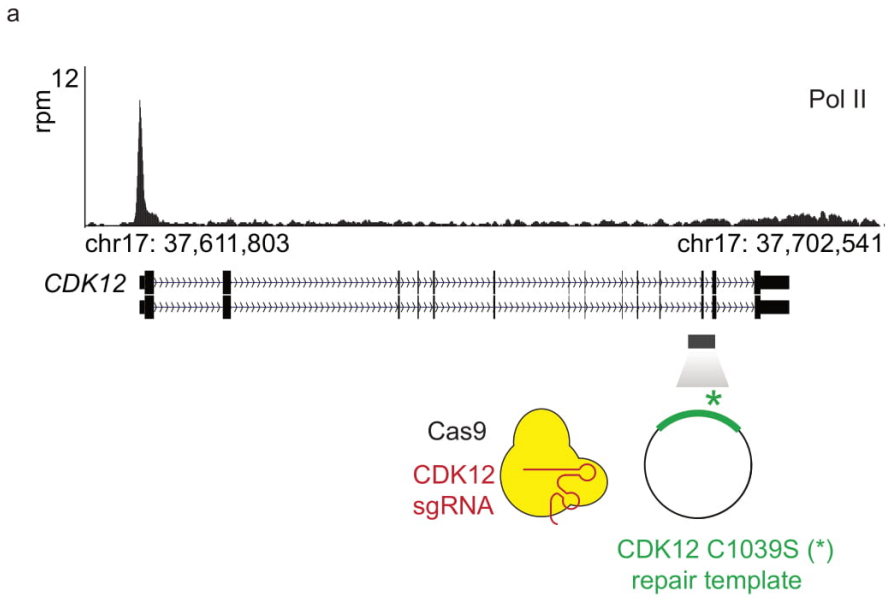
Nature Chemical Biology: doi:10.1038/nchembio.2166

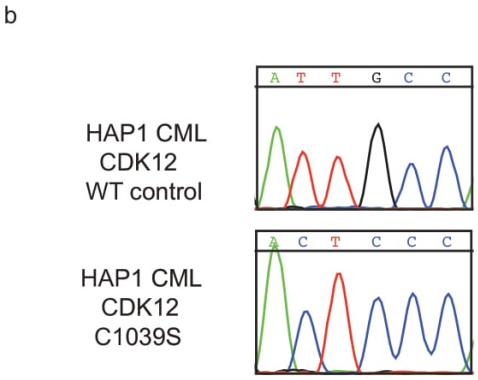


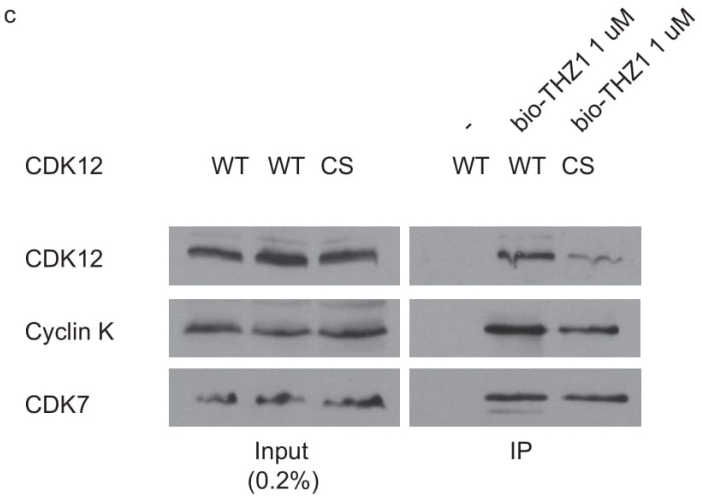
Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 5 | Mutation of Cys-1039 to serine reduces CDK12 covalent affinity and rescues THZ531 –induced proliferation defects. a, Gene track of CDK12 with schematic of CRISPR technique to mutate CDK12 allele. **b,** CRISPR technique mutates C1039 to serine (C1039S). Genomic DNA from CDK12 WT control and CDK12 C1039S HAP1 cells were Sanger sequenced. TGC (Cys) was successfully mutated to TCC (Ser). Other silent mutations were added to remove NGG CRISPR targeting sequence and to permit initial PCR screening of mutated alleles. **c,** CDK12 C1039S mutation prevents CDK12 pulldown with bioTHZ1. 25 million cells of WT control and C1039S HAP1 clones were lysed and probed with 1 μM bioTHZ1 at 4 degrees overnight. Interacting proteins were precipitated with streptavidin beads and probed with indicated antibodies. Uncut westerns blots are in Supplementary Fig. 10.

Nature Chemical Biology: doi:10.1038/nchembio.2166



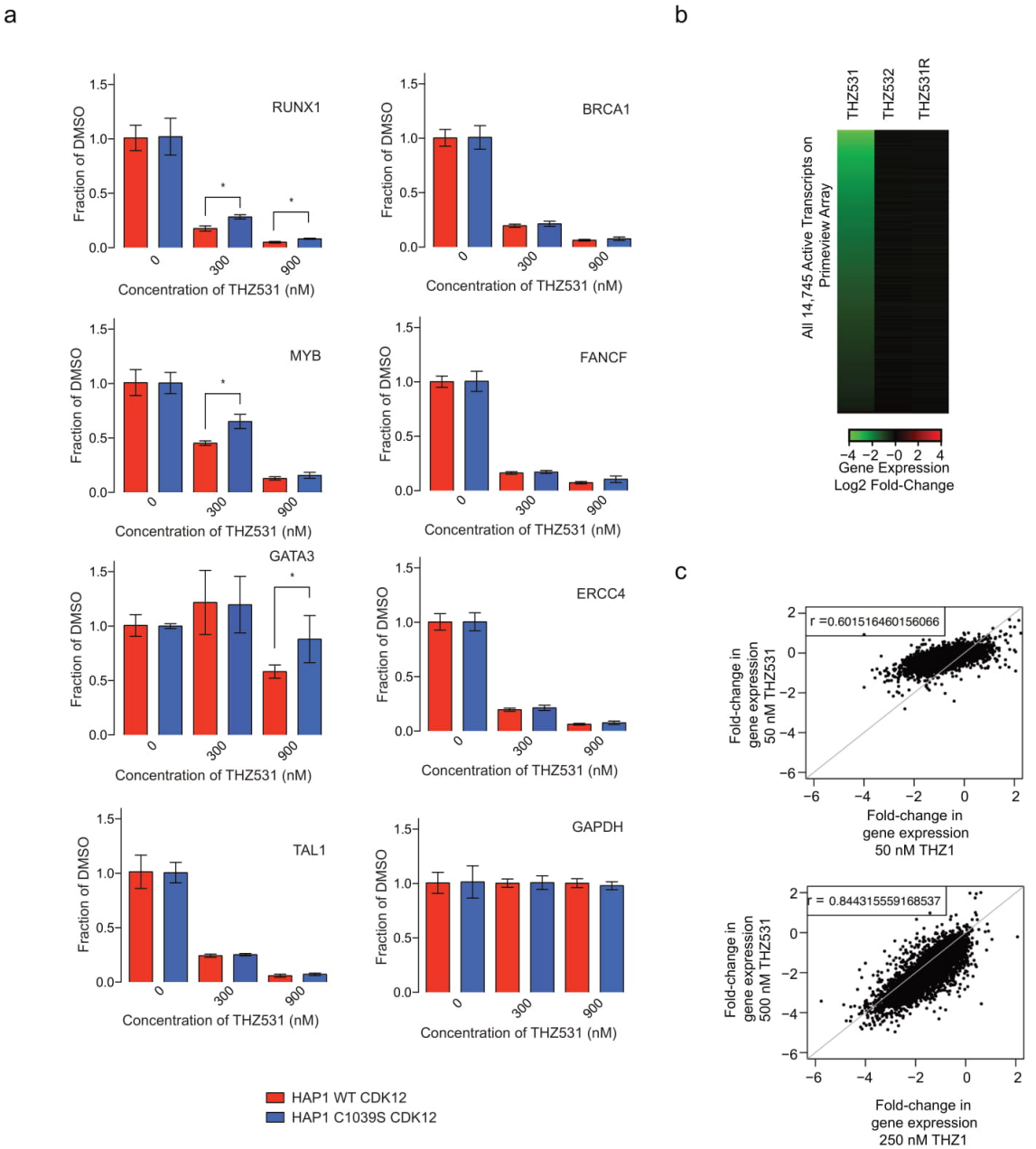




Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 6 | **THZ531 inhibits gene expression** **a,** Expression of CDK12 C1039S partially restores T-ALL transcription factor gene expression. RT qPCR of T-ALL transcription factors and DDR gene transcripts. RT qPCRs were performed in biological triplicate and error bars are +/- SD. **b,** THZ531R and THZ532 do not affect steady-state mRNA levels. Jurkat cells were treated with 500 nM THZ531, THZ531R, or THZ532 for 6 hrs. Heatmaps display the log2 fold-change in gene expression vs. DMSO for the 14,745 transcripts expressed in DMSO. **c,** THZ1 and THZ531 display similar yet distinct effects on the expression of 14,745 expressed genes (in DMSO). Log2 fold-change in gene expression for 50 nM THZ1 vs. THZ531 (left) and 250 nM THZ1 vs. 500 nM THZ531 (right). Pearson coefficient r = 0.60 and 0.84 respectively.

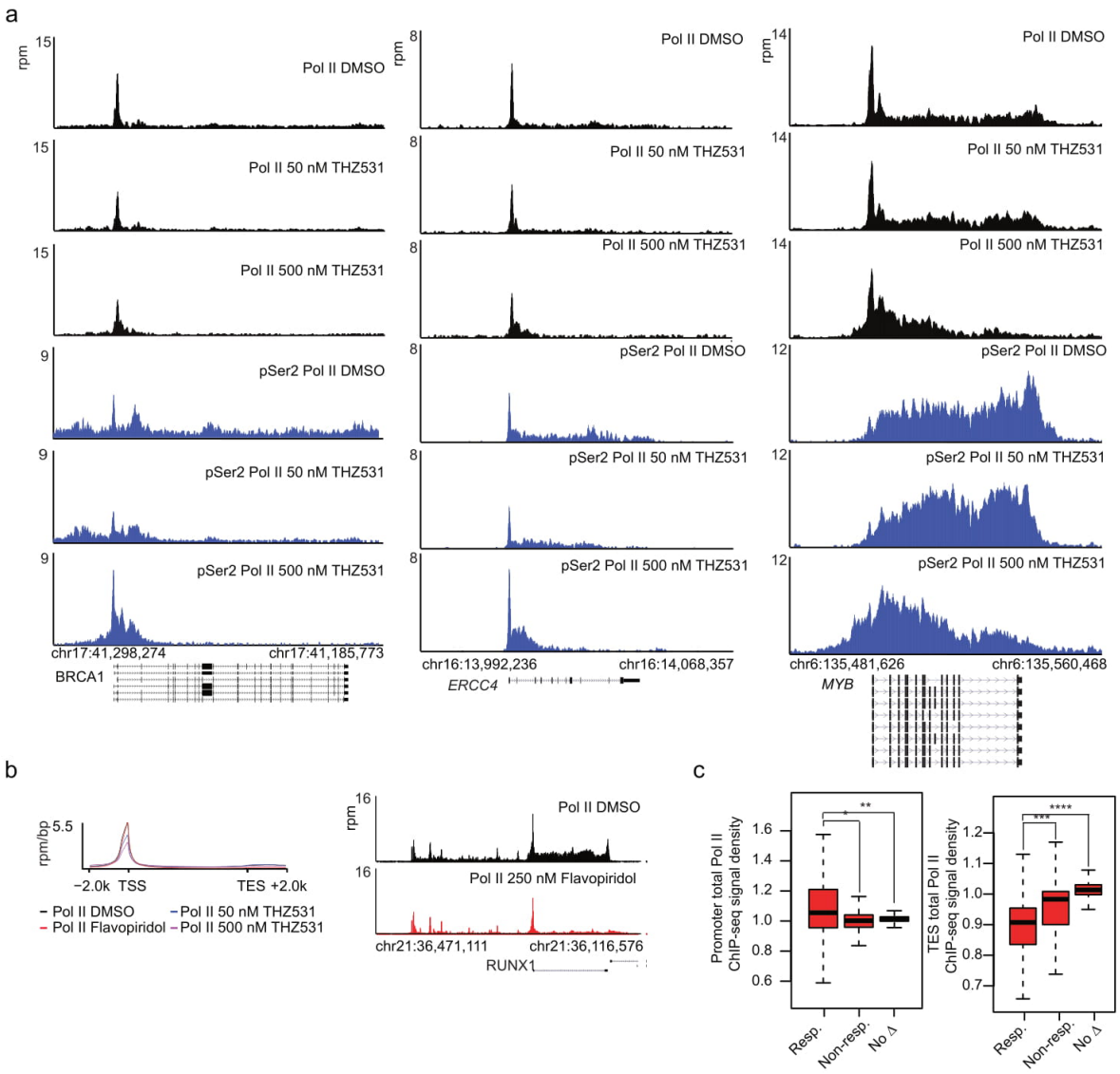
Nature Chemical Biology: doi:10.1038/nchembio.2166



Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 7 | THZ531 inhibits transcriptional elongation. a, Gene tracks whose expression is sensitive to THZ531. **b,** THZ531 and Flavopiridol show different effects on Pol II distributions. Metagene representation of global Pol II occupancy across gene bodies. Average ChIP-seq signal in 13906 genes expressed in 6h DMSO conditions in units of rpm/bp (left). Gene tracks of Pol II ChIP-seq at *RUNX1* gene locus following 250 nM Flavopiridol treatment for 6 hrs. (right). **c**, Flavopiridol significantly increases promoter–bound Pol II ChIP-seq signal and decreases elongating Pol II ChIP-seq signal at Flavopiridol–responsive genes relative to non-responsive genes. Box plots of Pol II ChIP-seq signal density at 5' transcriptional start sites (TSS) and 3' termination sites (TSS) at 2001 Flavopiridol –responsive genes (Resp.), 2001 non-responsive (Non-resp.), and 2001 genes whose expression doesn't change (no Δ). *p-value = 9.37e-54, **p-value = 2.60e-85, ***p-value = 2.27e-97, ****p-value = 1.34e-296. Responsive genes are defined as those having > log2 fold-change in gene expression.

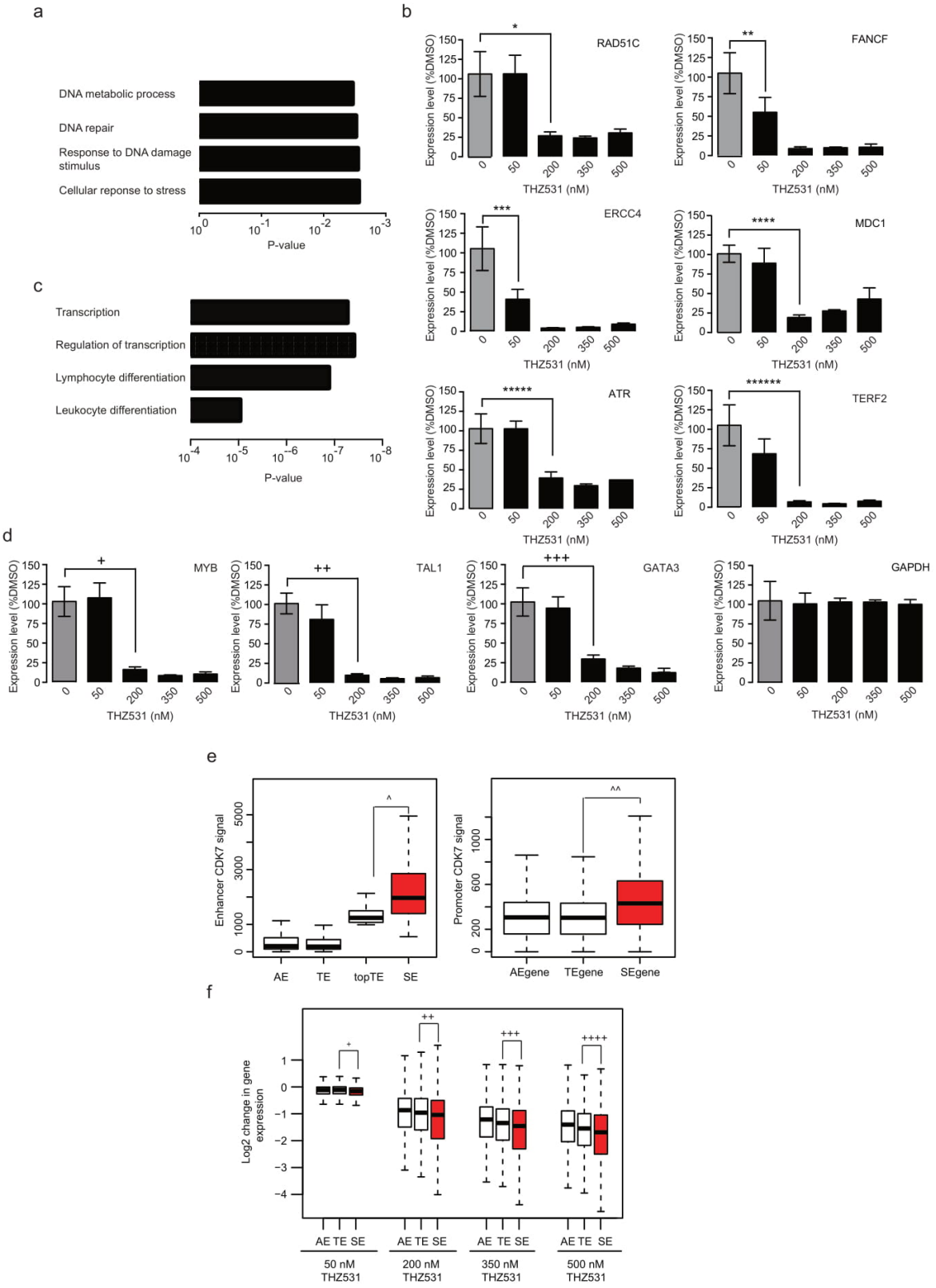
Nature Chemical Biology: doi:10.1038/nchembio.2166



Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 8 | THZ531 downregulates DDR and transcription factor gene expression **a**, The top 2% of genes downregulated with 50 nM show enrichment for genes encoding factors that regulate DDR. David gene ontology analysis, p-values supplied by David program. **b,** RT-qPCR of additional DDR genes transcript expression following THZ531 treatment. *p-value=2.48e-05, **p-value=1.93e-03, ***p-value=2.05e-04, ****p-value=6.71e-10, *****p-value=6.13e-06, ******p-value=1.04e-06. **c,** The top 2% of genes downregulated with 200 nM show enrichment for genes encoding factors that regulate transcription. David gene ontology analysis, p-values supplied by David program. **d,** RT-qPCR of additional T-ALL transcription factor gene transcripts following THZ531 treatment. +p-value=1.29e-07, ++p-value=1.11e-09, +++p-value=6.04e-07. **e,** Super –enhancers and promoters of their associated genes contain more CDK7 ChIP-seq signal compared to typical enhancers and their associated gene promoters. Boxplots demonstrating CDK7 ChIP-seq enhancer (left) and promoter (right) signal at all enhancers (AE), typical enhancers (TE), the top 818 TEs (Top TEs), and the 818 super –enhancers (SE). ^p-value = 4.68e-51, ^^p-value = 4.58e-142. **f**, Super–enhancer –associated gene expression is more sensitive to THZ531. Boxplots showing the fold-change in gene expression for those genes associated with AEs, TEs, and SEs. +p-value = 3.78e-06, ++p-value = 8.90e-06, +++p-value = 3.16e-06, ++++p-value = 1.79e-06, calculated with the two-tailed Student's t test. All RT-qPCR experiments were performed in biological triplicate and error bars are +/- SD. GAPDH gene expression was used as internal control for all RT-qPCR experiments. P-values were determined with a two-tailed Student's T test.

Nature Chemical Biology: doi:10.1038/nchembio.2166

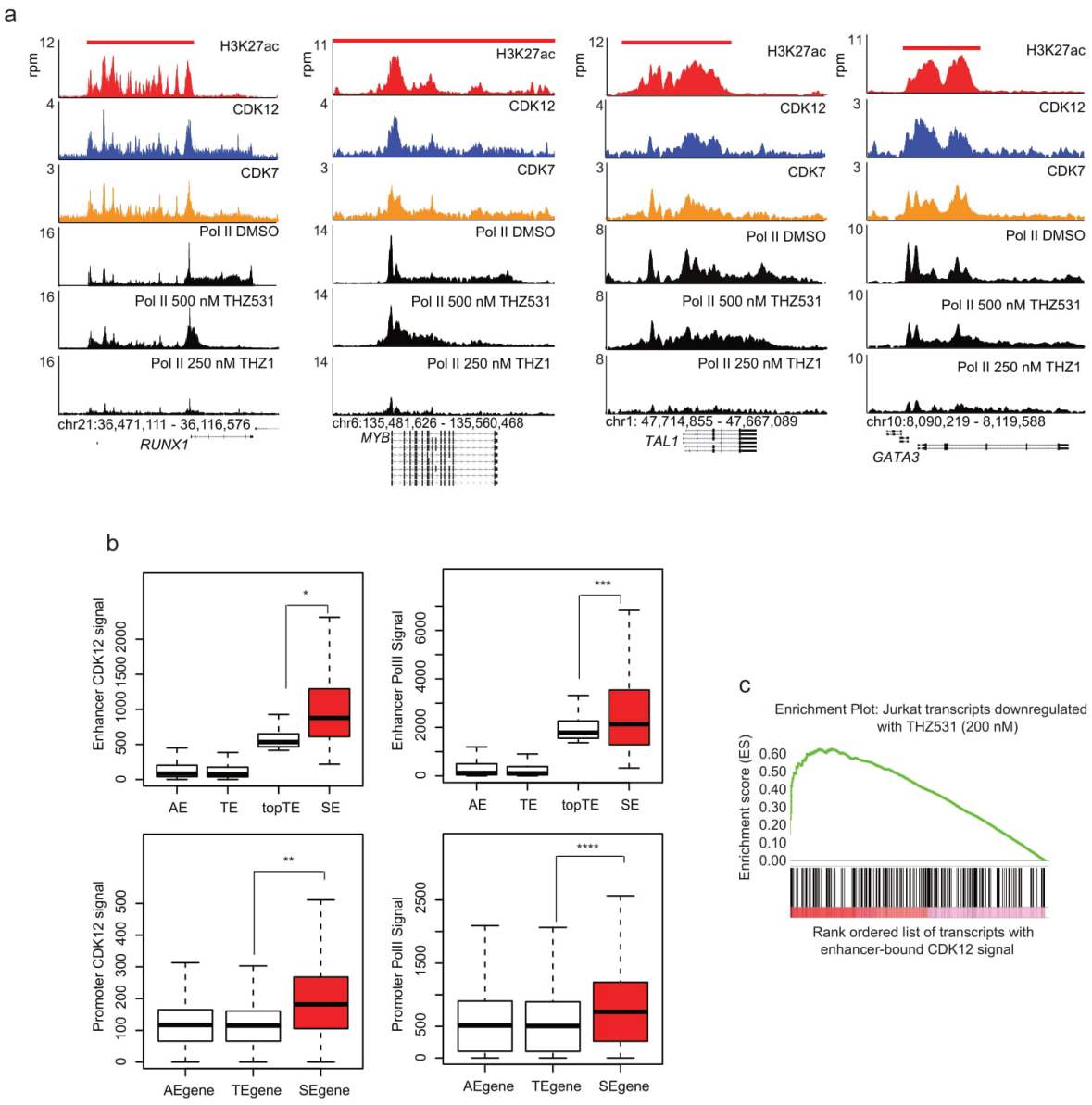


Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 9 | Super-enhancers contain exceptional amounts of CDK12.

a, *S*uper-enhancer genes contain large amounts of CDK7 (yellow), CDK12 signal (blue), and H3K27Ac (red). Pol II (black) elongation is impacted following treatment with both 250 nM THZ1 and 500 nM THZ531. The red bar indicates the genomic coordinates of a super –enhancer. **b,** SEs (top) and their associated gene promoters (bottom) contain more CDK12 and Pol II ChIP-seq signal compared to TEs and their associated gene promoters. Boxplots demonstrating the ChIP-seq signal for CDK12 (top left) and Pol II (top right) at AEs, TEs, 818 Top TEs , and 818 SEs. Boxplots demonstrating the ChIP-seq signal for CDK12 (bottom left) and Pol II (bottom right) at gene promoters associated with AEs, TEs, and SE. *p-value = 2.63e-124, **p-value = 3.16e-46, ***p-value = 4.58e-12, ****p-value = 9.39e-18. **f**, Transcripts down-regulated by 200 nM THZ531 are enriched for transcripts whose associated enhancers contain the highest levels of CDK12 ChIP-seq signal. Gene set enrichment analysis of top 500 transcripts downregulated following a 6-hour treatment with THZ531 (200 nM) in comparison to CDK12 ChIP-seq signal at enhancers for these transcripts GSEA-supplied p-value < 0.001.

Nature Chemical Biology: doi:10.1038/nchembio.2166

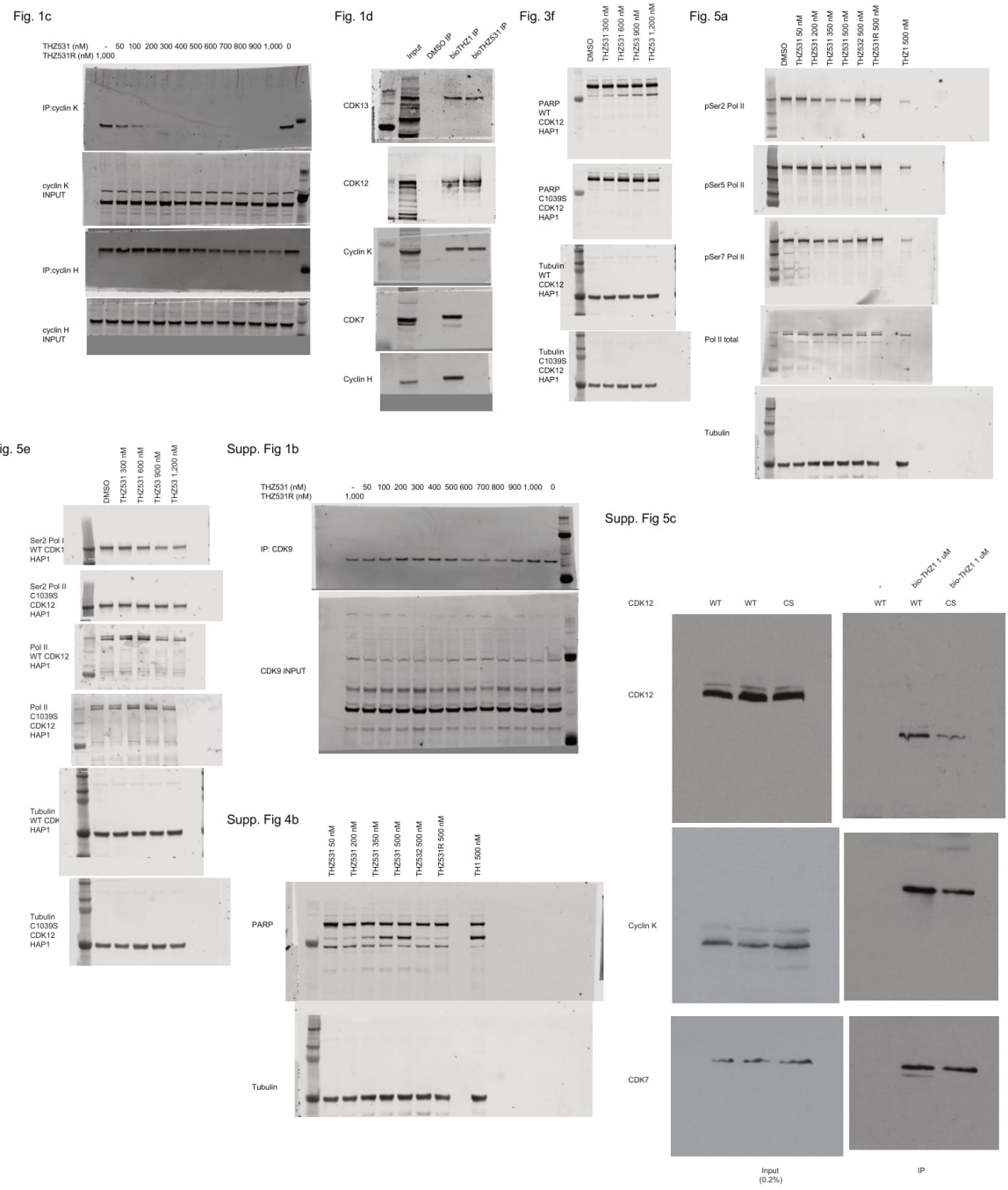


Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 10 | Uncut western blots

Uncut western blots corresponding to cropped western blots in main and supplementary figures.

Nature Chemical Biology: doi:10.1038/nchembio.2166



Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Figure 11 | CDK12 genomic sequence for genome editing

Modified CDK12 genomic sequence used as repair template in genomic editing experiments.

Nature Chemical Biology: doi:10.1038/nchembio.2166

The modified CDK12 genomic sequence (RefSeq Accession NM_016507) was cloned into pUC57-AMP by Genewiz and used as the repair template for genome editing:

GCCATGAGGAGTCGACATTACACTAGAATGTTGATACATTGAATATGACTTGAAACATATAAGGGTTTTA
CTGAAATTTGGGAACTCCTATTATAGAGAGATTTATAGTAAGATTTGACCTACTTGGTCTGATTGTATGA
GAGATTCTGGACTAAGTTATACGTAGTGCTGAACTCCACAGATGAGCAAGTTGATATCCAGTCATAGAA
ATTCTGTGGGCAGAAAGGGGAAAGAACTAGTCTTTGGTCCTCACAAACCAAATTACACATAAGTTGATTT
TGCACAGAGATGTTTTGATCATGATACATGGTTCCATGGTTTAAGTCACCGCCTTTCCTCGACCTTCTGT
ATACATTAACAAGCCAGCCATTACTGTCCTCGTCTTTATTCCCTACTGAAA
CTTGACCAATACTGTTCTCCTTTTTTGTTGTTATGCCAAGGAAAGACAGTATTTATATGGGAATTTATATA
GCTGGTCTGTACCTAGTATTAGCAAGATCCTCCTTTCTCACTATGTAACTTTCGTACTTTTATTTCCTCAG
ACTACCATCTGCCATGATTTATTTTACACTGCTGTTACCTACTTTTACTAACTTTTTTGGTTATTTTTGTTT
TCCTTATTCTTTACATTTCCCACAGTCTTTGCCTTCCCATTTTAATCCTTTGCTCCTCCATTCATTCACTG
CTGCATTCCCTTACATATTTCCCCTTTGCTTTGTCTTTTTCCAGCCTCCCCCACTGGCAAGACTCCCACG
AGCTCTGGAGTAAGAAACGCCGACGTCAGCGACAAAGTGGTGTTGTAGTCGAAGAGCCACCTCCATCC
AAAACTTCTCGAAAAGAAACTACCTCAGGGACAAGTACTGAGCCTGTGAAGAACAGCAGCCCAGCACC
ACCTCAGCCTGCTCCTGGCAAGGTGGAGTCTGGGGCTGGGGATGCAATAGGTCAGTGCCAGAATGGG
GCCTTTGTGCTTTTGCTAAGCGTATTTGGCAGGTTTTGAAGGTCAAGTGTAAGGAGTTTGTGTGTGTGT
CTGTTTTTTAGTTTTTGGTCCAGTGAACCATTAGGAAAGAGGTTAATGTCCTCATTATGGTGAGAAGGAG
ATTAACCAAGTTTCGTACATATGGTCCAGAAAGTCTACAATAGATCATTTTCCTTTCCTGAAATATGTACG
GAACACTGTATCTCAATTAGATGTGTGGCATAAATCAGTGAATTAGGTAAGCCAAATTTAATTTAAGGTT
AGTTCACCCATGACTTTATGAAACACATAATTTGGAGCCTTGTTTATATTAGATTTGATAAGACCTTCATA
GAATCTAGTAGTGTTAAATTACAATTGCTAAATGTTGTTCTATAGACTGTTATTTTGATCTAAATGTATCTG
AAACTATGTCATATTGAACTGTTATAACAAAGTAGTTCTCCTGTGTTATGACCAATTCTTCGGGACAAAA
GATACTCTTTTTCTGGTTCACTTTCCCCTCTTCGAATTCTCGAGTACCG

1. Green highlighting indicates the introduced desired TCC mutation, which codes for serine (C1039S), replacing TGC which codes for cysteine (C1039, WT)
2. Yellow highlighting indicates wobble mutations introduced to remove Cas9 –targeting cysteines, to prevent cutting of repair template.
3. Pink highlighting indicates Wobble mutations introduced for PCR-based screening to permit WT vs. mutated allele discrimination.
4. Red highlighting indicates Sal I and EcoRI sites used for pUC57 cloning

Sanger sequence of confirmed mutant C1039S allele:

TGATCCGCCTGCCTCTACCTCCCAAAGTGCTGGGATTACAGGCATGAGCTGCCGCGCCCAGCCAAATT
CATTTATTACACTAGAATGTTGATACATTGAATATGACTTGAAACATATAAGGGTTTTACTGAAATTTGGG
AACTCCTATTATAGAGAGATTTATAGTAAGATTTGACCTACTTGGTCTGATTGTATGAGAGATTCTGGAC
TAAGTTATACGTAGTGCTGAACTCCACAGATGAGCAAGTTGATATCCAGTCATAGAAATTCTGTGGGCA
GAAAGGGGAAAGAACTAGTCTTTGGTCCTCACAAACCAAATTACACATAAGTTGATTTTGCACAGAGAT
GTTTTGATCATGATACATGGTTCCATGGTTTAAGTCACCGCCTTTCCTCGACCTTCTGTATACATTAACA
AGCCAGCCATTACTGTCCTCGTCTTTATTCCCTACTGAAACTTGACCAATACTGTTCTCCTTTTTTGTTGT
TATGCCAAGGAAAGACAGTATTTATATGGGAATTTATATAGCTGGTCTGTACCTAGTATTAGCAAGATCC
TCCTTTCTCACTATGTAACTTTCGTACTTTTATTTCCTCAGACTACCATCTGCCATGATTTATTTTACACTG
CTGTTACCTACTTTTACTAACTTTTTTGGTTATTTTTGTTTTCCTTATTCTTTACATTTCCCACAGTCTTTGC
CTTCCCATTTTAATCCTTTGCTCCTCCATTCATTCACTGCTGCATTCCCTTACATATTTCCCCTTTGCTTT
GTCTTTTTCCAGCCTCCCCCACTGGCAAGACTCCCACGAGCTCTGGAGTAAGAAACGCCGACGTCAGC
GACAAAGTGGTGTTGTAGTCGAAGAGCCACCTCCATCCAAAACTTCTCGAAAAGAAAC

Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Tables

Supplementary Table 1: Intracellular KiNATiv™ profiling assay identifies CDK12 and CDK13 as major intracellular targets of THZ531.

Nature Chemical Biology: doi:10.1038/nchembio.2166

Jurkat cells treated with indicated compounds at 1μM

Kinase	Reference	Sequence	Labeling Site	THZ-1-31-1 1μM	THZ-1-31-1-R 1μM
CDK12	UniRef100_Q9NYV4	DIKCSNILLNNSGQIK	Lys2	98.1	8.1
CDK13	UniRef100_Q14004	DIKCSNILLNNR	Lys2	82.9	-5.6
AurA	UniRef100_O14965	DIKPENLLIGSAGELK	Lys2	54.4	-13.5
PLK1	UniRef100_P53350	DLKLGNLFLNEDLEVK	Lys2	53.3	16.1
AurA	UniRef100_O14965	FILAVKVFK	Lys2	51.7	-2.9
CLK1	UniRef100_P49759	LTHTDLKPENILFVQSDYTEAYNPK	Lys2	42.6	4.5
JNK1, JNK2, JNK3	UniRef100_P45983, UniRef100_P53779, UniRef100_P45984	DLKPSNIVVK	Lys2	42.3	2.6
PLK1	UniRef100_P53350	CFDSDADTKEVFAGKIVPK	Lys1	38.9	3
NEK4	UniRef100_P51957	DLKTQNVFLTR	Lys2	38.1	33.3
CDK7	UniRef100_P50613	DIKNTNQVAHK	Lys1	34	7
PCTAIRE2	UniRef100_Q00537	DLKPQNLLINEK	Lys2	32.9	-7.3
CDK7	UniRef100_P50613	DLKPNNLLDENGVLK	Lys2	32.8	6.1
SIK	UniRef100_Q9H2G2	DLKAGNILFTLDGDIK	Lys2	30.6	25.4
CHK1	UniRef100_B4DT73	LSKGDGLEFK	Protein Kinase Domain	25.3	6.3
MAP3K5	UniRef100_Q99683	DIKGDNVLNTYSGVLK	Lys2	24.8	17.8
IRAK4	UniRef100_Q9NWZ3	DIKSANILIDEAFTAK	Lys2	24.2	18.6
TAO1, TAO3	UniRef100_Q7L7X3, UniRef100_Q9H2K8	DIKAGNILLTEPDQVK	Lys2	23.3	32.5
GSK3A	UniRef100_P49840	DIKPQNLLVDPDTAVLK	Lys2	22.9	8.9
MAP3K2, MAP3K3	UniRef100_Q9Y2U5, UniRef100_Q99759	DIKGANILR	Lys2	22.5	18.2
AurA, AurB, AurC	UniRef100_O14965, UniRef100_Q9UQB9, UniRef100_Q96GD4	GKFGNVYLAR	ATP Loop	22.2	-16.5
GSK3B	UniRef100_P49841	DIKPQNLLIDPDTAVLK	Lys2	22	3.3
OSR1	UniRef100_O9BGI9, UniRef100_O95747	DVKAGNILLGEDGSVQIADFGVSAFLATGGDITR	Lys2	21.1	20.5
SNRK	UniRef100_Q9NRH2	DLKPENVVFFEK	Lys2	19.5	-3.9
MST4, YSK1	UniRef100_Q00506, UniRef100_Q9P289	DIKAANVLLSEGDIAK	Lys2	19.4	32
CDK9	UniRef100_P50750	DMKAANVLITR	Lys2	19.3	17.9
MAP2K3	UniRef100_P46734	DVKPSNVLINK	Lys2	19.3	6.6
MST3	UniRef100_Q9Y6E0	DIKAANVLLSEHGEVK	Lys2	18.9	7.1
GCK	UniRef100_Q12851	DIKGANLLTLQSDVK	Lys2	18.7	35.7
MAP3K4	UniRef100_Q9Y6R4	DIKGANFLTSGGLIK	Lys2	18.6	37.3
PRPK	UniRef100_Q96S44	FLSGLELVKQGAEAR	ATP Loop	17.4	14.4
KHS1	UniRef100_Q9Y4K4	DIKGANILLTDHGDVK	Lys2	17	14
MAST3	UniRef100_O60307	DLKPENILLTSLGHIK	Lys2	17	-29
HPK1	UniRef100_Q92918	DIKGANILINDAGEVR	Lys2	16.7	27
IKKb	UniRef100_O14920	DLKPENVLLDQGEQIR	Lys2	16.5	4.5
LOK	UniRef100_O94804	DLKAGNVLMTLEGDIR	Lys2	16.5	13.4
CHK2	UniRef100_O96017	DLKPENVLLSSQEEDCLIK	Lys2	16.4	-8
ITK	UniRef100_Q08881	FVLDDQYTSSTGTKFPVK	Activation Loop	16.1	-0.6
Erk2	UniRef100_P28482	DLKPSNLLINTTCDLK	Lys2	16	8.9
PHKB	UniRef100_Q9UBF8	VPHTQAVVLNSKDK	ATP	15.9	-10.8
IKKa	UniRef100_O15111	DLKPENVLQDVGGK	Lys2	15.6	4.8
MAP3K1	UniRef100_Q13233	DVKGANLLDSTGQIR	Lys2	15.2	20.9
MST1, MST2	UniRef100_Q13188, UniRef100_Q13043	DIKAGNILVTEGHAK	Lys2	14.9	7.2
CaMK2a, CaMK2b, CaMK2d, CaMK2g	UniRef100_Q13557B, UniRef100_Q13557, UniRef100_Q13555, UniRef100_Q9UQM7	DLKPENLLLASK	Lys2	14.3	8.4
Erk1	UniRef100_P27361	DLKPSNLLINTTCDLK	Lys2	14.3	5
TAO2	UniRef100_Q9UL54	DVKAGNILSEPGLVK	Lys2	14.3	27.7
GPRK5	UniRef100_P34947	DLKPENILLDDYGHIR	Lys2	14.2	0.7
RSK2 domain2	UniRef100_P51812	DLKPSNILYVDESGNPESIR	Lys2	14	-1.8
p38a	UniRef100_Q16539	DLKPSNLAVNEDCELK	Lys2	13.9	3.3
PIPSK3	UniRef100_Q9Y2I7	GGKSGAAFYATEDDRFLK	ATP	13.5	3.7
CDK13, CDK8	UniRef100_P49336, UniRef100_Q9BWU1	DLKPANILVMGEGPER	Lys2	13.3	-4.2
NLK	UniRef100_Q9UBE8	DIKPGNLLVNSNCVLK	Lys2	13.2	-1.7
MARK1, MARK2	UniRef100_Q7KZI7, UniRef100_Q9P0L2	EVAVRIDK	Lys1	12.9	-2.9
ZC3/HGK, ZC2/TNIK, ZC3/MINK	UniRef100_Q9Y419, UniRef100_Q9UKE5, UniRef100_Q8N4C8	DIKGQNVLLTENAEVK	Lys2	12.8	15
MARK2, MARK3	UniRef100_P27448, UniRef100_Q7KZI7	DLKAENLLLDADMNIK	Lys2	12.3	-11.7
PKN1	UniRef100_Q16512	VLLSEFRPGEIFAHKALK	Lys1	11.7	0.9
PKN2	UniRef100_Q16513	DLKLDNLLLDTEGFVK	Lys2	11.7	-2.5
MAP2K4	UniRef100_P45985	DIKPSNILDR	Lys2	11.3	-7.7
MSK1 domain1	UniRef100_O75582	DWLGLNILIDSNGHIKVVLTDFGLSK	Lys2	11.3	0.3
MST2	UniRef100_Q13188	ESGQVVAIKQVPVESDLQEIIK	Lys1	11	-9.8
PYK2	UniRef100_Q14289	YIEDEDYYKASVTR	Activation Loop	10.5	7.8
Erk5	UniRef100_Q13164	DLKPSNLLVNENCELK	Lys2	10	4.6
RSK1 domain2	UniRef100_Q15418	DLKPSNILYVDESGNPECLR	Lys2	10	-8.4
SGK3	UniRef100_Q96BR1	IVYKDLKPENILLDSVGHVKLTDFGLCK	Lys2	9.7	12.4
STLK3	UniRef100_Q9UEW8	DLKAGNILLGEDGSVQIADFGVSAFLATGGDVTR	Lys2	9.7	6.6
CaMKLd	UniRef100_Q8IL85	DLKPENLLYYSGDEESK	Lys2	9.6	-10
CCRK	UniRef100_Q8IZL9	DLKPANLLSASGGLK	Lys2	9.3	7.3
NEK8	UniRef100_Q86SG6	DLKTQNILLDK	Lys2	9.3	6.4
MARK4	UniRef100_Q96L34	DLKAENLLLDAEANIK	Lys2	9	-5.6
MST1	UniRef100_Q13043	ETGQEVAIKQIPVESDLQEIIK	Lys1	8.8	-6.9
p38a	UniRef100_Q16539	QELNKTIWEVPER	Protein Kinase Domain	8.8	-1.1
ROCK1, ROCK2	UniRef100_Q75116, UniRef100_Q13464	DVKPDNMLLDK	Lys2	8.8	-8.3
CDK10	UniRef100_Q15131	DLKVSNLLMTDK	Lys2	8.7	-0.7
ITK	UniRef100_Q08881	VAIKTIR	Lys1	8.5	-6.7
GCN2	UniRef100_Q9P2K8	LDGCCYAVKR	Lys1	8.2	-0.8
PKD2	UniRef100_Q9BZL6	DVAVKVIDK	Lys1	8.2	-4
RSK1 domain1, RSK2 domain1	UniRef100_P51812, UniRef100_Q15418, UniRef100_Q15349	DLKPENILLDEEGHIK	Lys2	8.1	-13.2
JAK1 domain2	UniRef100_P23458	IGDFGLTKAIETDKEYYTVK	Activation Loop	8	-16.6
PCTAIRE2, PCTAIRE3	UniRef100_Q00537, UniRef100_Q07002	SKLTENLVALKEIR	Lys1	8	-1
NEK3	UniRef100_P51956	SKNPLTGNQK	Activation Loop	7.9	-2.7
FER	UniRef100_P16591	QEDGGVYSSSGLKQIPIK	Activation Loop	7.8	6.7
CHK1	UniRef100_B4DT73	DIKPENLLLDER	Lys2	7.5	-28
DNAPK	UniRef100_P78527	KGGSWQEINVAEK	ATP	7.4	6.3
DRAK1	UniRef100_Q5UEE5	DSVHLDLKPQNILLTSESPLGDIK	Lys2	7.3	15.5
SNRK	UniRef100_Q9NRH2	VAVEVIDK	Lys1	7.1	-1
JNK2	UniRef100_P45984	YQQLKPIGSEGAQGIVCAAFDTVLGINVAVKK	Lys1	6.8	11
PKN1	UniRef100_Q16512	DLKLDNLLLDTEGYVK	Lys2	6.7	-5
SRPK1, SRPK2	UniRef100_Q96SB4, UniRef100_Q96S B4	FVAMKVVK	Lys1	6.6	-28.4
LKB1	UniRef100_Q15831	DIKPGNLLLTTGGTLK	Lys2	6.5	6.9
RSKL1	UniRef100_Q96S38	VLGVDKVLLVMDTR	ATP	6.5	-3.6
MLK1	UniRef100_P80192	DLKSSNILICDK	Lys2	6.2	10.3
CDK6	UniRef100_Q00534	DLKPQNILVTSSGQIK	Lys2	5.7	0.7
MAP3K2	UniRef100_Q9Y2U5	ELAVKQVQFDPDSPETSKEVNALECEIQLLK	Lys1	5.5	-11.4
LATS1	UniRef100_O95835	ALVATKTLR	Lys1	5.4	-27.3
MAP2K3	UniRef100_P46734	HAQSGTIMAVKR	Lys1	5.2	-7
NDR2	UniRef100_Q9Y2H1	DIKPQNLLLDAK	Lys2	5	-4.5
eEF2K	UniRef100_O00418	YIKYNSNSDVR	ATP	4.9	19.5
MAP2K4	UniRef100_P45985	MVHHPSQQIMAVKR	Lys1	4.8	-5.2
SGK3	UniRef100_Q96BR1	FVAVKVLQK	Lys1	4.8	1
DGKH	UniRef100_Q86XP1	ATFSFCVSPLLVFVNSKSGDNQGVK	ATP	4.7	4.6
ULK1	UniRef100_O75385	DLKPQNLLISPAGR	Lys2	4.5	-20.4
MAP2K6	UniRef100_P52564	DVKPSNVLINALGQVK	Lys2	4.4	1.8
MAP2K6	UniRef100_P52564	HVPSGQIMAVKR	Lys1	4.3	-7.3
NDR1	UniRef100_Q15208	DIKPQNLLLDSK	Lys2	4	-7.8
ILK	UniRef100_Q13418	ISMADVKFSFQCPGR	Protein Kinase Domain	3.4	10.1
TAK1	UniRef100_O43318	DLKPPNLLVAGGTVLK	Lys2	3.3	-12.3
EphB2	UniRef100_P29323	FLEDDTSDPTYTSALGGKIPIR	Activation Loop	3.2	5.5
NEK6, NEK7	UniRef100_Q8TDX7, UniRef100_Q9HC98	DIKPANVFITATGVVK	Lys2	3.2	-6.2
MAP3K15, MAP3K5, MAP3K6	UniRef100_Q99683, UniRef100_095382, UniRef100_Q8ZN16	IAKEIPER	Lys1	3.1	3.6
TLK1	UniRef100_Q9UKI8	YLNEIKPPIHYDLKPGNILVDGTACGEIK	Lys2	3.1	-8.9
GCN2	UniRef100_Q9P2K8	DLKPVNIFLDSDDHVK	Lys2	3	-2.7
FGR	UniRef100_P09769	LIKDDEYNPCQGSKFPIK	Activation Loop	2.9	4.6
PKCI	UniRef100_P41743	DLKLDNVLLDSEGHIK	Lys2	2.8	-11.3
MAP2K5	UniRef100_Q13163	DVKPSNMLVNTR	Lys2	2.7	5.2
MAP2K1, MAP2K2	UniRef100_P36507, UniRef100_Q02750	DVKPSNLVNSR	Lys2	2	2
MAP2K7	UniRef100_O14733	DVKPSNILLDER	Lys2	1.8	-9.6
RSK2 domain1	UniRef100_P51812	LTDFGLSKESIDHEKK	Activation Loop	1.7	2.8
CLK3	UniRef100_P49761	YEIVGNLGEGTFGKVVECLDHAR	ATP Loop	1.3	1.7
IKKb	UniRef100_O14920	WHNGETGEQIAIKQCR	Lys1	1.2	-1.3
PAK5	UniRef100_Q9P286	VMDPTRSLITGK	ATP	1.1	-6.9
PITSLRE	UniRef100_P21127	DLKTSNLLLSHAGILK	Lys2	1	-1.9
ZAP70	UniRef100_P43403	ISDFGLSKALGADDSYYTAR	Activation Loop	0.8	-6.4
CK1a	UniRef100_P48729	DIKPDNFLMGIGR	Lys2	0.7	-4.2
NEK9	UniRef100_Q8TD19	DIKTLNIFLTK	Lys2	0.5	-2.2
PIK3CD	UniRef100_O00329	TKVNWLAIHVSKDNRQ	ATP	0.5	-5.3
PHKB	UniRef100_Q9UBF8	LLSVVKCGDDLRQELLAPQVLK	ATP	0.3	7.3
RSK2 domain1	UniRef100_P51812	VLGQDSFGKVFLVK	ATP Loop	0.3	13.9
CK2a2	UniRef100_P19784	DVKPHNVMIDHQDK	Lys2	0.2	-7.9
PKR	UniRef100_P19525	DLKPSNFLVDTK	Lys2	0.2	5.1
AMPKa1, AMPKa2	UniRef100_P54646, UniRef100_Q96E92	DLKPENVLLDAHMNAK	Lys2	0.1	-2.6
MAP3K1	UniRef100_Q99759	ELASKQVQFDPDSPETSKEVSALECEIQLLK	Lys1	0.1	5.7
MPSK1	UniRef100_O75716	DLKPTNILLGDEGQPVLMDLGSMNQACIHVEGSR	Lys2	0.1	-1.7
CDK5	UniRef100_Q00535	DLKPQNLLINR	Lys2	0	-1.4
CSK	UniRef100_P41240	VAVKCIK	Lys1	-0.1	-9
CDK5	UniRef100_Q00535	NRETHEVALKR	Lys1	-0.3	1.2
MAST1	UniRef100_Q96GX5	GAFGKVYLGQK	ATP Loop	-0.3	-2.7
MARK3, MARK4	UniRef100_Q96L34, UniRef100_P27448	EVAIKIIDK	Lys1	-0.4	-7.1
TLK1	UniRef100_Q9UKI8	YAAYKHQLNK	Lys1	-0.6	6.4
JAK1 domain2	UniRef100_P23458	YDPEGDNTGEQVAVKSLKPESGGNHIADLKK	Lys1	-1	0.6
p70S6K, p70S6Kb	UniRef100_Q9UBS0, UniRef100_P23443	GGVGKVFQVR	ATP Loop	-1	8.4
TLK2	UniRef100_Q86UE8	YVAVKHQLNK	Lys1	-1.1	4.8
ULK3	UniRef100_Q3DWG7	EVVAKCVAK	Lys1	-1.3	-2.1
MLK3	UniRef100_Q16584	DLKSNNILLQPESDDMEHK	Lys2	-1.5	-23.4
PFK	UniRef100_Q96NZ5	DLKPSNIFFTMEDVVK	Lys2	-1.6	-1.3
TLK2	UniRef100_Q86UE8	YLNEIKPPIHYDLKPGNILVNGTACGEIK	Lys2	-1.7	-7.1
MST4	UniRef100_Q9P289	TQQVVAIKILEAEDEEIEDIQQEITVLSQCDSSYVTK	Lys1	-1.9	-22.3
PKD3	UniRef100_Q94806	NYHCDLKPENVLLASAEFPQVQK	Lys2	-2.5	4.2
CaMK2g	UniRef100_Q13555	TSTQEYAAKIINTK	Lys1	-2.6	14.3
CSK	UniRef100_P41240	VSDFGLTKEASSTQDTGKLPVK	Activation Loop	-2.7	-2.5
CK1d, CK1e	UniRef100_P48674, UniRef100_P49674	DVKPDNFLMGLGKK	Lys2	-2.9	-7.1
IKKe, TBK1	UniRef100_Q14164, UniRef100_Q9UHD2	DIKPGNIMM	Lys2	-3.2	-17.5
IRAK4	UniRef100_Q9NWZ3	GYVNNTTVAVKK	Lys1	-3.2	-1
LATS1	UniRef100_O95835	DIKPQNVLIDR	Lys2	-3.6	-31.5
PRP4	UniRef100_Q13523	AAGIGKDPKENPNLR	Protein Kinase Domain	-3.8	7.2
p38d, p38g	UniRef100_O15264, UniRef100_P53778	DLKPGNLAVNEDCELK	Lys2	-4.1	-5.9
GCK	UniRef100_Q12851	DTVTSELAAVKIVK	Lys1	-4.4	3.5
SIK	UniRef100_Q9H2G2	AQNIETSVLAAAKVIDTK	Lys1	-4.4	-5.7
CaMKLd	UniRef100_Q8IL85	LFAVKCIPK	Lys1	-4.5	-5.2
PCTAIRE1, PCTAIRE3	UniRef100_Q00536, UniRef100_Q07002	DLKPQNLLINER	Lys2	-4.6	2.5
MASTL	UniRef100_Q96GX5	LYAVKVVK	Lys1	-4.7	-7.3
JAK1 domain1	UniRef100_P23458	QLASALSYLEDKDLVHGNVCTKNLLLAR	Protein Kinase Domain	-5.1	3
AKT2, AKT3	UniRef100_Q9Y243, UniRef100_P31751	GTFGKVILVR	ATP Loop	-5.2	3.9
CaMKK2	UniRef100_Q96BR4	LAYNENDNTYYAMKVLSK	Lys1	-5.4	-2.2
p70S6K	UniRef100_P23443	DLKPENIMLNHQGHVK	Lys2	-5.4	5.2
p38b	UniRef100_Q15759	QELNKTVWEVPQR	Protein Kinase Domain	-5.5	-4.1
MARK3	UniRef100_P27448	EVAKIIDKTQLNPTSLQK	Lys1	-5.7	1.5
KHS2	UniRef100_Q8IVH8	NVNTGELAAKVIK	Lys1	-5.8	17.4
MARK2	UniRef100_Q7KZI7	EVAVKIIDKTQLNSSLQK	Lys1	-5.8	6.2
CDC2	UniRef100_Q5H9N4	DLKPQNLLIDDK	Lys2	-5.9	-8.7
STLK5	UniRef100_Q7RTN6	YSVKVLPWLSPYLQQNLQGYDAK	Activation Loop	-6	0.8
MLK2	UniRef100_Q02779	DLKSNNILILEAIENHNLADTVLK	Lys2	-6.1	2.3
LATS2	UniRef100_Q9NRM7	DIKPQNILIDLDGHIK	Lys2	-6.3	-33.2
AMPKa1	UniRef100_Q96E92	KSHYLGDTLGVGTFGKVK	ATP Loop	-6.4	2.2

Labeling Site Key

Lys1	Conserved Lysine 1
Lys2	Conserved Lysine 2
ATP Loop	ATP binding loop.
Activation Loop	Activation loop
ATP	ATP site in non-canonical kinase (e.g. lipid kinase)
Protein Kinase Domain	Other lysine within kinase domain, possibly not in ATP binding site
Other	Labeling of residue outside of the protein kinase domain, possibly not in ATP binding site

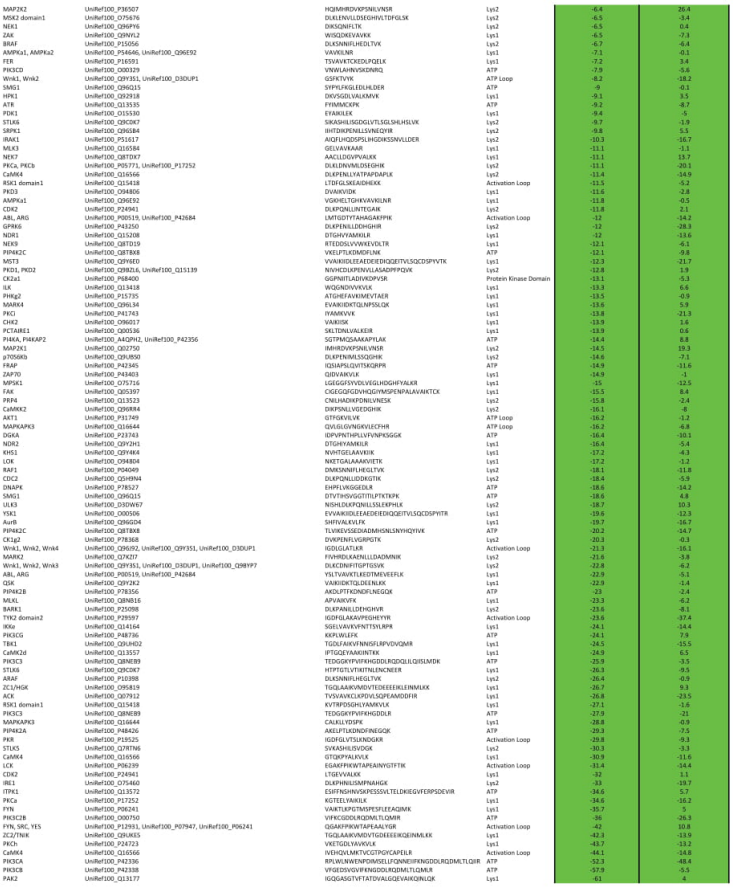
	>90% Inhibition
	75 - 90% Inhibition
	50 - 75% Inhibition
	35 - 50% Inhibition
	No change
	> 100% increase in MS signal
Not colored	Variable or weak data not reaching significance despite >35% inhibition

Note: This KINativ dataset is the result of an analysis of duplicate treated samples and either duplicate or quadruplicate control samples. The % changes in MS signals being reported are statistically significant (Student T-test score <0.04). Additionally, the KINativ method is performed using biological matrices (cells/tissues), which can be inherently variable, followed by a complex procedure of sample preparation and mass spec analysis. Thus, we recommend the use of independent, biological replicates to determine the inter-sample variability of the dataset (i.e., repeating the experimental conditions more than one time and with an independent sample of the biological matrix in order to validate the dataset). In addition, we strongly recommend that customers obtain supporting data through additional independent KINativ studies or through orthogonal approaches prior to making critical project decisions.

The data reported was performed in a non-GLP manner and was not intended for regulatory submission. It was generated to provide scientific data for informational purposes only.

KINativ project managers are available to discuss the results and assist customers in understanding the strengths and limitations of this KINativ dataset. ActivX Biosciences, Inc. bears no responsibility for decisions made by customers based on this KINativ dataset or for any experimental suggestions offered by KINativ project managers.

Nature Chemical Biology: doi:10.1038/nchembio.2166



MAP2K2	UniRef100_P36507	HQIMHRDVKPSNILVNSR	Lys2	-6.4	26.4
MSK2 domain1	UniRef100_O75676	DLKLENVLLDSEGHIVLTDFGLSK	Lys2	-6.5	-3.4
NEK1	UniRef100_Q96PY6	DIKSQNIFLTK	Lys2	-6.5	0.4
ZAK	UniRef100_Q9NYL2	WISQDDKEVAVKK	Lys1	-6.5	-7.3
BRAF	UniRef100_P15056	DLKSNNIFLHEDLTVK	Lys2	-6.7	-6.4
AMPKa1, AMPKa2	UniRef100_P54646, UniRef100_Q96E92	VAVKILNR	Lys1	-7.1	-0.1
FER	UniRef100_P16591	TSVAVKTCKEDLPQELK	Lys1	-7.2	3.4
PIK3CD	UniRef100_O00329	VNWLAVHKYSKDNRQ	ATP	-7.9	-5.6
Wnk1, Wnk2	UniRef100_Q9Y3S1, UniRef100_D3DUP1	GSFKTVYK	ATP Loop	-8.2	-18.2
SMG1	UniRef100_Q96Q15	SYPYLFKGLEDLHLDER	ATP	-9	-0.1
HPK1	UniRef100_Q92918	DIKPSQDLVALKMVK	Lys1	-9.1	3.5
ATR	UniRef100_Q13535	FYIMMCKPK	ATP	-9.2	-8.7
PDK1	UniRef100_O15530	FYAMLEK	Lys1	-9.4	-5
STLK6	UniRef100_Q8C0K7	SIKASHLIGSDGLVTLSGLSHLHSLVK	Lys2	-9.7	-1.9
SRPK1	UniRef100_Q96SB4	IIHTDIKPENILLSVNEQYIR	Lys2	-9.8	5.5
IRAK1	UniRef100_P51617	AIQFLHQDSKPSLHRGDIKSSNVLLDER	Lys2	-10.3	-16.7
MLK3	UniRef100_Q16584	GELVAVKAAR	Lys1	-11.1	-1.1
NEK7	UniRef100_Q8TDX7	AACLLDGVPVALKK	Lys1	-11.1	13.7
PKCa, PKCb	UniRef100_P05771, UniRef100_P17252	DLKLENVMLDSEGHIK	Lys2	-11.1	-20.1
CaMK4	UniRef100_Q16566	DLKPENLLYATPAPDAPLK	Lys2	-11.4	-14.9
RSK1 domain1	UniRef100_Q15418	LTDFGLSKEAIDHEKK	Activation Loop	-11.5	-5.2
PKD3	UniRef100_O94806	DVAIKVIDK	Lys1	-11.6	-2.8
AMPKa1	UniRef100_Q96E92	VGKHELTGHKVAVKILNR	Lys1	-11.8	-0.5
CDK2	UniRef100_P24941	DLKPQNLLINTEGAIK	Lys2	-11.8	2.1
ABL, ARG	UniRef100_P00519, UniRef100_P42684	LMTGDTYTAHAGAKFPIK	Activation Loop	-12	-14.2
GPRK6	UniRef100_P43250	DLKPENILLDDEHGHIR	Lys2	-12	-28.3
NDR1	UniRef100_Q15208	DTGHVYAMKILR	Lys1	-12	-13.6
NEK9	UniRef100_Q8TD19	RTEDDSLVVVKEVDLTR	Lys1	-12.1	-6.1
PIP4K2C	UniRef100_Q8TBX8	VKELPTLKDMDFLNK	ATP	-12.1	-9.8
MST3	UniRef100_Q9Y6E0	VVVAIKIDLEEAEDEIEDIQQEITVLSQCDSPYVTK	Lys1	-12.3	-21.7
PKD1, PKD2	UniRef100_Q9BZL6, UniRef100_Q15139	NIVHCDIKPENVLLASADPFPQVK	Lys2	-12.8	1.9
CK2a1	UniRef100_P68400	GGPNITLADFGVAVKDPVSR	Protein Kinase Domain	-13.1	-5.3
ILK	UniRef100_Q13418	WQGNDIVVVKVLK	Lys1	-13.3	6.6
PHKg2	UniRef100_P15735	ATGHEFAVKIMEVTAER	Lys1	-13.5	-0.9
MARK4	UniRef100_Q96L34	EVAIKIDNTQLNPTSLQK	Lys1	-13.6	5.9
PKCI	UniRef100_P41743	IYAMKVVK	Lys1	-13.8	-21.3
CHK2	UniRef100_O96017	VAAKISK	Lys1	-13.9	1.6
PCTAIRE1	UniRef100_Q00536	SKLTDNLVALKEIR	Lys1	-13.9	0.6
PI4KA, PI4KAP2	UniRef100_A4QPH2, UniRef100_P42356	SGTPMQSAAKAPYLAK	ATP	-14.4	8.8
MAP2K1	UniRef100_Q02750	IMHRDVKPSNILVNSR	Lys2	-14.5	19.3
p70S6Kb	UniRef100_Q9UBS0	DLKPENIMLSSQGHIK	Lys2	-14.6	-7.1
FRAP	UniRef100_P42345	IQSAPSLQVTSKQRPR	ATP	-14.9	-11.6
ZAP70	UniRef100_P43403	GIDVAVKVLK	Lys1	-14.9	-1
MPSK1	UniRef100_O75716	LGEGGFSYVDLVEGLHDGHFYALKR	Lys1	-15	-12.5
FAK	UniRef100_Q05397	CGEGQGFGDVHQGIYMSPENPALAVAKTCK	Lys1	-15.5	8.4
PRP4	UniRef100_Q13523	CNILHADIKPDNILVNESK	Lys2	-15.8	-2.4
CaMKK2	UniRef100_Q96RR4	DIKPSNLLVGEDGHIK	Lys2	-16.1	-8
AKT1	UniRef100_P31749	GTFGKVILVK	ATP Loop	-16.2	-1.2
MAPKAPK3	UniRef100_Q16644	QVLGLGVNGKVLECFHR	ATP Loop	-16.2	-6.8
DGK6	UniRef100_P37743	IDPVPNTHPLLVFLVNPKSGGK	ATP	-16.4	-10.1
NDR2	UniRef100_Q9Y2H1	DTGHYAMKILR	Lys1	-16.4	-5.4
KHS1	UniRef100_Q9Y4K4	NVHTGELAAVKIIK	Lys1	-17.2	-4.3
LOK	UniRef100_O94804	NKETGALAAAKVIETK	Lys1	-17.2	-1.2
RAF1	UniRef100_P04049	DMKSNNIFLHEGLTVK	Lys2	-18.1	-11.8
CDC2	UniRef100_Q5HYN4	DLKPQNLLIDDKGTIK	Lys2	-18.4	-5.9
DNAPK	UniRef100_P78527	EHPFLVKSGEDLR	ATP	-18.6	-14.2
SMG1	UniRef100_Q96Q15	DTVTHSVSGGTITLPTKTKPK	ATP	-18.6	4.8
ULK3	UniRef100_Q8IWX7	NISHLDLKPDNILLSSLEKPHLK	Lys2	-18.7	10.3
YSK1	UniRef100_O00506	EVVAKIIDLEEAEDEIEDIQQEITVLSQCDSPYITR	Lys1	-19.6	-12.3
AurB	UniRef100_Q96GD4	SHIIVALKVLFK	Lys1	-19.7	-16.7
PIP4K2C	UniRef100_Q8TBX8	TLVKEVSSEDAADMHSNLSNYHQYIVK	ATP	-20.2	-14.7
CK1g2	UniRef100_P78368	DVKPENFLVGRPGTK	Lys2	-20.3	-0.3
Wnk1, Wnk2, Wnk4	UniRef100_Q9UIG0, UniRef100_Q9Y3S1, UniRef100_D3DUP1	IGDLGLATLKR	Activation Loop	-21.3	-16.1
MARK2	UniRef100_Q7KZI7	FIVHRDLKAENLLLDADMNIK	Lys2	-21.6	-3.8
Wnk1, Wnk2, Wnk3	UniRef100_Q9Y3S1, UniRef100_D3DUP1, UniRef100_Q9BYP7	DLKCDNIFTGPTGSVK	Lys2	-22.8	-6.2
ABL, ARG	UniRef100_P00519, UniRef100_P42684	YSLTVAVKTLKEDTMEVEEFLK	Lys1	-22.9	-5.1
QSK	UniRef100_Q9Y2K2	VAIKIIDKTQLDEENLKK	Lys1	-22.9	-1.4
PIP4K2B	UniRef100_P78356	AKDLPTTKDNDFLNEGQK	ATP	-23	-2.4
MLKL	UniRef100_Q8NB16	APVAKKVFK	Lys1	-23.3	-6.2
BARK1	UniRef100_P25098	DLKPANILLDEHGHVR	Lys2	-23.6	-8.1
TYK2 domain2	UniRef100_P29597	IGDFGLAKAVPEGHEYYR	Activation Loop	-23.6	-37.4
IKKe	UniRef100_Q14164	SGELVAVKVFNTTSYLRPR	Lys1	-24.1	-14.4
PIK3CG	UniRef100_P48736	KKPLWLEFK	ATP	-24.1	7.9
TBK1	UniRef100_Q9UHD2	TGDLFAKKVFNNISFLRPVDVQMR	Lys1	-24.5	-15.5
CaMK2d	UniRef100_Q13557	IFTGQEYAAKIINTKK	Lys1	-24.9	6.5
PIK3C3	UniRef100_Q8NEB9	TEDGGKYFVIFKHGDDLRQDQLILQIISLMDK	ATP	-25.9	-3.5
STLK6	UniRef100_Q8C0K7	HTPTQTLVTFKITNLENCNEER	Lys1	-26.3	-9.5
ARAF	UniRef100_P10398	DLKSNNIFLHEGLTVK	Lys2	-26.4	-0.9
ZC1/HGK	UniRef100_Q9S819	TGQLAAKVMDVTEDEEEEIKLEINMLKK	Lys1	-26.7	9.3
ACK	UniRef100_Q07912	TVSVAVKCLKPDVLSQPEAMDDFIR	Lys1	-26.8	-23.5
RSK1 domain1	UniRef100_Q15418	KVTRPDSGHLYAMKVLK	Lys1	-27.1	-1.6
PIK3C3	UniRef100_Q8NEB9	TEDGGKYFVIFKHGDDLR	ATP	-27.9	-21
MAPKAPK3	UniRef100_Q16644	CALKLLYDSPK	Lys1	-28.8	-0.9
PIP4K2A	UniRef100_P48426	AKELPTLKDNDFINEGQK	ATP	-29.3	-7.5
PKR	UniRef100_P19525	IGDFGLVTSLKNDGKR	Activation Loop	-29.8	-9.3
STLK5	UniRef100_Q78TN6	SVKASHLISVDGK	Lys2	-30.3	-3.3
CaMK4	UniRef100_Q16566	GTQEPPYALPVLK	Lys1	-30.9	-11.6
LCK	UniRef100_P06239	EGAKFPIKWTAPEAINYGTFTIK	Activation Loop	-31.4	-14.4
CDK2	UniRef100_P24941	LTGEVVALKK	Lys1	-32	1.1
IRE1	UniRef100_O75460	DLKPHNILISMPNAHGK	Lys2	-33	-19.7
ITPK1	UniRef100_Q13572	ESIFFNSHNVSKPESSSVLTELDKIEGVFERPSDEVIR	ATP	-34.6	5.7
PKCa	UniRef100_P17252	KGTEELYAKKILK	Lys1	-34.6	-10.2
FYN	UniRef100_P06241	VAIKTLKPGTMSPESFLEEAQIMK	Lys1	-35.7	5
PIK3C2B	UniRef100_O00750	VFKCGDDLRQDMLTLQMIR	ATP	-36	-26.3
FYN, SRC, YES	UniRef100_P12931, UniRef100_P07947, UniRef100_P06241	QSAKFPIKWTAPEAALYGR	Activation Loop	-42	10.8
ZC2/TNIK	UniRef100_Q8IUKE5	TGQLAAKVMDVTGDEEEEIKQEINMLKK	Lys1	-42.3	-13.9
PKCh	UniRef100_P24723	VKETGDLYAVKVLK	Lys1	-43.7	-13.2
CaMK4	UniRef100_Q16566	IVEHQVLMKTVCGTPGYCAPEILR	Activation Loop	-44.1	-14.8
PIK3CA	UniRef100_P42336	RPLWLNWENPDIMSELFQNNEIIFKNGDDLRQDMLTLQIIR	ATP	-52.3	-48.4
PIK3CB	UniRef100_P42338	VFQEDSVGVIPFKNGDDLRQDMLTLQMLR	ATP	-57.9	-5.5
PAK2	UniRef100_Q13177	IGQSASGTVFTATDVALGQEVAVKQMNLQK	Lys1	-61	4

Supplementary Table 2: Table of diffraction data collection and refinement statistics for CDK12-cyclin K and THZ531 co-crystal structure.

Nature Chemical Biology: doi:10.1038/nchembio.2166

Table 2 Data collection and refinement statistics (molecular replacement)

	$CDK12^{715-1052}$/cyclin K^{11-267}
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	*49.8, 148.7, 91.6*
α, β, γ (°)	90.0, 93.8, 90.0
Resolution (Å)	41.3-2.7(12.1-2.7)*
R_{merge}	0.071 (0.569)
$I / \sigma I$	12.5 (1.9)
Completeness (%)	98.7 (99.0)
Redundancy	3.0 (3.1)
Refinement	
Resolution (Å)	41.3-2.7
No. reflections	34,240 (2581)
R_{work} / R_{free}	22.1/26.2
No. atoms	
Protein	8993
Ligand/ion	80
B-factors	
Protein	61.1
Ligand/ion	87.0
R.m.s. deviations	
Bond lengths (Å)	0.0072
Bond angles (°)	1.1369

*Single crystal.
Values in parentheses are for highest-resolution shell.

Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Table 3: GEO upload files

Nature Chemical Biology: doi:10.1038/nchembio.2166

ChIP-seq samples	Figure	GEO
Jurkat DMSO Pol II	4a,b,c,d; 5f,h; 6f; SF5; SF7a,b,c; SF9a,b	GSM1850204 (NEW)
Jurkat H3K27ac	4a,b,c,d; 6e,f; SF8e,f; SF9a,b	GSM1296384
Jurkat CDK7	SF8e; SF9a	GSM1296385
Jurkat CDK12	4a,b,c,d; 6f,g; SF9a,b,c	GSM1850203 (NEW)
Jurkat 50nM THZ531 Pol II	5f,h; 6f; SF7a,b	GSM1850205 (NEW)
Jurkat 500nM THZ531 Pol II	5f,h; 6f; SF7a,b; SF9a	GSM1850206 (NEW)
Jurkat Input DNA		GSM1296386
Jurkat Flavo Pol II	SF7b,c	GSM1224787
Jurkat THZ1 Pol II	SF9a	GSM1224785
Jurkat DMSO pSer2 Pol II	5b,f,g; SF7a	GSE72023
Jurkat 50 nM THZ531 pSer2 Pol II	5b,f,g; SF7a	GSE72023
Jurkat 500 nM THZ531 pSer2 Pol II	5b,f,g; SF7a	GSE72023
Expression microarrays	5c,d,g,h; 6a,c,g; SF6b,c; SF7c; S8a,c,f; SF9c	GSE72022
Previous THZ1 data from THZ1 paper (both 50 and 250 nM data)	SF6c	GSM1224822, GSM1224826, GSM1224827, GSM1224818, GSM1224819, GSM1224820, GSM1224821
Previous Flavopiridol data from THZ1 paper (250 nM data)	SF7c	GSM1224823, GSM1224822

Nature Chemical Biology: doi:10.1038/nchembio.2166

Supplementary Data Sets

Supplementary Data Set 1: Mass spectrometry identifies CDK12-cyclin K and CDK13-cyclin K complexes as major targets of bioTHZ531 in Jurkat cell lysates.

See accompanying excel file

Supplementary Data Set 2: *In vitro* Ambit^{TM} binding assay shows THZ531 potently inhibits CDK13.

See accompanying excel file

Supplementary Data Set 3: Gene expression microarray data of THZ531, Flavopiridol, THZ1 – treated cells.

See accompanying excel file

Supplementary Data Set 4: Jurkat enhancers and super –enhancers identified by H3K27Ac ChIP-seq.

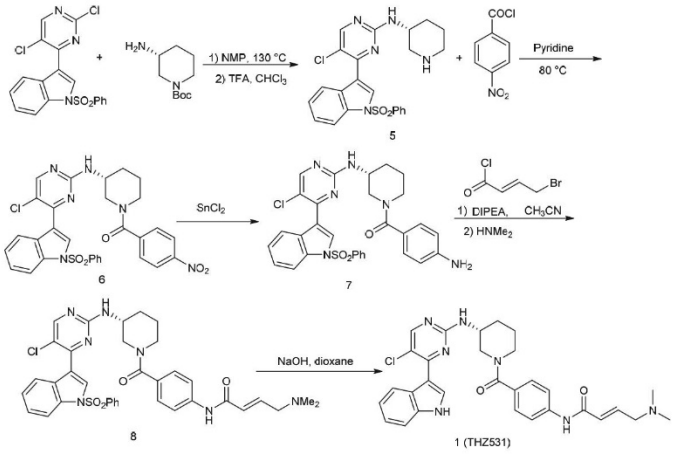
See accompanying excel file

Nature Chemical Biology: doi:10.1038/nchembio.2166

SUPPLEMENTARY NOTE

Chemistry

All solvents and reagents were used as obtained. 1H NMR spectra were recorded with a Varian Inova 600 NMR spectrometer and referenced to dimethylsulfoxide. Chemical shifts are expressed in ppm. Mass spectra were measured with Waters Micromass ZQ using an ESI source coupled to a Waters 2525 HPLC system operating in reverse mode with a Waters Sunfire C18 5 µm, 4.6 mm x 50 mm column. Purification of compounds was performed with either a Teledyne ISCO CombiFlash Rf system or a Waters Micromass ZQ preparative system. The purity was analyzed on an above-mentioned Waters LC-MS Symmetry (C18 column,4.6 mm x 50 mm, 5 µM) using a gradient of 5-95% methanol in water containing 0.05% trifluoroacetic acid (TFA). Detailed synthetic schemes and characterization data are presented below and elsewhere.[1]

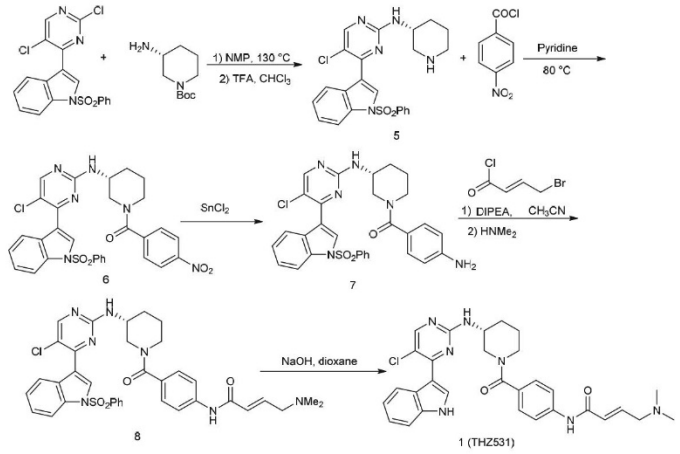


Scheme 1. The synthesis of THZ531

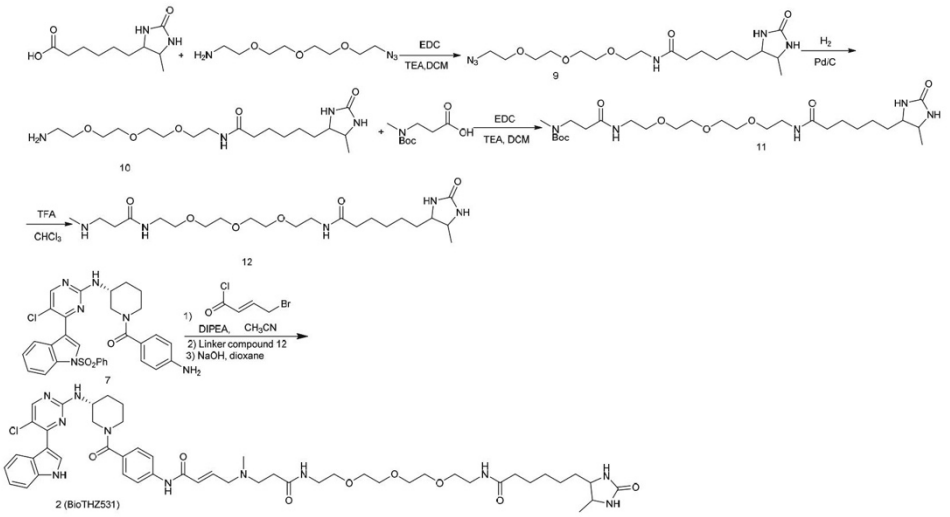
SUPPLEMENTARY NOTE

Chemistry

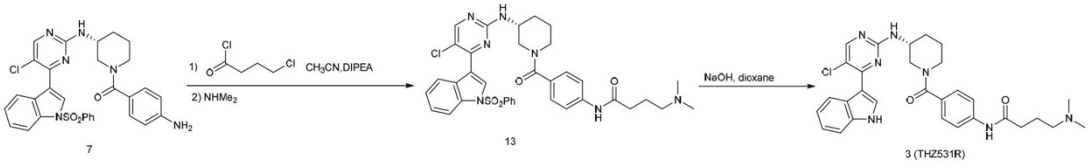
All solvents and reagents were used as obtained. 1H NMR spectra were recorded with a Varian Inova 600 NMR spectrometer and referenced to dimethylsulfoxide. Chemical shifts are expressed in ppm. Mass spectra were measured with Waters Micromass ZQ using an ESI source coupled to a Waters 2525 HPLC system operating in reverse mode with a Waters Sunfire C18 5 µm, 4.6 mm x 50 mm column. Purification of compounds was performed with either a Teledyne ISCO CombiFlash Rf system or a Waters Micromass ZQ preparative system. The purity was analyzed on an above-mentioned Waters LC-MS Symmetry (C18 column,4.6 mm x 50 mm, 5 µM) using a gradient of 5-95% methanol in water containing 0.05% trifluoroacetic acid (TFA). Detailed synthetic schemes and characterization data are presented below and elsewhere.[1]



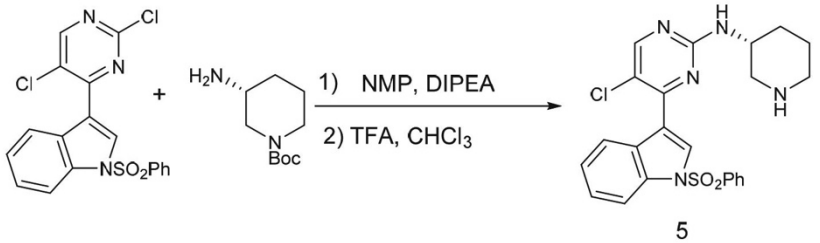
Scheme 1. The synthesis of THZ531



Scheme 2. The synthesis of BioTHZ531



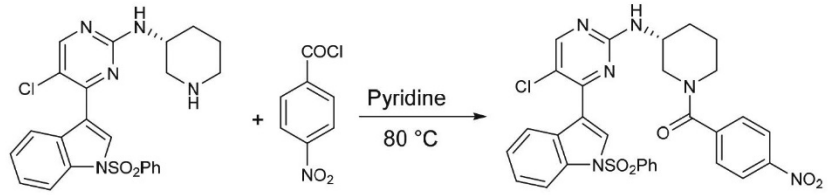
Scheme 3. The synthesis of THZ531R



(R)-5-chloro-4-(1-(phenylsulfonyl)-1H-indol-3-yl)-N-(piperidin-3-yl)pyrimidin-2-amine (5)

To a solution of 3-(2, 5-dichloropyrimidin-4-yl)-1-(phenylsulfonyl)-1H-indole (402mg) in NMP (5.0mL) was added tert-butyl (R)-3-aminopiperidine-1-carboxylate (200 mg, 1.0 equiv) and diisopropylethylamine (129 mg, 1.0 equiv). The solution was heated for 2 h at 130 ^{O}C.The

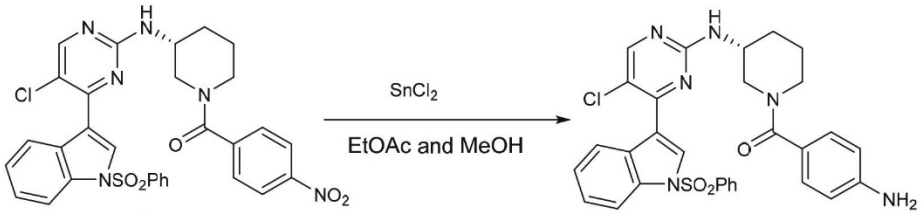
cooled solution was diluted with 100 mL of $CHCl_3$ and *i*-PrOH (4:1, 100 mL)) and then washed with water. After removing solvent, the crude was dissolved in $CHCl_3$ (10 mL) and then was treated with TFA (5 mL). After 30 min stirring at room temperature, the solvent was then removed. The separation by silica gel chromatography column with CH_2Cl_2/methanol (10/1) to give the product **5** (350 mg, 76%, two steps). LC-MS: *m/z* (M+H) 468.^{1}H NMR (500 MHz, DMSO) δ 8.70 (s, 2H), 8.54 (s, 1H), 8.48 (s, 1H), 8.29 (d, *J* = 8.2 Hz, 1H), 8.07 – 8.03 (m, 2H), 8.01 (d, *J* = 8.4 Hz, 1H), 7.73 (t, *J* = 7.5 Hz, 1H), 7.63 (t, *J* = 7.9 Hz, 2H), 7.45 (dd, *J* = 11.3, 4.1 Hz, 2H), 7.37 (t, *J* = 7.6 Hz, 1H), 4.18 (td, *J* = 13.5, 6.9 Hz, 1H), 3.41 (s, 1H), 3.22 – 3.17 (m, 1H), 2.91 (dd, *J* = 21.6, 11.2 Hz, 2H), 2.04 (dd, *J* = 12.6, 3.8 Hz, 1H), 1.92 (d, *J* = 4.4 Hz, 1H), 1.80 – 1.56 (m, 2H).^{13}C NMR (125 MHz, DMSO) ^{13}C NMR (126 MHz, DMSO) δ 160.19, 158.72, 156.61, 137.33, 135.47, 134.50, 130.47 (2C), 129.85, 128.99, 127.35 (2C), 126.03, 124.61, 123.56, 117.82, 116.04, 113.51, 47.06, 45.65, 43.76, 28.57, 20.97.



5 6

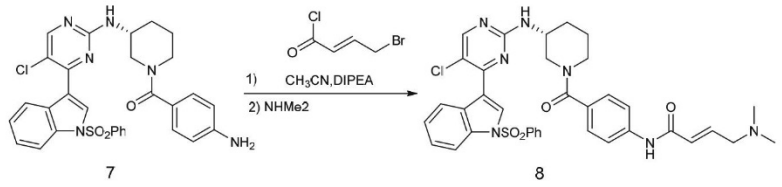
(R)-(3-((5-chloro-4-(1-(phenylsulfonyl)-1H-indol-3-yl)pyrimidin-2-yl)amino)piperidin-1-yl)(4-nitrophenyl)methanone (6) To a stirred solution of the compound 5 (467 mg) in pyridine (10 mL) was added 4-nitrobenzoyl chloride (180 mg, 1.0 equiv) and then was heated to 80 ^{O}C. The reaction mixture was stirred for 2 hours and then was concentrated under reduced pressure. The resulting crude was purified by silica gel column with CH_2Cl_2/methanol (10/1) to give the product **6** (450 mg, 73%). LC-MS: *m/z* (M+H) 617. ^{1}H NMR (500 MHz, DMSO) δ 8.49 (s, 1H), 8.31 (d, *J* = 41.7 Hz, 2H), 8.05 (d, *J* = 7.7 Hz, 2H), 8.00 (d, *J* = 8.3 Hz, 1H), 7.73 (t, *J* = 7.4 Hz, 1H), 7.63 (t, *J* = 7.8 Hz, 2H), 7.55 (s, 1H), 7.44 (t, *J* = 7.7 Hz, 1H), 7.37 (dd, *J* = 16.3, 8.5 Hz, 2H), 3.90 (s, 1H), 3.52 (s, 2H), 3.28 (s, 2H), 2.10 – 2.00 (m, 1H), 1.91 (s, 1H), 1.75 (s, 1H), 1.58

(s, 1H).^{13}C NMR (125 MHz, DMSO) δ 168.01, 160.29, 158.51, 156.27, 148.12, 143.03, 137.36, 135.43, 134.48, 130.45 (2C), 129.71, 128.95, 128.49 (2C), 127.31 (2C), 126.00, 124.57, 123.62 (2C), 117.85, 115.58, 113.45, 50.52, 48.04, 42.61, 29.88, 22.67.



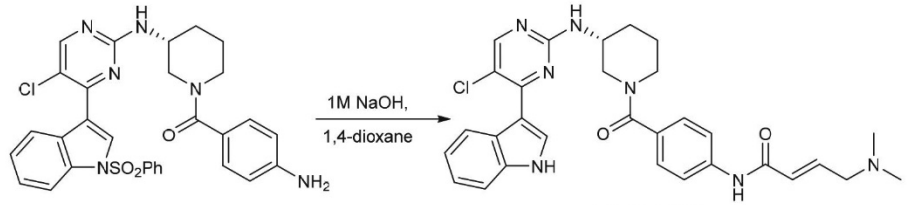
6 7

(R)-(4-aminophenyl)(3-((5-chloro-4-(1-(phenylsulfonyl)-1H-indol-3-yl)pyrimidin-2-yl)amino)piperidin-1-yl)methanone (7) The nitro compound 6 (617 mg) was suspended in ethyl acetate/methanol (5:1, 30 mL) and was treated with $SnCl_2$ (380 mg, 2.0 equiv). After stirring for 2 hours at 80 ^{O}C, the reaction mixture was cooled down to room temperature and poured onto saturated aqueous $NaHCO_3$. The mixture was stirred for 10 min followed by extraction with chloroform and 2-propanol (4:1, 100 mL). The organic layer was washed with water and brine, dried over sodium sulfate, filtered through a pad of celite and concentrated under reduced pressure. The resulting crude was purified by flash column with CH_2Cl_2/methanol (10/1) to provide the compound **7** (290 mg, 50%). LC-MS: *m/z* (M+H) 587. ^{1}H NMR (500 MHz, DMSO) δ 8.54 (s, 1H), 8.40 (s, 1H), 8.32 (d, *J* = 7.9 Hz, 1H), 8.09 – 8.03 (m, 2H), 8.00 (d, *J* = 8.3 Hz, 1H), 7.72 (t, *J* = 7.5 Hz, 1H), 7.62 (t, *J* = 7.9 Hz, 2H), 7.47 – 7.40 (m, 1H), 7.34 (t, *J* = 7.3 Hz, 1H), 7.20 (d, *J* = 8.2 Hz, 2H), 6.88 (s, 4H), 6.70 (d, *J* = 7.9 Hz, 2H), 4.07 (d, *J* = 11.7 Hz, 1H), 3.92 (s, 1H), 3.80 (d, *J* = 12.5 Hz, 1H), 3.10 (dt, *J* = 22.1, 7.5 Hz, 2H), 2.12 – 1.96 (m, 1H), 1.82 (dd, *J* = 9.0, 4.4 Hz, 1H), 1.67 (ddd, *J* = 16.5, 11.9, 3.8 Hz, 1H), 1.59 – 1.45 (m, 1H).^{13}C NMR (125 MHz, DMSO) δ 170.41, 160.37, 158.53, 156.41, 146.65, 137.37, 135.40, 134.49, 130.44 (2C), 129.78, 129.20 (2C), 129.05, 127.33 (2C), 126.32, 125.95, 124.54, 123.72, 117.94, 115.57 (2C), 115.44, 113.41, 49.66, 48.26, 45.39, 30.52, 23.96.



7 8

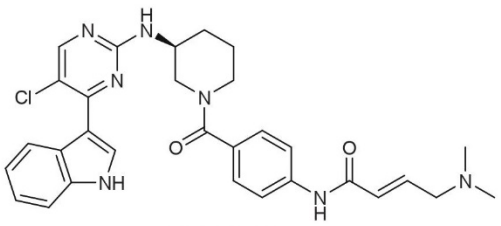
(R,E)-N-(4-(3-((5-chloro-4-(1-(phenylsulfonyl)-1H-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)-4-(dimethylamino)but-2-enamide (8) To the solution of the compound 7 (60 mg) in acetonitrile (10 mL) was added diisopropylethylamine (13 mg, 1.0 equiv). The reaction mixture was cooled down to 0 ^{O}C and then treated with 4-chlorobut-2-enoyl chloride (54 mg, 3.0 equiv) in CH_2Cl_2. The reaction mixture was stirred for 10 min at 0 ^{O}C followed by adding dimethylamine in THF (1M, 2.0 mL). The reaction mixture was then warmed up to room temperature, stirred for 1 hour and then concentrated under reduced pressure. The resulting crude was purified by preparative HPLC to give the product **8** (38 mg, 55%) LC-MS: *m/z* (M+H) 698. ^{1}H NMR (500 MHz, DMSO) δ 10.38 (s, 1H), 8.53 (s, 1H), 8.37 (s, 1H), 8.30 (d, *J* = 6.9 Hz, 1H), 8.08 – 8.02 (m, 2H), 7.99 (d, *J* = 8.3 Hz, 1H), 7.72 (t, *J* = 7.5 Hz, 1H), 7.70 – 7.58 (m, 4H), 7.47 – 7.40 (m, 1H), 7.39 – 7.27 (m, 4H), 6.88 – 6.80 (m, 1H), 6.54 (d, *J* = 15.4 Hz, 1H), 5.95 (s, 1H), 3.95 (d, *J* = 6.6 Hz, 2H), 3.91 (s, 1H), 3.67 (d, *J* = 25.4 Hz, 1H), 3.29 – 3.09 (m, 2H), 2.81 (s, 6H), 2.56 (d, *J* = 8.7 Hz, 1H), 2.11 – 1.98 (m, 1H), 1.85 (s, 1H), 1.70 (dd, *J* = 9.4, 3.5 Hz, 1H), 1.54 (dd, *J* = 9.7, 3.8 Hz, 1H). ^{13}C NMR (125 MHz, DMSO) δ 169.74, 162.73, 160.33, 158.51, 156.38, 140.22, 137.35, 135.41, 134.47, 132.74, 131.95, 131.80, 130.44 (2C), 129.81, 129.04, 128.15 (2C), 127.32 (2C), 125.94, 124.55, 123.68, 119.42 (2C), 117.89, 115.50, 113.41, 57.35, 49.48, 48.17, 44.88, 42.54 (2C), 30.27, 23.65



7 1 (THZ531)

(R,E)-N-(4-(3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)-4-(dimethylamino)but-2-enamide (1) The compound 8 (70 mg) was dissolved in 1,4-dioxane (5 mL) and aqueous NaOH (1.0 M, 5 mL). The solution was allowed to stir at room temperature for 2 hours followed by neutralization with HCl (1M, 5 mL). The solution was then extracted with chloroform and 2-propanol (4:1, 30 mL) and was washed with water. The removal of solvent provided the crude which was purified by HPLC to give the final product **1** (25 mg, 43%). LC-MS 558 (M+1), ^{1}H NMR (500 MHz, DMSO-*d6*) 11.83 (s, 1H), 10.94 (s, 1H), 10.41 (s, 1H), 8.53 (s, 1H), 8.44 (s, 1H), 8.25 (s, 1H), 7.67 (d, *J* = 7.0 Hz, 1H), 7.52 (d, *J* = 8.0 Hz, 1H), 7.36 (d, *J* = 8.5 Hz, 2H), 7.22 (t, *J* = 8.0 Hz, 1H), 7.16 (d, *J* = 8.0 Hz, 1H), 6.83 (m, 1H), 6.52 (d, *J* = 15.0 Hz, 1H), 4.02 (br, 2H), 3.93-3.80 (m, 3H), 3.20 (br, 2H), 2.80 (s, 6H), 2.10 (m, 1H), 1.86 (m, 1H), 1.71 (m, 1H), 1.61 (m, 1H). ^{13}C NMR (126 MHz, DMSO) δ 169.89, 162.78, 160.36, 157.54, 153.27, 140.36, 136.78, 132.99, 132.70, 131.99, 131.65, 128.11 (2C), 126.74, 123.20, 122.96, 121.69, 119.42 (2C), 113.66, 112.59, 111.53, 57.19, 49.12, 48.30, 45.07, 42.31 (2C), 30.21, 23.51.

(*S*,*E*)-*N*-(4-(3-((5-chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)-4-(dimethylamino)but-2-enamide

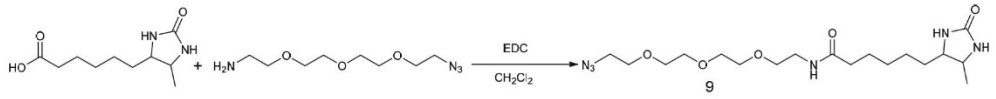


4 (THZ532)

Same synthetic method as compound 1 (THZ531)

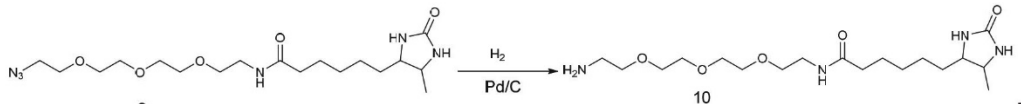
(S,E)-N-(4-(3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)-4-(dimethylamino)but-2-enamide (4) LC-MS 558 (M+1), ^{1}H NMR (500 MHz, DMSO-*d6*) 11.83 (s, 1H), 10.94 (s, 1H), 10.41 (s, 1H), 8.53 (s, 1H), 8.44 (s, 1H), 8.25 (s, 1H), 7.67 (d, *J* = 7.0 Hz, 1H), 7.52 (d, *J* = 8.0 Hz, 1H), 7.36 (d, *J* = 8.5 Hz, 2H), 7.22 (t, *J* = 8.0 Hz, 1H), 7.16 (d, *J* = 8.0 Hz, 1H), 6.83 (m, 1H), 6.52 (d, *J* = 15.0 Hz, 1H), 4.02 (br, 2H), 3.93-

3.80 (m, 3H), 3.20 (br, 2H), 2.80 (s, 6H), 2.10 (m, 1H), 1.86 (m, 1H), 1.71 (m, 1H), 1.61 (m, 1H). ^{13}C NMR (126 MHz, DMSO) δ 169.89, 162.78, 160.36, 157.54, 153.27, 140.36, 136.78, 132.99, 132.70, 131.99, 131.65, 128.11 (2C), 126.74, 123.20, 122.96, 121.69, 119.42 (2C), 113.66, 112.59, 111.53, 57.19, 49.12, 48.30, 45.07, 42.31 (2C), 30.21, 23.51.



9

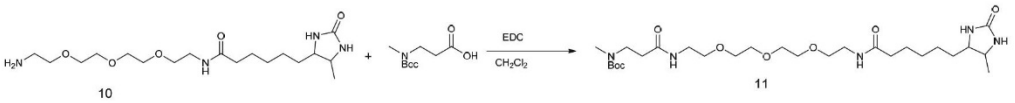
N-(2-(2-(2-(2-aminoethoxy)ethoxy)ethoxy)ethyl)-6-(5-methyl-2-oxoimidazolidin-4-yl)hexanamide (10) Desthiobiotin (430 mg, 2.0 mmol), 11-Azido-3,6,9-trioxaundecan-1-amine (435 mg, 2.0 mmol) and 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (576 mg, 3.0 mmol), were dissolved in methylene chloride (10 mL) and trimethylamine (380 uL, 3.0 mmol) was added at 0°C. The mixture was stirred at room temperature for 6 hours. The reaction solution was added with water and extracted with chloroform and isopropanol (4:1). The organic layer was dried with sodium sulfate and the solvent was removed under the reduced pressure to give a crude product which was used directly in the next step.



9 10

N-(2-(2-(2-(2-aminoethoxy)ethoxy)ethoxy)ethyl)-6-(5-methyl-2-oxoimidazolidin-4-yl)hexanamide the crude from above reaction was dissolved in methanol (10 mL) and then 10% Pd on carbon (50 mg) was added. The mixture solution was stirred for 3 hours with hydrogen balloon at room temperature. The solution was filtered to remove catalyst and solvent was removed under reduced pressure. The product was purified by silica gel column with CH_2Cl_2/methanol (10/1) to provide the product **10** (233 mg, 30%). LC-MS: *m/z* (M+H) 389. ^{1}H NMR (500 MHz, MeOD) δ 3.85 (dd, *J* = 7.8, 6.5 Hz, 1H), 3.76 – 3.70 (m, 1H), 3.71 – 3.64 (m, 8H), 3.62 – 3.50 (m, 4H), 3.40 (d, *J* = 5.5 Hz, 1H), 3.34 (dt, *J* = 3.3, 1.6 Hz, 1H), 2.81 (dd, *J* =

7.2, 3.4 Hz, 2H), 2.25 (dd, *J* = 15.2, 7.9 Hz, 2H), 1.72 – 1.60 (m, 2H), 1.59 – 1.50 (m, 2H), 1.50 – 1.27 (m, 4H), 1.14 (d, *J* = 6.5 Hz, 3H).

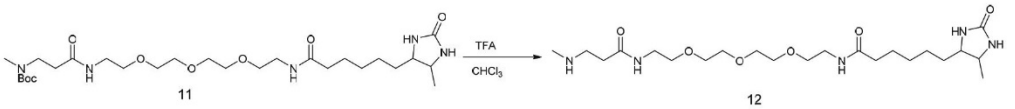


10 11

tert-butyl methyl(22-(5-methyl-2-oxoimidazolidin-4-yl)-3,17-dioxo-7,10,13-trioxa-4,16-diazadocosyl)carbamate (11) The compound **10** (388 mg, 1.0 mmol) 3-((tert-butoxycarbonyl)(methyl)amino)propanoic acid (260 mg, 1.3 mmol) and 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (370 mg, 2.0 mmol) in methylene chloride (15 mL) and then trimethylamine (270 uL, 2.0 mmol) was added at 0°C. The mixture solution was stirred for 6 hours. The solvent was removed under the reduced pressure and the residue was purified by purified by silica gel column with CH_2Cl_2/methanol (10/1) to provide the product **11** (120mg, 21%) LC-MS: *m/z* (M+H) 574.37

^{1}H NMR (500 MHz, MeOD) δ 3.85 (dd, *J* = 7.9, 6.5 Hz, 1H), 3.76 – 3.62 (m, 8H), 3.59 – 3.50 (m, 5H), 3.40 (d, *J* = 5.5 Hz, 3H), 3.34 (dt, *J* = 3.3, 1.6 Hz, 1H), 3.21 – 3.12 (m, 1H), 2.89 (s, 3H), 2.45 (t, *J* = 6.9 Hz, 2H), 2.36 (s, 2H), 2.24 (t, *J* = 7.5 Hz, 2H), 1.69 (dq, *J* = 29.6, 7.5 Hz, 2H), 1.53 (ddd, *J* = 7.5, 5.2, 2.7 Hz, 2H), 1.49 (s, 9H), 1.45 – 1.30 (m, 3H), 1.15 – 1.10 (m, 3H).

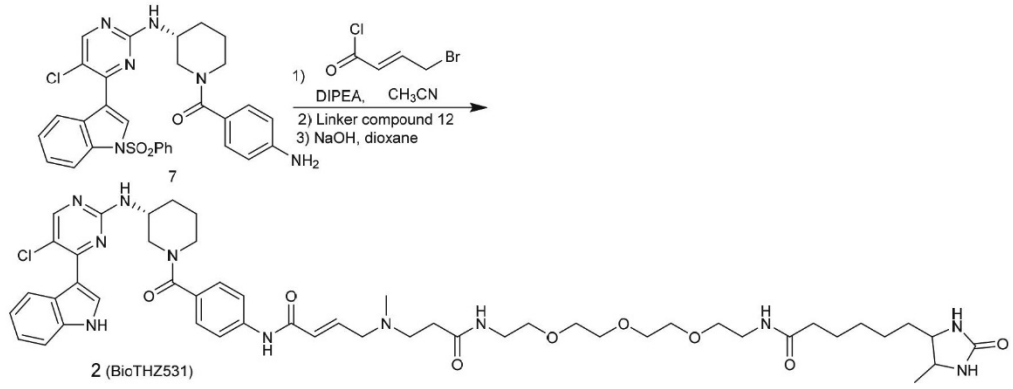
.



11 12

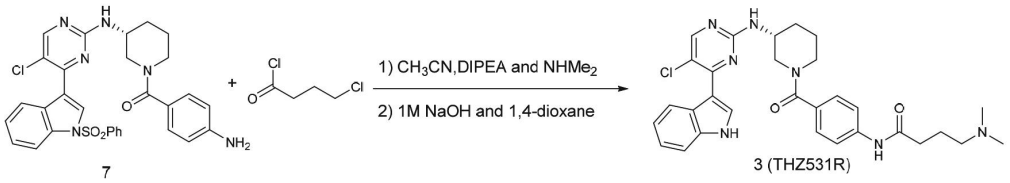
6-(5-methyl-2-oxoimidazolidin-4-yl)-N-(5-oxo-9,12,15-trioxa-2,6-diazaheptadecan-17-yl)hexanamide Crude compound 11 was dissolved in chloroform (10 mL) and then trifluoroacetic acid (5 mL) was added. The solution was stirred at room temperature for 1 hour and then solvent was removed under the reduced pressure. The residue was purified by

preparative LC-MS with methanol and water (1% TFA) as eluent to provide compound 11 (165 mg, 35% for two steps). LC-MS: *m/z* (M+H) 474.



N-((E)-20-((4-((R)-3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)amino)-16-methyl-13,20-dioxo-3,6,9-trioxa-12,16-diazaicos-18-en-1-yl)-6-(5-methyl-2-oxoimidazolidin-4-yl)hexanamide (2) To the solution of the compound **7** (60 mg) in acetonitrile (10 mL) was added diisopropylethylamine (13 mg, 1.0 equiv). The reaction mixture was cooled down to 0 ℃ and then treated with 4-chlorobut-2-enoyl chloride (54 mg, 3.0 equiv) in CH_2Cl_2. The reaction mixture was stirred for 10 min at 0 ^{O}C and propanol (4:1, 30 mL) and was washed with water. The removal of solvent provided the crude which was used then dissolved in THF (3.0 mL). To this solution, the free amino compound 12 (94.0 mg, 0.2 mmol) was added and the solution were then heated for 2 hours at 50℃. The reaction mixture was then cooled down to room temperature. The solvent was removed under reduced pressure and the residue was dissolved again in 1,4-dioxane (2.0 mL) followed by adding 1M NaOH solution (2.0 mL). The solution was stirred at room temperature for 2 hours and then was quenched with 1M HCl solution (2.0mL), extracted with chloroform/2-propanol (4/1, vol/vol, 20 mL). The organic layer was washed with water, brine and dried over sodium sulfate. The crude after removing the solvent was purified by HPLC to provide BioTHZ531 as TFA salt (15.0 mg, 15%) LC-MS: *m/z* (M+H) 987,^{1}H NMR (500 MHz, DMSO-*d6*) 11.67 (s, 1H), 10.23 (s, 1H), 8.54 (s, 1H), 8.39 (s,

1H), 8.21 (s, 1H), 8.05 (s, 1H), 7.65 (d, *J* = 7.0 Hz, 2H), 7.55 (br, 1H), 7.51 (d, *J* = 8.0 Hz, 1H), 7.38 (d, *J* = 8.0 Hz, 2H), 7.22 (t, *J* = 8.0 Hz, 1H), 7.16 (t, *J* = 8.0 Hz, 1H), 7.06 (br, 1H), 6.83 (m, 1H), 6.52 (d, *J* = 15.0 Hz, 1H), 4.20-3.20 (m, 20H), 2.80 (s, 3H), 2.65 (t, *J* = 7.0 Hz, 2H), 2.08 (m, 3H), 1.86 (m, 1H), 1.71 (m, 1H), 1.58 (m, 1H), 1.51 (m, 2H), 1.40-1.22 (m, 4H), 1.00 (d, *J* = 6.0 Hz, 3H). ^{13}C NMR (125 MHz, DMSO-d6) δ 173.76, 170.70, 169.77, 163.10, 160.22, 158.68, 158.38, 157.41, 140.41, 136.63, 131.44, 131.08, 128.15, 126.81, 123.27, 122.76, 121.01, 119.18, 114.07, 112.24, 111.66, 55.75, 55.54, 55.20, 50.81, 48.27, 35.77, 34.38, 33.41, 30.63, 29.99, 29.21, 25.96, 25.30, 24.00, 19.75, 15.90.



(R)-N-(4-(3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)piperidine-1-carbonyl)phenyl)-4-(dimethylamino)butanamide (3) The free amino compound **7** (60mg) was dissolved in acetonitrile (2.0 mL) was added diisopropylethylamine (20.0 uL) and 4-chlorobutanoyl chloride (30 mg, 2.0 equiv) in dichloromethane (1.0 mL). After stirring for 5 min, dimethylamine (1M, 2.0 mL)) in THF was added and the reaction mixture was then heated to 50 °C for 2 hours. After removal of solvent, the crude was then treated with 1,4-dioxane (2.0 mL) and 1 M NaOH solution (2.0 mL) and then the solution was stirred at room temperature for 2 hours. After quench with 1M HCl (2.0 mL), the solution was extracted with chloroform/2-propanol (4:1, 20 mL). The organic layer was washed with water, brine and dried with $MgSO_4$.The concentration under reduced pressure to give the crude produce which was purified by HPLC to provide THZ531R (**3**) as TFA salt (16.0 mg, 60%) LC-MS: *m/z* (M+H) 560 ^{1}H NMR (500 MHz, DMSO-*d6*) 11.67 (s, 1H), 9.95 (s, 1H), 9.34(br, 1H), 8.53 (s, 1H), 8.40 (s, 1H), 8.21 (s, 1H), 7.67 (d, *J* = 7.0 Hz, 1H), 7.52 (d, *J* = 8.0 Hz, 1H), 7.36 (d, *J* = 8.5 Hz, 2H), 7.22 (t, *J* = 8.0 Hz, 1H), 7.16 (d, *J* = 8.0 Hz, 1H), 7.04 (br, 1H), 3.99 (br, 2H), 3.13 (m, 5H), 2.80 (s, 6H), 2.44 (t, J =7.0 Hz), 2.10

(m, 1H), 1.98 (m, 2H), 1.85 (m, 1H), 1.71 (m, 1H), 1.58 (m, 1H). ^{13}C NMR (126 MHz, DMSO) δ 170.66, 169.81, 160.24, 158.66, 157.46, 140.40, 136.62, 131.43, 131.07, 128.14 (2C), 126.80, 123.25, 122.78, 121.03, 119.18 (2C), 114.09, 112.24, 111.65, 57.14, 49.40, 48.27, 45.06, 42.99 (2C), 33.36, 30.61, 23.97, 20.17.

6.3 References

[1] Doonan, J. H.; Kitsios, G. Functional Evolution of Cyclin-Dependent Kinases. *Mol. Biotechnol.* **2009**, *42* (1), 14-29.

[2] Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **2014**, *15* (6), 122.

[3] Wood, D. J.; Endicott, J. A. Structural insights into the functional diversity of the CDK – cyclin family. *Open Biol.* **2018**, *8* (9), 180112.

[4] Malumbres, M.; Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **2005**, *30* (11), 630-641.

[5] Hochegger, H.; Takeda, S.; Hunt, T. Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (11), 910-916.

[6] Palmer, N.; Kaldis, P. Less-well known functions of cyclin/CDK complexes. *Semin. Cell Dev. Biol.* **2020**, *107*, 54-62.

[7] Parua, P. K.; Fisher, R. P. Dissecting the Pol II transcription cycle and derailing cancer with CDK inhibitors. *Nat. Chem. Biol.* **2020**, *16* (7), 716-724.

[8] Lienhard Schmitz, M.; Kracht, M. Cyclin-Dependent Kinases as Coregulators of Inflammatory Gene Expression. *Trends Pharmacol. Sci.* **2016**, *37* (2), 101-113.

[9] Hydbring, P.; Malumbres, M.; Sicinski, P. Non-canonical functions of cell cycle cyclins and cyclin-depent kinases. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (5), 280-292.

[10] Rice, A. P. Cyclin-Dependent Kinases as Therapeutic Targets for HIV-1 Infection. *Expert Opin. Ther. Targets* **2016**, *20* (12), 1453-1461.

[11] Wells, A. D.; Morawski, P. A. New roles for cyclin-dependent kinases in T cell biology: linking cell division and differentiation. *Nat. Rev. Immunol.* **2014**, *14* (4), 261-270.

[12] Jirawatnotai, S.; Dalton, S.; Wattanapanitch, M. Role of cyclins and cyclin-dependent kinases in pluripotent stem cells and their potential as a therapeutic target. *Semin. Cell Dev. Biol.* **2020**, *107*, 63-71.

[13] Krasinska, L.; Fisher, D. Non-Cell Cycle Functions of the CDK Network in Ciliogenesis: Recycling the Cell Cycle Oscillator. *Bioessays* **2018**, *40* (6), e1800016.

[14] Palmer, N.; Talib, S. Z. A.; Kaldis, P. Diverse roles for CDK-associated activity during spermatogenesis. *FEBS Lett.* **2019**, *593* (20), 2925-2949.

[15] Cortés, N.; Guzmán-Martínez, L.; Andrade, V.; González, A.; Maccioni, R. B. CDK5: A Unique CDK and Its Multiple Roles in the Nervous System. *J. Alzheimers Dis.* **2019**, *68* (3), 843-855.

[16] Bhounsule, A. S.; Bhatt, L. K.; Prabhavalkar, K. S.; Oza, M. Cyclin dependent kinase 5: A novel avenue for Alzheimer's disease. *Brain Res. Bull.* **2017**, *132*, 28-38.

[17] Chohan, T. A.; Qayyum, A.; Rehman, K.; Tariq, M.; Akash, M. S. H. An insight into the emerging role of cyclin-dependent kinase inhibitors as potential therapeutic agents for the treatment of advanced cancers. *Biomed. Pharmacother.* **2018**, *107*, 1326-1341.

[18] Asghar, U.; Witkiewicz, A. K.; Turner, N. C.; Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat. Rev. Drug Discov.* **2015**, *14* (2), 130-146.

[19] Vijayaraghavan, S.; Moulder, S.; Keyomarsi, K.; Layman, R. M. Inhibiting CDK in Cancer Therapy: Current Evidence and Future Directions. *Target Oncol.* **2018**, *13* (1), 21-38.

[20] Greenleaf, A. L. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millennium. *Transcription* **2019**, *10* (2), 91-110.

[21] Kohoutek, J.; Blazek, D. Cyclin K goes with Cdk12 and Cdk13. *Cell Div.* **2012**, *7*, 12.

[22] Bösken, C. A.; Farnung, L.; Hintermair, C.; Merzel Schachter, M.; Vogel-Bachmayr, K.; Blazek, D.; Anand, K.; Fisher, R. P.; Eick, D.; Geyer, M. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* **2014**, *5*, 3505.

[23] Dixon-Clarke, S. E.; Elkins, J. M.; Cheng, S.-W. G.; Morin, G. B.; Bullock, A. N. Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding. *Sci. Rep.* **2015**, *5*, 17122.

[24] Greifenberg, A. K.; Hönig, D.; Pilarova, K.; Düster, R.; Bartholomeeusen, K.; Bösken, C. A.; Anand, K.; Blazek, D.; Geyer, M. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep.* **2016**, *14* (2), 320-331.

[25] Blazek, D.; Kohoutek, J.; Bartholomeeusen, K.; Johansen, E.; Hulinkova, P.; Luo, Z.; Cimermancic, P.; Ule, J.; Peterlin, B. M. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **2011**, *25* (20), 2158-2172.

[26] Dubbury, S. J.; Boutz, P. L.; Sharp, P. A. Cdk12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* **2018**, *564* (7734), 141-145.

[27] Li, X.; Chatterjee, N.; Spirohn, K.; Boutros, M.; Bohmann, D. Cdk12 Is A Gene-Selective RNA Polymerase II Kinase That Regulates a Subset of the Transcriptome, Including Nrf2 Target Genes. *Sci. Rep.* **2016**, *6*, 21455.

[28] Henry, K. L.; Kellner, D.; Bajrami, B.; Anderson, J. E.; Beyna, M.; Bhisetti, G.; Cameron, T.; Capacci, A. G.; Bertolotti-Ciarlet, A.; Feng, J.; Gao, B.; Hopkins, B.; Jenkins, T.; Li, K.; May-Dracka, T.; Murugan, P.; Wei, R.; Zeng, W.; Allaire, N.; Buckler, A.; Loh, C.; Juhasz, P.; Lucas, B.; Ennis, K. A.; Vollman, E.; Cahir-McFarland, E.; Hett, E. C.; Ols, M. L. CDK12-mediated transcriptional regulation of noncanonical NF-κB components is essential for signaling. *Sci. Signal.* **2018**, *11* (541), eaam8216.

[29] Choi, S. H.; Martinez, T. F.; Kim, S.; Donaldson, C.; Shokhirev, M. N.; Saghatelian, A.; Jones, K. A. CDK12 phosphorylates 4E-BP1 to enable mTORC1-dependent translation and mitotic genome stability. *Genes Dev.* **2019**, *33* (7-8), 418-435.

[30] Manavalan, A. P. C.; Pilarova, K.; Kluge, M.; Bartholomeeusen, K.; Rajecky, M.; Oppelt, J.; Khirsariya, P.; Paruch, K.; Krejci, L.; Friedel, C. C.; Blazek, D. CDK12 controls G1/S progression by regulating RNAII processivity at core DNA replication genes. *EMBO Rep.* **2019**, *20* (9), e47592.

[31] Fan, Z.; Devlin, J. R.; Hogg, S. J.; Doyle, M. A.; Harrison, P. F.; Todorovski, I.; Cluse, L. A.; Knight, D. A.; Sandow, J. J.; Gregory, G.; Fox, A.; Beilharz, T. H.; Kwiatkowski, N.; Scott, N. E.; Tufegdzic Vidakovic, A.; Kelly, G. P.; Svejstrup, J. Q.; Geyer, M.; Gray, N. S.; Vervoort, S. J.; Johnstone, R. W. CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Sci. Adv.* **2020**, *6* (18), eaaz5041.

[32] Chen, H.-R.; Lin, G.-T.; Huang, C.-K.; Fann, M.-J. Cdk12 and Cdk13 regulate axonal elongation through a common signaling pathway that modulates Cdk5 expression. *Exp. Neurol.* **2014**, *261*, 10-21.

[33] Quereda, V.; Bayle, S.; Vena, F.; Frydman, S. M.; Monastyrskyi, A.; Roush, W. R.; Duckett, D. R. Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell* **2019**, *36* (5), 545-558.e7.

[34] Zeng, M.; Kwiatkowski, N. P.; Zhang, T.; Nabet, B.; Xu, M.; Liang, Y.; Quan, C.; Wang, J.; Hao, M.; Palakurthi, S.; Zhou, S.; Zeng, Q.; Kirschmeier, P. T.; Meghani, K.; Leggett, A. L.; Qi, J.; Shapiro, G. I.; Liu, J. F.; Matulonis, U. A.; Lin, C. Y.; Konstantinopoulos, P. A.; Gray, N. S. Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13. *eLife* **2018**, *7*, e39030.

[35] Geng, M.; Yang, Y.; Cao, X.; Dang, L.; Zhang, T.; Zhang, L. Targeting CDK12-mediated transcription regulation in anaplastic thyroid carcinoma. *Biochem. Biophys. Res. Commun.* **2019**, *520* (3), 544-550.

[36] Nguyen, B.; Mota, J. M.; Nandakumar, S.; Stopsack, K. H.; Weg, E.; Rathkopf, D.; Morris, M. J.; Scher, H. I.; Kantoff, P. W.; Gopalan, A.; Zamarin, D.; Solit, D. B.; Schultz, N.; Abida, W. Pan-cancer Analysis of *CDK12* Alterations Identifies a Subset of Prostate Cancers with Distinct Genomic and Clinical Characteristics. *Eur. Urol.* **2020**, *78* (5), 671-679.

[37] Wang, C.; Wang, H.; Lieftink, C.; du Chatinier, A.; Gao, D.; Jin, G.; Jin, H.; Beijersbergen, R. L.; Qin, W.; Bernards, R. CDK12 inhibition mediates DNA damage and is synergistic with sorafenib treatment in hepatocellular carcinoma. *Gut* **2020**, *69* (4), 727-736.

[38] Iniguez, A. B.; Stolte, B.; Wang, E. J.; Conway, A. S.; Alexe, G.; Dharia, N. V.; Kwiatkowski, N.; Zhang, T.; Abraham, B. J.; Mora, J.; Kalev, P.; Leggett, A.; Chowdhury, D.; Benes, C. H.; Young, R. A.; Gray, N. S.; Stegmaier, K. EWS/FLI confers tumor cell synthetic lethality to CDK12 inhibition in Ewing sarcoma. *Cancer Cell* **2018**, *33* (2), 202-216.e6.

[39] Liu, M.; Fan, H.; Li, T.; Sihong, L..; Qiao, S.; Bi, J. Low expression of CDK12 in gastric cancer is correlated with advanced stage and poor outcome. *Pathol. Res. Pract.* **2020**, *216* (7), 152962.

[40] Liang, S.; Hu, L.; Wu, Z.; Chen, Z.; Liu, S.; Xu, X.; Qian, A. *CDK12*: A Potent Target and Biomarker for Human Cancer Therapy. *Cells* **2020**, *9* (6), 1483.

[41] Paculová, H.; Kohoutek, J. The emerging roles of CDK12 in tumorigenesis. *Cell Div.* **2017**, *12*, 7.

[42] Choi, S. H.; Kim, S.; Jones, K. A. Gene expression regulation by CDK12: a versatile kinase in cancer with functions beyond CTD phosphorylation. *Exp. Mol. Med.* **2020**, *52* (5), 762-771.

[43] Chilà, R.; Guffanti, F.; Damia, G. Role and therapeutic potential of CDK12 in human cancers. *Cancer Treat. Rev.* **2016**, *50*, 83-88.

[44] Krajewska, M.; Dries, R.; Grassetti, A. V.; Dust, S.; Gao, Y.; Huang, H.; Sharma, B.; Day, D. S.; Kwiatkowski, N.; Pomaville, M.; Dodd, O.; Chipumuro, E.; Zhang, T.; Greenleaf, A. L.; Yuan, G.-C.; Gray, N. S.; Young, R. A.; Geyer, M.; Gerber, S. A.; George, R. E. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat. Commun.* **2019**, *10* (1), 1757.

[45] Manz, T. „Synthesis of biotinylated CDK12 inhibitors for ALPHA-Screenings and first steps in the development of potent and selective dual CDK12/EGFR inhibitors for novel anti-tumor treatment“; Diplom thesis, Universität des Saarlandes, **2015**.

7 Novel Chemical Probes for the Study of the Peptidyl-prolyl Isomerase Pin1

7.1 Introduction

7.1.1 Determining Proteome-wide Selectivity of a Covalent Probe *via* a Novel Desthiobiotin-mediated Pulldown Technique

Covalent protein binders, being equipped with a reactive, electrophilic warhead, are important chemical probes for the labeling and study of their targets and pose interesting starting points in the development of novel small molecule treatments for various diseases (see also Chapters 5 and 6). While the covalent binding of probes can generally increase their selectivity, e.g., by targeting a unique cysteine, the inherent reactivity of certain warheads can also lead to unintended binding effects within a cell. Thus, proteome-wide selectivity profiling of reactive probes is essential to uncover potential off-target activities.

A novel approach for determining the cellular selectivity of covalent probes is the CITe-Id technology, which was introduced by Browne *et al.* in 2019 and quantifies a probe's proteome-wide cysteine reactivity using respective desthiobiotin-tagged probe analogs[1] (the biotin/desthiobiotin-streptavidin technology was already introduced in Chapter 5). Pre-treatment of live cells or lysates with the unlabeled parent compound in a competition format further enables the differentiation of specific and non-specific labeling events[1]. In contrast to classical ABPP techniques (further described in Chapter 10), in which competition with promiscuous cysteine-reactive probes such as iodoacetamide-alkyne reflects the proteome-wide cysteine selectivity of covalent compounds, CITe-Id enriches for compound-specific binding sites, resulting in a more direct and simplified approach for profiling covalent protein binders[1].

Figure 7.1.1 gives an overview of the general experimental set-up of a CITe-Id profiling. Cell lysates are treated with a compound of interest at varying doses, followed by addition of its respective desthiobiotin-labeled analog[1]. The samples are then subjected to tryptic digest and labeled protein fragments are enriched *via* streptavidin-mediated pulldown of the desthiobiotin-labeled target-probe complexes[1]. Using iTRAQ methods (see also Chapter 10), the different aliquots are further labeled with stable isotope reagents, which can then be analyzed by mass spectrometry[1]. Using this approach, the dose-dependent binding of the covalent parent compound for specific cysteine-thiols is determined and is used to differentiate between specific and non-specific binding effects[1].

Importantly, Browne and coworkers showed that the CITe-Id method was not only able to detect targets that are identified using other screening methods (including ABPP) but also that it revealed novel target proteins generally not identified with common ABPP techniques, highlighting the value of this method[1]. Moreover, due to the LC-MS/MS analysis step, which is also found among classical ABPP techniques, this approach can precisely pinpoint the labeled cysteine residues within the target proteins, which further helps to inform probe development and optimization purposes[1].

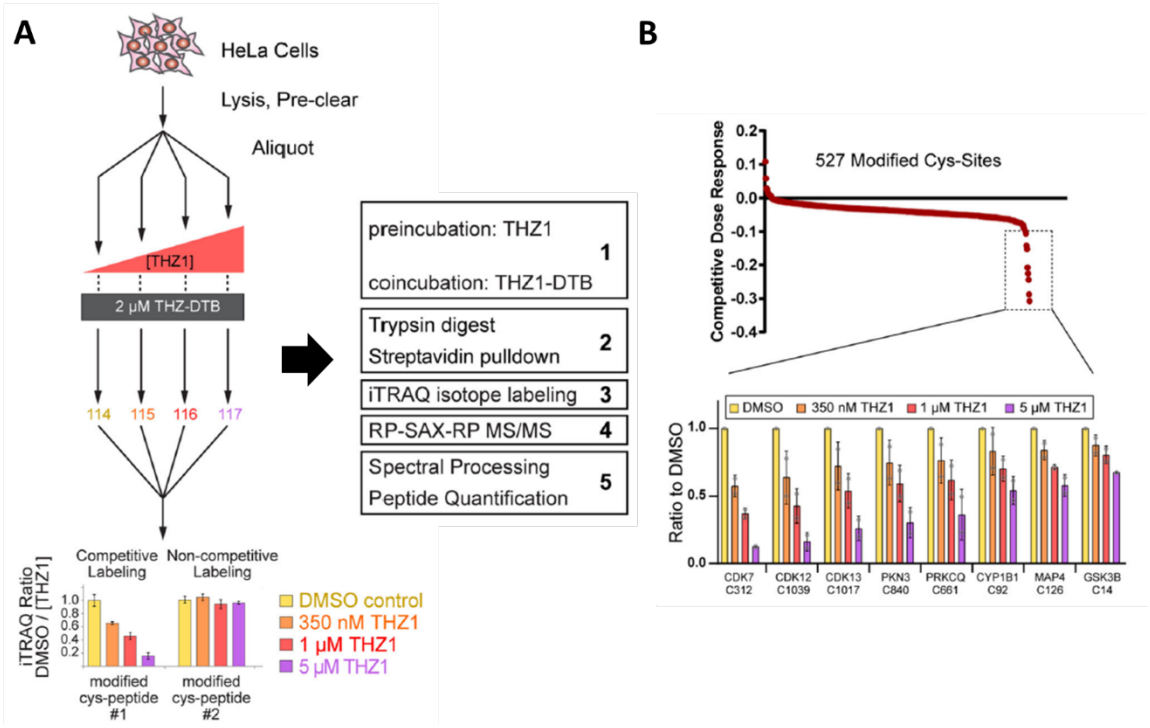


Figure 7.1.1 | Example of the CITe-Id Workflow[a]. A. Overview of the CITe-Id experimental set-up (here with the known kinase inhibitor THZ1 and its desthiobiotin analog, THZ1-DTB). In the first step, the parent compound (THZ1) was pre-incubated with aliquots of cell lysate (here: HeLa cells) at different concentrations. After addition and co-incubation with THZ1-DTB, the protein samples were trypsin digested, followed by a streptavidin-mediated pulldown of labeled protein fragments. Using the iTRAQ strategy, the resulting samples were labeled with isotope reagents consisting of varying isobaric tags, which label the N-termini and side chain amines of the peptides in the samples. Due to the differently labeled peptides, the samples of the different treatments can be combined and analyzed *via* mass spectrometry as one sample. **B.** The quantified read-outs are further analyzed in accordance with dose-response effects.
[a] Adapted from Browne *et al.*, *J. Am. Chem. Soc.* **2019**, *141* (1), 191-203[1].

7.1.2 The Peptidyl-prolyl Isomerase Pin1 – A Key Regulator of Cellular Processes

Peptidyl-prolyl isomerases (PPIases) catalyze the conformational switch between *cis*- and *trans*-isomers of peptidyl-prolyl peptide bonds[2,3] – a feature unique to proline (Pro) (Figure 7.1.2). This isomerization effect, which was first discovered in the 1980s, dictates protein folding processes and regulates the shape of native proteins[2]. There are three known families of PPIases, including the cyclophilins, the FK506-binding proteins (FKBPs), and the parvulins[2].

The Unique Case of Never-in-Mitosis (NIMA)-interacting Peptidyl-prolyl cis-trans Isomerase 1 (Pin1)
While cyclophilins and FKBPs typically form multidomain complexes, parvulins generally consist of only one catalytic PPIase domain[2], resulting in relatively small-sized proteins, which is the origin of their name (from Latin "*parvulus*", meaning "tiny")[3]. In the 1990s, a variety of parvulins were identified, including the yeast and human analogs Ess1 (essential protein 1) and Pin1, respectively[3]. Generally, Pin1 is found in all human tissues

at low levels, but is mainly expressed in the brain[3]. Intracellularly, Pin1 is normally localized in the nucleus, though in the context of certain cancers, Pin1 can also be found in the cytoplasm[3]. In addition, two more human parvulins, Par14 and Par17 (reflecting their sizes of 14 and 17 kDa, respectively), were identified, though they differ from Pin1 (~18 kDa[4]) in their substrate specificity as well as catalytic activity[3].

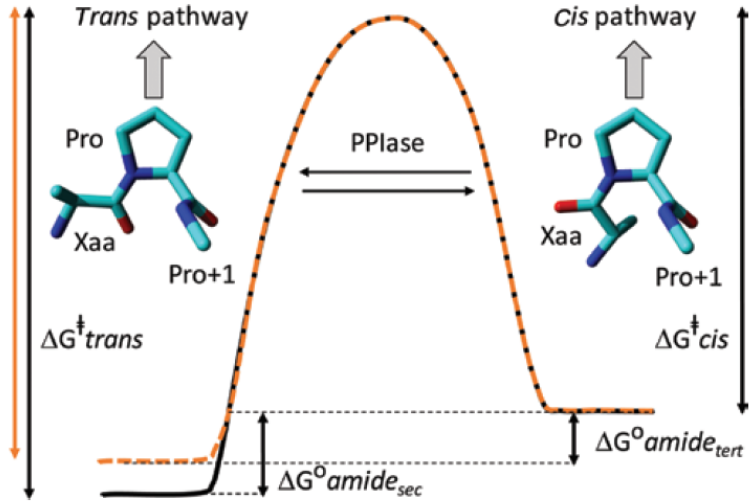


Figure 7.1.2 | Function of Peptidyl-prolyl Isomerases. PPIases catalyze the transitioning between the *cis*/*trans*-state of a prolyl moiety within a peptide bond. Taken from Matena *et al*., *Biol. Chem.* **2018**, *399* (2), 101-125[3].

Pin1 (Figure 7.1.3) has unique substrate-recognition specificity, being the only PPIase that acts on phosphorylated Ser/Thr-Pro motifs, thereby connecting kinase activity with PPIase function[5]. This selectivity for phosphorylated substrates is in part determined by Pin1's WW domain (derived from "Wechselwirkung", the German word for "interaction"), which mediates phosphate binding *via* a specific arginine residue[2]. The WW domain is linked to the C-terminal PPIase domain *via* a flexible linker region forming the catalytic cleft, which is structurally conserved among human parvulins[3]. The catalytic cleft is composed of hydrophobic amino acids, such as leucine, methionine, phenylalanine, and isoleucine, which make up the prolyl-binding pocket, as well as charged and polar residues, which interact with the phosphorylated serine or threonine residue of Pin1's endogenous substrates[3]. The active site also contains a cysteine (Cys) residue, Cys113, which is unique to Pin1[3] (Figure 7.1.4).

The catalytic activity of Pin1 (and human parvulins in general) was originally thought to be mediated by a covalent mechanism, here, *via* nucleophilic attack of the catalytic Cys113, to form a tetrahedral intermediate[3]. However, a non-covalent theory, in which the negative charge of the cysteine stabilizes the sp^2-character of the peptide carbonyl, thus supporting the sp^3-hybridization of the amide bond and enabling the rotation of the substrate peptide bond, was recently confirmed[3]. This mechanism also explains why parvulins can tolerate substitution of the cysteine for an aspartate residue[3]. Notably, besides Cys113, His59, Ser154, and His157 also support the catalytic mechanism by forming a defined hydrogen-bonding network[7].

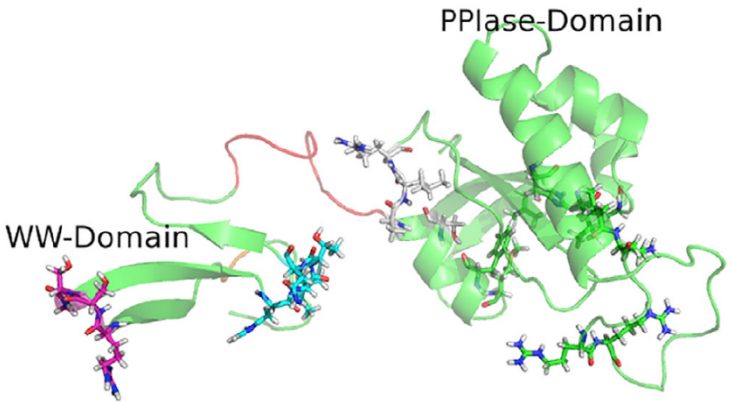


Figure 7.1.3 | Structure of Human Pin1. The PPIase and WW domains of the human peptidyl-prolyl isomerase Pin1 are connected *via* a flexible linker (red). Residues within the WW domain determine substrate specificity of the isomerase (magenta sticks) or facilitate domain-domain binding (cyan sticks). Catalytic and substrate-directing moieties (green sticks) are found within the PPIase domain. Residues depicted in gray sticks coordinate the formation of the domain-domain interface. Taken from Olsson *et al*., *Structure* **2016**, *24* (9), 1464-1475[6].

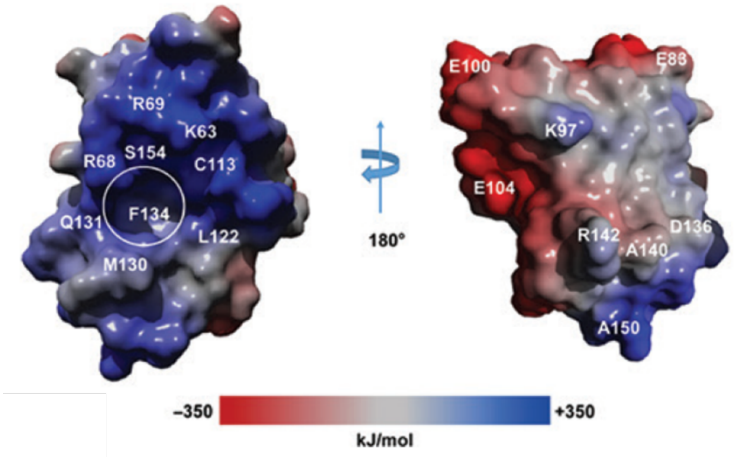


Figure 7.1.4 | Surface Charges of the Human Parvulin Pin1. The surface charges of the crystal structure of Pin1 (with a bound substrate; not shown in this graphic) are depicted as blue for positively charged and red for negatively charged areas. Neutral surface structures are shown in gray. Adapted from Matena *et al*., *Biol. Chem.* **2018**, *399* (2), 101-125[3].

As a key regulator of proline-directed phosphorylation signaling networks, Pin1's activity, stability, and localization are in turn regulated by various post-translational modifications including phosphorylation, oxidation, acetylation, ubiquitination, and SUMOylation (SUMO stands for "small ubiquitin-like modifiers")[7] (Figure 7.1.5). Interestingly, a single modification can lead to a vast variety of effects. For example, while the phosphorylation of Ser16 by protein kinase A (PKA) reduces Pin1's ability to bind certain substrates and thereby inhibits cell cycle progression, in a different context, Ser16 phosphorylation mediated by the cancer Osaka thyroid (COT) kinase increases Pin1's ability to bind other substrates, which ultimately results in increased cyclin D1 levels and elevated tumorigenesis[7]. These findings highlight the complexity of this regulatory system, which

orchestrates Pin1 function and is highly dependent on enzyme localization, cell type, and cell signaling networks[7].

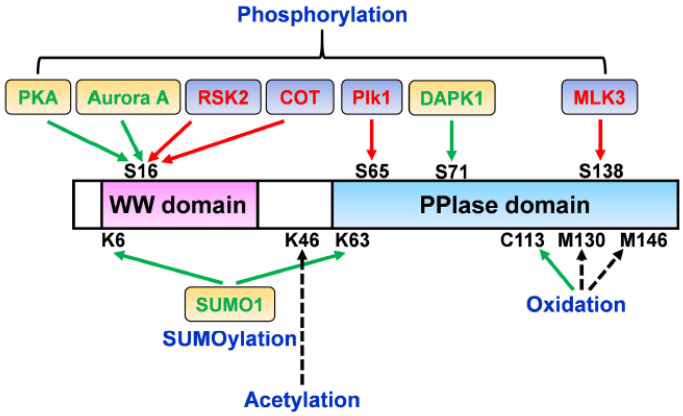


Figure 7.1.5 | Post-translational Modifications of Pin1. Pin1 activity is directed by a variety of different post-translational modifications including phosphorylation, SUMOylation, acetylation, and oxidation. Given are some examples of enzymes that regulate Pin1 activity, either having a positive (green) or negative (red) effect on Pin1 function. Taken from Chen *et al.*, *Front. Cell Dev. Biol.* **2020**, *8*, 179[7].

Pin1 Function in Health and Disease

Given the abundance of proline-adjacent Ser/Thr-motifs in proteins and the importance of proline-directed kinases and phosphatases, including CDKs, ERKs, Jun-N-terminal kinases (JNKs), PP2A, and others, Pin1 provides another layer of regulation for numerous phosphoproteins[8,9]. Pin1-mediated conformational changes in substrate proteins have been linked to several cellular effects such as cell cycle progression, stress and immune responses, as well as germ and neuronal cell development[8]. Consequently, deregulation of Pin1 has been implicated in a variety of related pathological conditions, including aging, asthma, infectious diseases, neurodegenerative conditions, and cancer[8].

As Pin1 was originally identified by its affinity for the essential mitotic kinase NIMA, thus controlling entry into mitosis in *Aspergillus nidulans*, the potential role of Pin1 in regulating cell cycle progression has thus been investigated[4]. In fact, several Pin1 substrates involved in the G2-M phase transition have been identified[10]. For example, GSK3β activity promotes hBora function in G2[10]. hBora in turn activates Aurora A, which stimulates PLK1 to activate the phosphatase CDC25, ultimately leading to activation of the cyclin-B/CDK1 complex and mitotic entry[10]. Pin1 regulates multiple nodes within this pathway; for instance, by facilitating hBora degradation, and by directly inhibiting CDC25 function[10]. Within the G2 phase, activated Aurora A leads to Pin1 inactivation *via* phosphorylation, which facilitates cell cycle progression[10]. Furthermore, Pin1-mediated inhibition of Wee1 kinase, a negative regulator of CDK1 function, promotes mitotic entry[10].

In addition to its role in the G2-M phase transition, Pin1 also regulates the G1-S phase transition[10]. For example, Pin1 activates and stabilizes cyclin D1, which is required for progression through G1[10]. Furthermore, Pin1 inactivates proteins that block cell cycle progression, such as the Rb protein, which binds and inhibits E2F to prevent the G1-S phase transition, and p27, a CDK2 inhibitor[10]. Notably, Pin1 also stabilizes p53 expression by

disrupting its interaction with the E3 ligase MDM2, therefore initiating cell cycle arrest or apoptosis[10].

Besides its role in controlling cell cycle progression, Pin1 also regulates transcription by affecting the localization, stability, and DNA-binding ability/affinity of transcription factors such as cyclin D1, β-catenin, NF-κB, the nuclear factor of activated T cells (NF-AT), Notch1, c-MYC, p53, and c-Jun[11]. Furthermore, Pin1 directly binds and isomerizes several motifs within the RNAPII CTD to regulate RNAPII activity during transcription initiation, elongation, and termination[12]. It has further been shown that Pin1 activity affects histone function[12] and is able to directly influence mRNA stability[13]. In summary, these findings highlight that Pin1 regulates multiple aspects of transcription.

Furthermore, Pin1 was found to play a role in metabolism[14]. For instance, it was shown that Pin1 promotes insulin signaling through its interaction with insulin receptor substrate 1 (IRS-1), leading to the activation of the PI3K/Akt pathway and ultimately to an increase in glucose uptake, for example in liver cells[14]. Interestingly, by activating JNK and S6 kinase (S6K), which inhibit IRS-1 activity, Pin1 also triggers insulin resistance under high-stress conditions[14]. Recent studies further implicated Pin1 function in promoting insulin secretion and the proliferation of pancreatic β-cells[15]. Moreover, some studies indicate that Pin1 is also involved in the regulation of gluconeogenesis to balance glucose levels in cell states under nutrient-deprived conditions as well as in adipogenesis through its interaction with peroxisome proliferator activated receptor γ (PPARγ) and AMPK[14]. Recently, it was shown that Pin1 facilitates degradation of the PR domain 16 (PRDM16) protein in adipocytes, which results in the downregulation of non-shivering thermogenesis[16].

Over the past decade, other tissue-specific effects of Pin1 have been described. In muscle cells, Pin1 was found to be involved in myogenesis and muscle regeneration processes by interacting with Smad3, a key player in regulating myoblast differentiation and fusion[17]. Through regulation of Smads, NF-κB, Wnt, TGF-β, and others, Pin1 orchestrates bone cell development and balances osteoclastogenesis as well as osteoblastogenesis[18]. Pin1 further facilitates spermatogenesis by promoting cell cycle progression[19].

Certain tissue-specific Pin1 activities have been associated with pathological events. For example, increased Pin1 expression in the myocardium as a result of cardiac injury, e.g. caused by increased pressure on the heart muscle, is connected to hypertrophy and is linked to Pin1's activating effects on Akt and MEK[20]. Furthermore, Pin1 was shown to be involved in hyperglycemia-mediated cardiovascular dysfunctions[21] and the development of atherosclerosis through regulation of NF-κB[22]. Pin1's function in regulating metabolism further suggests a role for Pin1 in facilitating the development of type 2 diabetes mellitus[15] as well as high-fat diet-induced obesity and non-alcoholic steatohepatitis[16,23]. In contrast, in neurodegenerative conditions such as Alzheimer's disease (AD), studies have implicated Pin1 as a key player in sustaining tau function and preventing tau aggregation, while also regulating GSK3β-induced Aβ-secretion, two essential processes that are dysregulated in AD[24]. Thus, reduced levels of functional Pin1 in the brain are associated with the development of AD[23]. Similarly, insufficient Pin1 activity in bone tissue is linked to osteoporosis[18].

Pin1 also regulates the immune system. For example, in dendritic cells, Pin1 affects CD8 (cluster of differentiation 8 antigen) expression and therefore the innate immune response[25]. Upregulated Pin1 function is thus associated with the development of systemic

lupus erythematosus, a severe autoimmune disease[26], asthma, and allergies[27]. In the context of viral infections, including feline coronaviruses[28], the human cytomegalovirus[29], and hepatitis B[30], Pin1 was found to facilitate viral replication. Another recent study suggests that Pin1 plays an essential role in lipopolysaccharide (LPS)-mediated activation of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase in neutrophils, causing pathogen-induced inflammation, which can develop into sepsis[31].

Finally, a lot of research has been dedicated to investigate Pin1's role in the development and severity of cancers such as breast[32-34], cervical[35], ovary[36], and prostate cancer[37], brain tumors[38,39], gastric[40], colorectal[41], pancreatic[42,43], and hepatocellular carcinoma[44,45], B-cell lymphoma[46], melanoma[47,48], and others. In cancer, Pin1 is often overexpressed[49] and localized in both the cytoplasm and the nucleus, which is generally linked with a poor clinical prognosis[41,47,49]. In addition, several activating *PIN1* mutations are associated with a higher risk of developing cancer, while other mutations, leading to reduced cellular Pin1 levels, are correlated with lower cancer risk[49]. Even though these studies suggest that Pin1 is an oncogene, other studies show that Pin1 can also act as a tumor suppressor depending on the cell type and state[49].

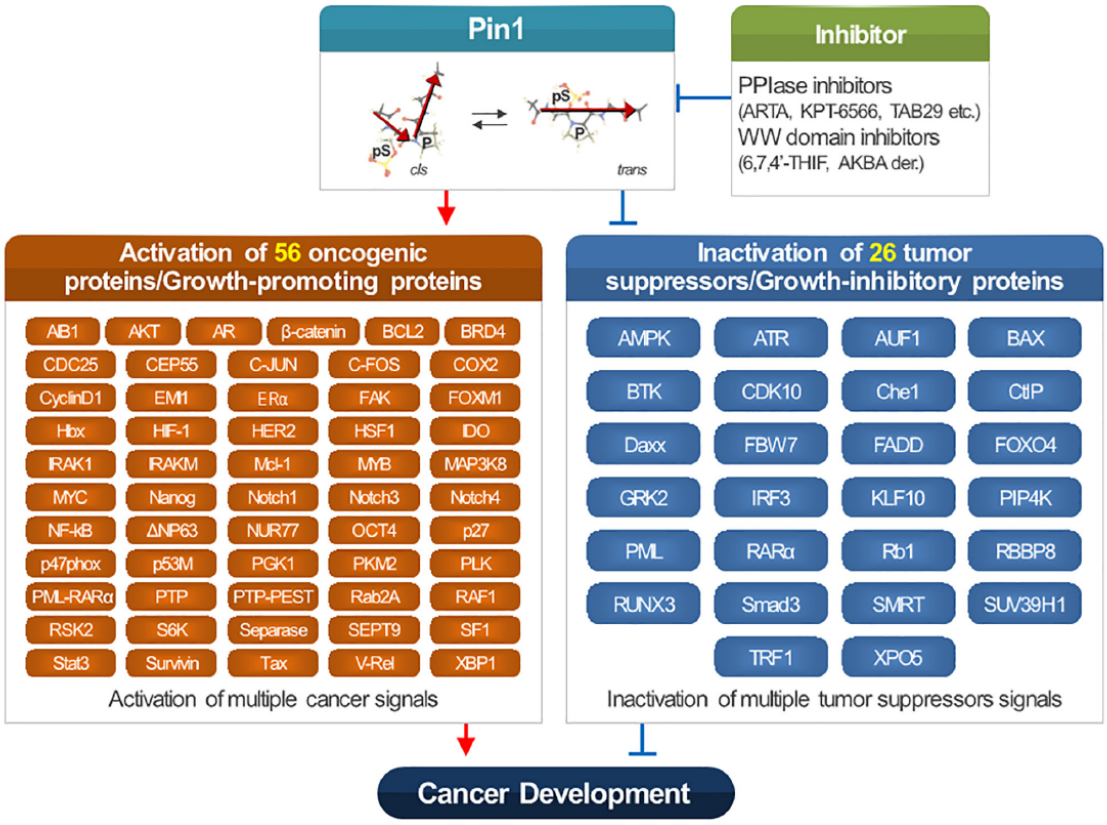


Figure 7.1.6 | Pin1 and Its Role in Cancer. An overview of important Pin1 protein targets involved in either tumor promoting (dark orange) or tumor suppressing (blue) functions is shown. Some examples of reported Pin1 inhibitors are listed in the green box. As Pin1 mainly acts to activate oncogenes and to inactivate tumor suppressors, Pin1 activity facilitates cancer development. Taken from Yu *et al.*, *Front. Cell Dev. Biol.* **2020**, *8*, 120[54].

Taken together, Pin1 orchestrates diverse cellular functions ranging from metabolism to regulation of numerous transcription factors, kinases, phosphatases, E3 ligases, nuclear receptors, and other proteins[50]. Therefore, Pin1 hyperactivation leads to an imbalance in these signaling pathways, which is often associated with the development of cancer and metastasis[51] (Figure 7.1.6). By simultaneously promoting the effects of several oncogenes and silencing tumor suppressor functions, overexpression of Pin1 drives oncogenesis, e.g. by controlling cell cycle events, cellular growth, apoptosis, protein transcription, and various cellular signaling pathways[50]. Transcription of Pin1 is predominantly activated by the E2F and Notch proteins, and its expression and function are further regulated by several miRNAs and post-translational mechanisms, respectively, as described above[52]. Consequently, while Pin1 expression and function are tightly regulated by a complex network of events, cancer-characteristic triggers, including cellular stress signals as well as hormonal and nutrient stimuli, can promote Pin1 expression and hyperactivation, which in turn facilitates differentiation of a cell towards a cancerous state[53].

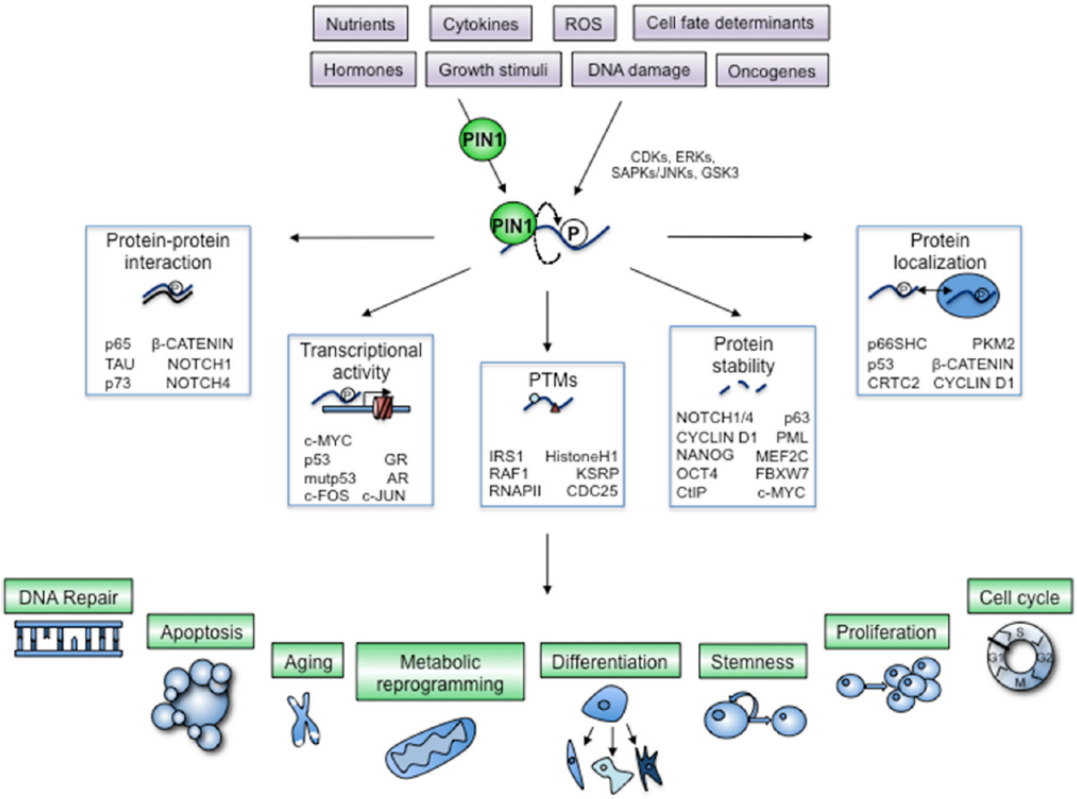


Figure 7.1.7 | Regulation and Function of Pin1. Pin1 regulates a variety of cellular processes *via* its effects on PPIs, protein transcription, post-translational modifications, as well as protein stability and localization of its target proteins. Several hormonal, nutrient, and cell stress signals induce Pin1 activity, which is further directed by Ser/Thr-Pro-directed kinase function. Ultimately, Pin1 was found to be involved in several cellular mechanisms, such as aging, metabolic reprogramming, differentiation, proliferation, and cell cycle function. Taken from Zannini *et al*., *Front. Oncol.* **2019**, *9*, 94[53].

As many of the aforementioned cellular mechanisms and Pin1 targets were summarized in previous chapters on cancer (Chapter 5) and considering the multitude of cellular pathways affected by Pin1 function (Figure 7.1.7), it is clear that Pin1 poses a highly interesting target for the treatment of a variety of diseases, including cancer. In particular, given Pin1's ability to direct "non-druggable" targets such as transcription factors, a Pin1 inhibitor could enable regulation of these targets that are classically difficult to modulate with small molecules[51]. Furthermore, Pin1 knockdown reduces cancer growth in mice, and Pin1 knockout mice have been shown to be resistant to tumorigenesis[51]. These studies, combined with Pin1's function in angiogenesis and metastasis, suggest that Pin1 inhibitors could prevent tumor initiation[54]. Notably, *Pin1*$^{-/-}$ mice develop normally, suggesting that a Pin1 inhibitor might have minimal toxicity, which is in contrast to the majority of anti-tumor treatments[51].

In summary, Pin1 is a master regulator of various cellular mechanisms by orchestrating the function, localization, and stability of its target proteins. Given Pin1's ability to intervene in multiple stages of the same pathway as well as its own regulation *via* translational and post-translational modifications, Pin1's effects could be extremely context- and cell line-dependent, which further adds to the complexity of studying this unique isomerase. However, its hyperactivation in diseases such as cancer, makes it an attractive target for the development of small molecule inhibitors as novel therapeutics.

Small Molecule Pin1 Inhibitors – An Overview

The first small molecule inhibitor of Pin1 was reported by Hennig and coworkers in 1998[55]. The group identified the small, covalent natural product juglone during a screen of inhibitors of bacterial parvulins[55]. However, juglone is highly promiscuous and covalently labels active-site cysteines in many other proteins, which makes it not suitable for use as an informative Pin1 probe[51]. In the following years, substrate-based strategies enabled the design of potentially more selective Pin1 inhibitors, yielding a series of phosphorylated or carboxylated, reversible Pin1 inhibitors, predominantly peptidomimetics[56]. One of the most potent of these early competitive Pin1 inhibitors was pTide, with a biochemical IC_{50} value of 1.2 nM[51,57]. However, due to the charged phosphate or carboxylate functions of these and other potent Pin1 inhibitors, they are not cell-permeable and therefore their potent biochemical activity does not translate into cells[51]. As a result, some octaarginine-decorated cyclic peptide analogs were developed and showed some antiproliferative activity in breast cancer cells[58]. Moreover, PiB – another reversible but non-phosphorylated isomerase inhibitor – was found in a low-throughput screen and was shown to target human parvulins with a biochemical IC_{50} value of ~1.5 μM[59]. However, PiB exhibited equivalent antiproliferative activity in Pin1 wild-type and Pin1$^{-/-}$ cells[59], highlighting its off-target effects.

It was further reported that the natural product epigallo-catechin-3-gallate (EGCG) also targets Pin1[51]. However, besides its weak inhibitory activity with a biochemical IC_{50} value of 22 μM, EGCG is also highly promiscuous[51]. In addition, all-*trans* retinoic acid (ATRA) has been shown to inhibit Pin1 and lead to its degradation; however, its therapeutic efficacy in patients with leukemia points to a multi-targeted mechanism, in particular through its activity against the retinoic acid receptors (RARs)[60]. Similarly, a recent study suggests that arsenic trioxide causes inhibition and degradation of Pin1; however, this compound is also highly promiscuous and exhibits toxicity at high doses[61].

In 2017, Campaner *et al.* identified another covalent Pin1 inhibitor, KPT-6566, *via* a high-throughput screen[62]. While the compound showed growth inhibition in wild-type cells, these effects were rescued in Pin1 knock-out cells[61]. Interestingly, the group proposed a unique mechanism for KPT-6566, in which the compound covalently labels Pin1, releasing a DNA-damaging agent that then leads to cytotoxic effects[62]. More recently, 4,6-bis(benzyloxy)-3-phenylbenzofuran-based[63] as well as dibenzothiophene-based[64] scaffolds have been investigated as potential, reversible Pin1 inhibitors, showing antiproliferative activity in cancer cells; however, the selectivity of these compounds for Pin1 was not demonstrated[63,64].

Figure 7.1.8 shows the structure of the Pin1 inhibitor scaffolds discussed in this chapter. In summary, while there have been many medicinal chemistry campaigns to develop potent Pin1 inhibitors, most of the reported scaffolds are either not potent, show no selectivity for Pin1, or are not cell-permeable. These findings demonstrate the difficulty in designing a suitable Pin1 inhibitor, potentially due to the shallow binding pocket and the highly charged phosphate-binding residues within Pin1's active site. Notably, while none of the reported Pin1 inhibitors exhibits suitable properties to be used as a probe to interrogate Pin1 function in cells or *in vivo*, many have been used as such in the past, likely leading to biased findings regarding Pin1 function and the effect of small molecule-based Pin1 inhibition in health and disease.

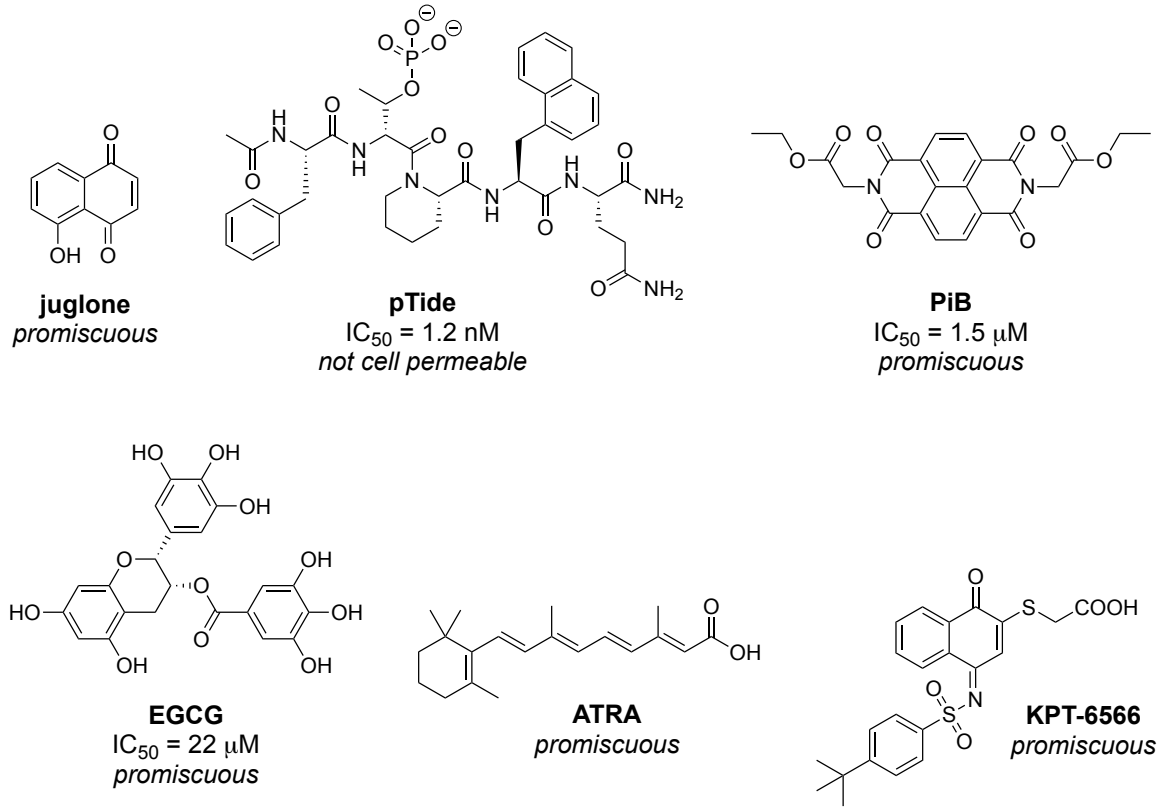


Figure 7.1.8 | Reported Pin1 Inhibitors. Overview of selected Pin1 inhibitor scaffolds mentioned in this chapter.

7.2 Publication E: *Identification of a potent and selective covalent Pin1 inhibitor*

7.2.1 Specific Aims

Starting from the reported peptide-based Pin1 inhibitor pTide, thorough lab-intern SAR studies led to the development of BJP-06-005-3, a potent and cell-permeable Pin1 inhibitor. When joining this project, the main aim was to support the chemistry and cell-biological characterization efforts.

7.2.2 Main Outcomes and Context

The presented study reports the development and characterization of the first potent, highly selective, and cell-permeable Pin1 inhibitor. BJP-06-005-3 was developed from the nM Pin1 modulator pTide, which was the first reported potent, biochemical Pin1 inhibitor. However, like many other Pin1-targeting compounds, pTide is equipped with a phosphate moiety to target the charged phosphate-binding site of the PPIase domain of Pin1 and it therefore is cell-impermeable. After extensive SAR studies and iterative optimization, BJP-06-005-3 was obtained not only as an improved non-phosphorylated Pin1 binder but also as a covalent modulator of the isomerase, specifically targeting C113.

Thorough characterization of the probe *via* mass spectrometric analysis and co-crystallization confirmed the covalent binding mode of the inhibitor. A desthiobiotin-labeled analog of BJP-06-005-3 (BJP-DTB) was further used to confirm the selectivity of the covalent probe by using the novel CITe-Id technology, showing that Pin1's C113 is the only dose-dependently labeled cysteine in HEK293 cell lysate samples. BJP-DTB was also highly useful in tracking cellular target engagement of BJP-06-005-3 using streptavidin bead-based pulldown assays.

Lastly, the consequences of Pin1 loss in pancreatic ductal adenocarcinoma (PDAC) were investigated. Genetic studies were performed to interrogate the phenotype of Pin1 overexpression, knockout, and mutation. NIH/3T3 cells were transduced with doxycycline (dox)-inducible Pin1; dox treatment successfully induced Pin1 overexpression and promoted cell transformation in the context of a $KRAS^{G12V}$ oncoprotein. It was further noted that $Pin1^{-/-}$ PATU 8988T cells showed a significant reduction in their growth rate as compared to Pin1 wild-type cells. Pin1 mutants with impaired catalytic activity, including W34A (WW domain), K63A (PPIase domain), and C113S (PPIase domain), all led to the same reduced cell growth phenotype as the knockout cell line. Similarly, long-term BJP-06-005-3 treatment in PATU 8988T cells led to time-dependent growth defects in a Pin1-dependent manner and was synergistic with $KRAS^{G12V}$ loss. BJP-06-005-3's mode of action was further analyzed and was linked to Pin1's effects on cell cycle progression.

In summary, the study presents the development and characterization of a high-quality chemical probe to interrogate Pin1 function *in cellulo* and provides evidence of the therapeutic potential of Pin1 inhibition in PDAC. The presented work underlies the importance and necessity of high-quality, well-characterized probes for the study of biological targets, and demonstrates another important application of desthiobiotin-labeled probes, here, using CITe-Id, which was critical in characterizing the proteome-wide selectivity of BJP-06-005-3.

Methods for cellular assays are found at the end of the article. Information regarding the chemical synthesis is provided in the Supplementary Information.

7.2.3 Published Article

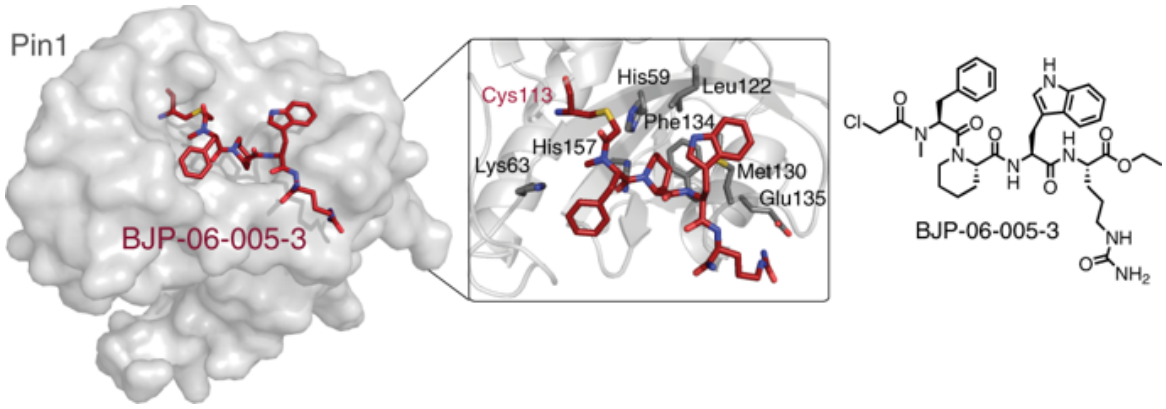
The following peer-reviewed publication was originally published as:

Identification of a potent and selective covalent Pin1 inhibitor

Benika J. Pinch, Zainab M. Doctor, Behnam Nabet, Christopher M. Browne, Hyuk-Soo Seo, Mikaela L. Mohardt, Shingo Kozono, Xiaolan Lian, Theresa D. Manz, Yujin Chun, Shin Kibe, Daniel Zaidman, Dina Daitchman, Zoe C. Yeoh, Nicholas E. Vangos, Ezekiel A. Geffken, Li Tan, Scott B. Ficarro, Nir London, Jarrod A. Marto, Stephen Buratowksi, Sirano Dhe-Paganon, Xiao Zhen Zhou, Kun Ping Lu and Nathanael S. Gray.

Nat. Chem. Biol. **2020**, *16* (9), 979-987.

DOI: https://doi.org/10.1038/s41589-020-0550-9



This article, including Supporting Information, was reproduced with kind permission from the journal *Nature Chemical Biology* © 2020, Springer Nature.

nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-020-0550-9

Check for updates

Identification of a potent and selective covalent Pin1 inhibitor

Benika J. Pinch[1,2,3], Zainab M. Doctor[2,3,10], Behnam Nabet[2,3,10], Christopher M. Browne[2,3,9], Hyuk-Soo Seo[2,3], Mikaela L. Mohardt[2], Shingo Kozono[4], Xiaolan Lian[4], Theresa D. Manz[2,3,5], Yujin Chun[3], Shin Kibe[4], Daniel Zaidman[6], Dina Daitchman[6], Zoe C. Yeoh[2,3], Nicholas E. Vangos[2,3], Ezekiel A. Geffken[2,3], Li Tan[2], Scott B. Ficarro[2,7], Nir London[6], Jarrod A. Marto[7,8], Stephen Buratowski[3], Sirano Dhe-Paganon[2,3], Xiao Zhen Zhou[4], Kun Ping Lu[4] ✉ and Nathanael S. Gray[2,3] ✉

Peptidyl-prolyl *cis/trans* isomerase NIMA-interacting 1 (Pin1) is commonly overexpressed in human cancers, including pancreatic ductal adenocarcinoma (PDAC). While Pin1 is dispensable for viability in mice, it is required for activated Ras to induce tumorigenesis, suggesting a role for Pin1 inhibitors in Ras-driven tumors, such as PDAC. We report the development of rationally designed peptide inhibitors that covalently target Cys113, a highly conserved cysteine located in the Pin1 active site. The inhibitors were iteratively optimized for potency, selectivity and cell permeability to give BJP-06-005-3, a versatile tool compound with which to probe Pin1 biology and interrogate its role in cancer. In parallel to inhibitor development, we employed genetic and chemical-genetic strategies to assess the consequences of Pin1 loss in human PDAC cell lines. We demonstrate that Pin1 cooperates with mutant KRAS to promote transformation in PDAC, and that Pin1 inhibition impairs cell viability over time in PDAC cell lines.

Proline-directed phosphorylation controls numerous cellular processes, including cell cycle progression, transcription and differentiation. Proline (Pro) is unique among the amino acids in that its imidic peptide bond can populate either the *cis* or the *trans* conformation[1,2], and the intrinsic isomerization rate of Pro-containing motifs is slow relative to biological signaling, requiring catalysis by peptidyl-prolyl isomerases (PPIases). As the only known phosphorylation-dependent PPIase, Pin1 acts in tandem with Pro-directed kinases and phosphatases, which are conformation-specific, to control the stability, localization and activity of their common targets[3].

Pin1 is frequently overexpressed in cancer[4], including PDAC, and Pin1 overexpression correlates with poor prognosis[5]. Pin1 ablation is reported to prevent MMTV-Ras-driven mouse mammary gland carcinoma, indicating that Pin1 is a key effector in Ras signaling[6]. While mutations in *KRAS* are observed in 90–95% of PDAC cases[7], it has historically proven difficult to develop inhibitors of mutant KRAS, spurring efforts to target proteins that facilitate Ras-mediated transformation, such as Pin1. Notably, while the yeast homolog of Pin1 (Ess1) is essential[8], Pin1-null mice develop normally, though with decreased body weight[9], suggesting that Pin1 inhibition could reduce tumorigenic potential with limited toxicity.

Several Pin1 inhibitors have been described to date, including juglone[10], all-*trans* retinoic acid (ATRA)[11], arsenic trioxide (ATO)[12], KPT-6566 (ref. [13]) and (*S*)-2 (ref. [14]). However, these compounds lack the potency and/or selectivity to accurately assess Pin1-mediated phenotypes[15]. Notably, while reported Pin1 inhibitors demonstrate potent cellular cytotoxicity, Pin1 was not identified as a cancer dependency in genome-wide CRISPR screens conducted across hundreds of cancer cell lines (Cancer Dependency Map), thus highlighting a contradiction in our understanding of Pin1 biology. A selective, potent and cell-permeable Pin1 inhibitor is required to effectively evaluate the role of Pin1 in oncogenesis.

Targeted covalent inhibitors allow for improved potency and selectivity, and prolonged duration of action[16,17]. Oxidation of the conserved cysteine in the Pin1 active site, Cys113, eliminates its isomerase activity and impairs its cellular function[18], suggesting that covalent binders would irreversibly inhibit Pin1. Also, previously reported inhibitors juglone[10], KPT-6566 (ref. [13]) and (*S*)-2 (ref. [14]) provide precedent for targeting Pin1 covalently.

In this study, we describe the structure-guided design of covalent peptides targeting Pin1 Cys113 by modifying the high-affinity but cell-impermeable substrate-mimetic inhibitor, D-PEPTIDE[19]. We demonstrate that the resulting inhibitor, BJP-06-005-3 (**1**), is potent, cell-permeable and highly selective. We further show that Pin1 cooperates with $KRAS^{G12V}$ to promote cellular transformation. By leveraging covalent Pin1 inhibition, as well as targeted Pin1 degradation using the degradation tag (dTAG) strategy[20], we demonstrate that Pin1 loss impairs cell growth and transformation in PDAC model systems. Ultimately, this work showcases a set of

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [2]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. [4]Division of Translational Therapeutics, Department of Medicine and Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [5]Department of Pharmacological and Medicinal Chemistry, Saarland University, Saarbruecken, Germany. [6]Department of Organic Chemistry, Weizmann Institute of Science, Rehovot, Israel. [7]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA, USA. [8]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [9]Present address: Discovery Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Boston, MA, USA. [10]These authors contributed equally: Zainab M. Doctor, Behnam Nabet.
✉e-mail: klu@bidmc.harvard.edu; Nathanael_Gray@dfci.harvard.edu

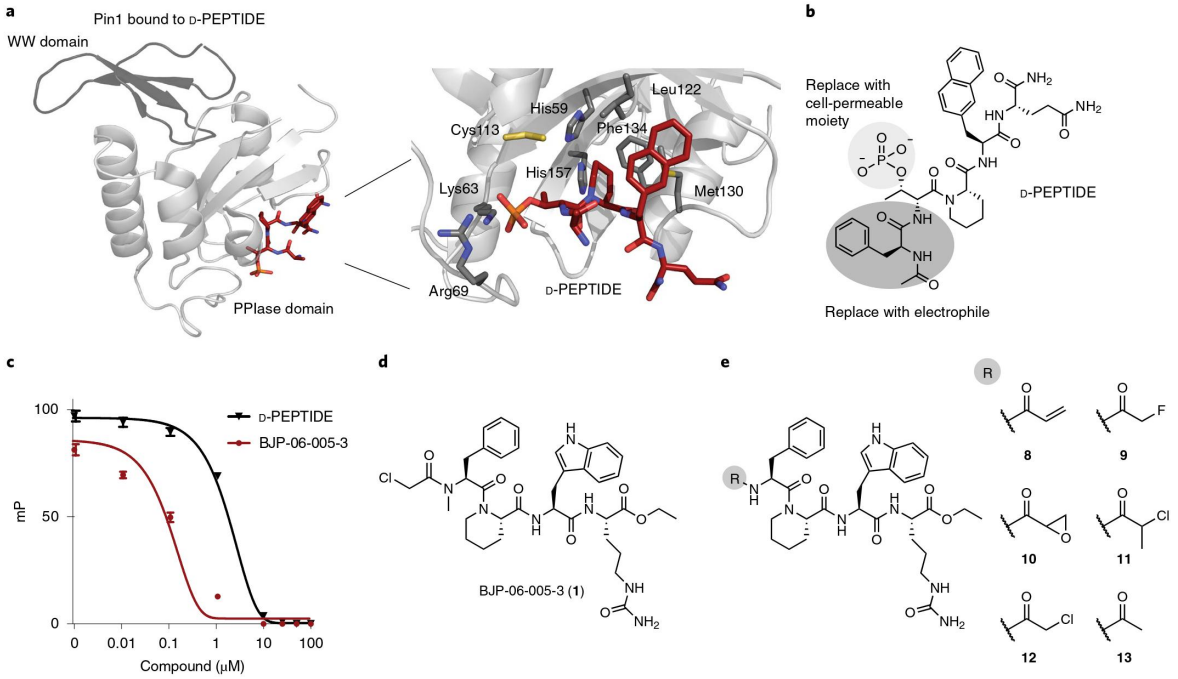


Fig. 1 | Structure-guided modification of D-PEPTIDE to yield optimized Pin1 binders. a, Structure of D-PEPTIDE (red) bound to Pin1, with the PPIase domain in light gray and the WW domain in dark gray (left). Zoomed structure of D-PEPTIDE (red) in the PPIase active site, with key residues for binding highlighted in dark gray, and Cys113 in yellow (right). PDB 2ITK. **b**, Plan for structure-guided optimization of D-PEPTIDE to achieve cell permeability and covalency. **c**, FP in units of millipolarization (mP) following a 12-h incubation with Pin1. Data points are plotted as the average of $n=3$ independent samples ±S.E.M., and are representative of $n=3$ independent experiments. **d**, Chemical structure of BJP-06-005-3. **e**, Chemical structures of derivatives featuring alternate electrophiles.

robust chemical, genetic and chemical-genetic tools with which to study the role of Pin1 in oncogenesis.

Results

Structure-guided design of Pin1 inhibitors. To develop a Pin1-targeted probe, we modified D-PEPTIDE (Ac-Phe-D-phos.Thr-Pip-Nal-Gln-NH_2; $K_i=20$ nM), which is a potent Pin1 inhibitor that cannot enter cells[19]. We sought to replace the phosphate of D-PEPTIDE with a cell-permeable moiety, while compensating for any resulting loss in potency with a covalent bond to Cys113. Based on the Pin1/D-PEPTIDE cocrystal structure (PDB 2ITK)[19], the N-terminus of D-PEPTIDE provided the best access to Cys113 and therefore served as the attachment site for a cysteine-targeting electrophile (Fig. 1a,b). We chose an α-chloroacetamide as the electrophile based on a previous proteome-wide screen to quantify cysteine reactivity, which identified several α-chloroacetamide-containing fragments that labeled Pin1 (ref. [21]). An α-chloroacetamide is also featured in KPT-6566 (ref. [13]).

We profiled each synthesized inhibitor in a fluorescence polarization (FP) assay[11] following a 12-h incubation of the inhibitor with recombinant Pin1 to allow sufficient time for covalent labeling. Given precedent that a carboxylate can successfully replace a phosphate in Pin1 inhibitors[22], we initially substituted the phos.Thr of D-PEPTIDE with a carboxylic acid. While an Asp-containing derivative (**2**) did not bind Pin1, the extra methylene of a Glu enabled Pin1 binding with an apparent K_i of 2 μM (**3**). Replacing the Gln with Arg improved potency by 12-fold to give BJP-02-118-2 (**4**). We obtained a cocrystal structure of BJP-02-118-2 bound to Pin1, which revealed a charge–charge interaction with Pin1 Glu135 (PDB 6O33; Supplementary Table 1). We then further enhanced potency by substituting the Arg with citrulline (Cit) (**5**) and the naphthylalanine (Nal) with Trp, improving the apparent K_i to 30 nM (**6**) (Supplementary Fig. 1a–c).

Despite having successfully replaced the phosphate with a carboxylate, the resulting inhibitors had negative calculated log (*n*-octanol/water partition coefficient) (cLogP) values and therefore low lipophilicity, which governs passive membrane permeability[23] (Supplementary Fig. 1d). We therefore increased lipophilicity by capping the C terminus with an ethyl ester, methylating the N-terminal nitrogen to eliminate a hydrogen bond donor and replacing the Glu with Phe to give BJP-06-005-3, with an apparent K_i of 15 nM (Fig. 1c,d). We hypothesized that the surprising affinity of BJP-06-005-3 might be explained by a cation–pi interaction between the Phe benzene and Arg68 or Arg69 in the Pin1 phosphate-binding pocket. Notably, while D-PEPTIDE features a D-phos.Thr residue, the covalent binding of BJP-06-005-3 flips the preferred stereochemistry at that position such that a D-Phe substitution (**7**) eliminates Pin1 binding (Supplementary Fig. 2a,b).

We next explored whether alternate thiol-targeting electrophiles were tolerated by substituting the α-chloroacetamide with an acrylamide (**8**), α-fluoroacetamide (**9**), epoxide (**10**) or 2-chloropropionamide (**11**). For these studies, we used a derivative of BJP-06-005-3 lacking the methylated N-terminal nitrogen (**12**, apparent $K_i=7.4$ nM). As a control, we also synthesized the unreactive *N*-methylacetamide (**13**) (Fig. 1e). None of the compounds featuring alternate electrophiles bound Pin1, highlighting that the α-chloroacetamide is necessary to achieve covalency (Supplementary Fig. 2c). Similarly, an analog

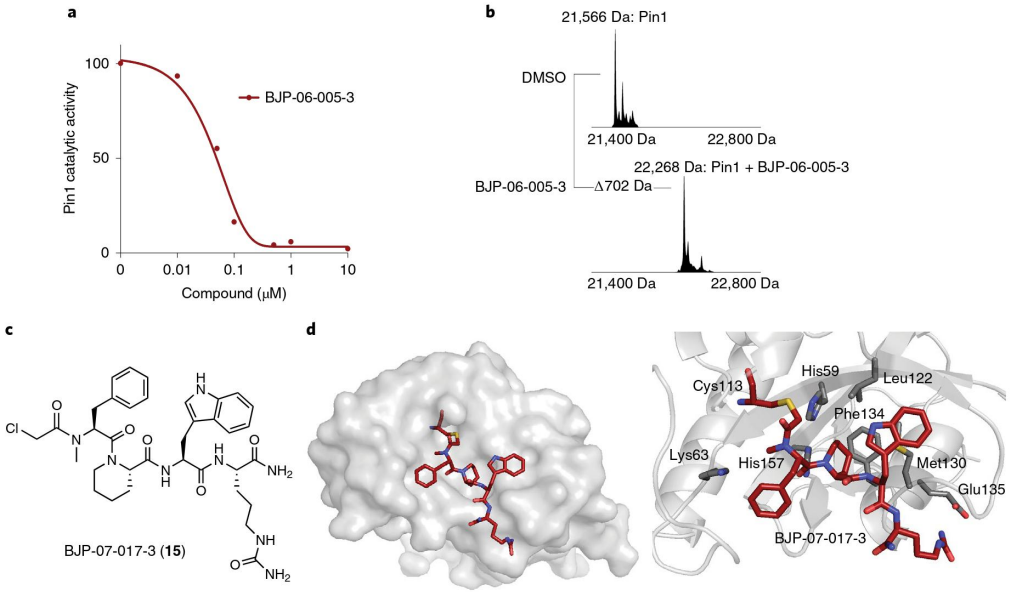


Fig. 2 | Biochemical and structural characterization of BJP-06-005-3. a, PPIase assay results for BJP-06-005-3 (IC_{50} = 48 nM) after a 12-h incubation with Pin1. Data points are plotted as the mean of n = 2 independent experiments, with each experiment having n = 1 independent samples. **b**, Intact mass spectrometry of Pin1 after incubation with DMSO or BJP-06-005-3 for 1 h at room temperature. **c**, Chemical structure of BJP-07-017-3. **d**, 1.6-Å cocrystal structure of BJP-07-017-3 (red) covalently bound to Pin1 (light gray) via Cys113, with key binding residues highlighted in dark gray (PDB 6O34).

of BJP-06-005-3 in which the α-chloroacetamide was replaced with an *N*-methylacetamide (BJP-R (**14**); Supplementary Fig. 2a,b) lost all binding to Pin1 indicating that, despite optimizing the noncovalent interactions, the binding of BJP-06-005-3 is primarily driven by its covalency. BJP-R was subsequently used as a negative control compound. In summary, by making a series of structure-guided modifications to D-PEPTIDE and leveraging a covalent binding mode, we developed BJP-06-005-3 as a high-affinity Pin1 binder.

BJP-06-005-3 is a potent and covalent Pin1 inhibitor. We next evaluated the inhibitory activity, covalency and binding mode of BJP-06-005-3 (Supplementary Fig. 3a). We assessed the ability of BJP-06-005-3 to inhibit Pin1 enzymatic activity using a chymotrypsin-coupled PPIase assay. In this spectrophotometric assay, the peptidic substrate (Succ-Ala-p.Ser-Pro-Phe-pNA) is only cleaved by chymotrypsin when the peptidyl-prolyl bond is in the *trans* conformation[11]. BJP-06-005-3 potently inhibited Pin1 catalytic activity with an apparent K_i of 48 nM (Fig. 2a). Next, to assess the covalency of BJP-06-005-3, recombinant full-length Pin1 protein was analyzed by intact mass spectrometry following incubation with BJP-06-005-3 or DMSO. BJP-06-005-3 showed 100% covalent labeling of Pin1 as indicated by a 702-Da molecular weight increase, corresponding to modification of Pin1 by BJP-06-005-3 upon loss of its chloride (Fig. 2b). Trypsin digest confirmed the site of covalent modification as Cys113 (Supplementary Fig. 3b). We next performed the FP assay in a dose- and time-dependent manner (0–30 min) to assess the k_{inact}/K_i of BJP-06-005-3, in which K_i describes the reversible binding and k_{inact} the maximum rate of inactivation[24]. We determined the k_{inact} to be 0.08 ± 0.01 min^{-1} and the K_i to be 1,800 ± 670 nM (k_{inact}/K_i = 740 M^{-1} s^{-1}) (Supplementary Fig. 3c). The apparent K_i of BJP-06-005-3 decreases with increasing incubation time, reaching 15 nM following a 12-h incubation with Pin1.

To evaluate the binding mode of BJP-06-005-3, we sought to obtain a cocrystal structure with Pin1. Unfortunately, BJP-06-005-3 precipitated out of the crystallization buffer, prompting us to synthesize a more hydrophilic and water soluble derivative of BJP-06-005-3 by replacing its C-terminal ethyl ester with an amide to give BJP-07-017-3 (**15**), with an apparent K_i of 15 nM (Fig. 2c and Supplementary Fig. 3d). This enabled us to obtain a 1.6-Å cocrystal structure of BJP-07-017-3 bound to full-length Pin1 (PDB 6O34), which showed good ligand occupancy and clear electron density to Cys113, consistent with covalent modification (Supplementary Fig. 3e and Supplementary Table 1). In this structure, the pipecolinic acid (Pip) of BJP-07-017-3 acts as a proline mimetic, making hydrophobic contacts with Leu122, Met130, Phe134, His59 and His157 in the prolyl-binding pocket. The Cit of BJP-07-017-3 is engaged in hydrogen bonds to the backbone amide of Met130 and Glu135, while the Trp interacts with the sidechain of Met130 in a methionine-aromatic motif[25] (Fig. 2d). Given that both Arg68 and Arg69 were disordered, we were unable to conclude whether the Phe of BJP-07-017-3 is engaged in a cation–pi interaction, as per our original hypothesis. Collectively, these biochemical and structural data confirmed that BJP-06-005-3 is a potent and covalent Pin1 inhibitor.

BJP-06-005-3 is selective and engages cellular Pin1. To evaluate the selectivity of BJP-06-005-3 and its ability to covalently modify Pin1 in cells, we synthesized a desthiobiotin (DTB)-labeled derivative (BJP-DTB, **16**), which retained high affinity for Pin1 (Supplementary Fig. 4a,b). BJP-DTB exhibited potent and rapid pull-down of Pin1 from PATU-8988T cell lysates (Supplementary Fig. 4c,d), and pull-down was selectively outcompeted following pre-incubation with BJP-06-005-3 (Supplementary Fig. 4e). In addition, BJP-DTB pulled down wild-type (WT) Pin1 but not a C113S Pin1 mutant, validating its covalent binding mode (Supplementary Fig. 4f).

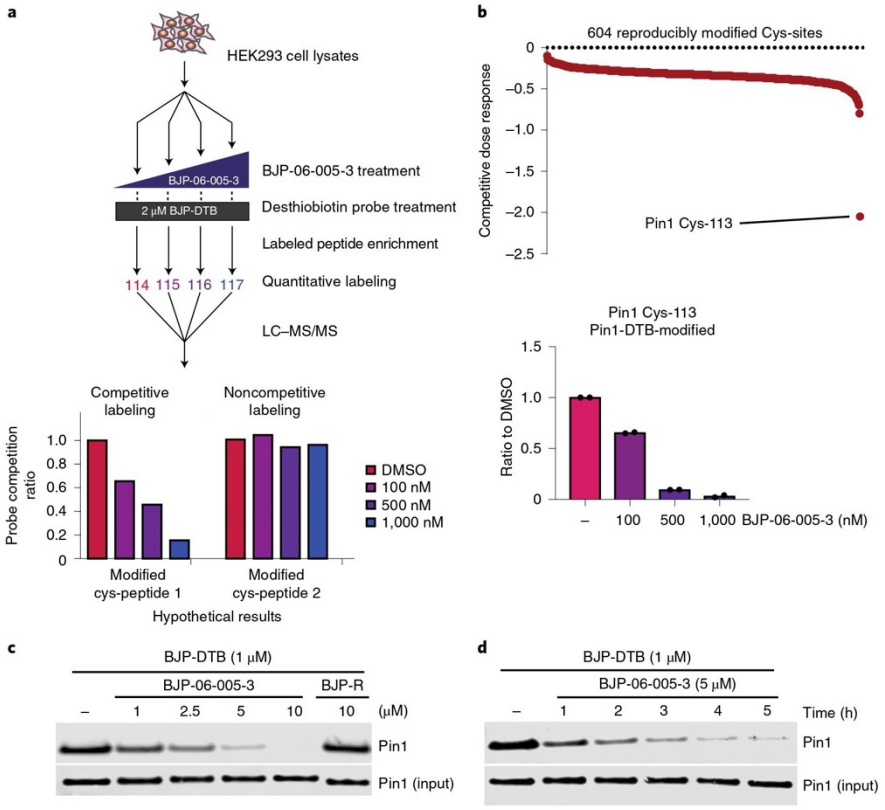


Fig. 3 | BJP-06-005-3 is selective and engages cellular Pin1. a, Schematic depiction of CITe-Id workflow, showing hypothetical results. CITe-Id directly identifies BJP-DTB modification sites across the proteome and quantifies their affinity for BJP-06-005-3 (ref. [26]). **b**, Waterfall plot of the competitive dose responses for the 604 cysteine sites reproducibly modified by BJP-DTB in $n=2$ independent experiments. Pin1 Cys113 competitive dose response is 11 s.d. from the next-strongest competing site, and it is the only site with >50% competition at the two highest concentrations (90% competition at 500 nM BJP-06-005-3, 97% competition at 1 μM BJP-06-005-3). The mean peak heights of $n=2$ biologically independent samples are shown as dot-plots. **c**, PATU-8988T cells treated in a competition format with the indicated concentrations of BJP-06-005-3 or BJP-R for 5 h, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down and immunoblot analysis. **d**, PATU-8988T cells treated with 5 μM BJP-06-005-3 for the indicated times, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down and immunoblot analysis. Uncropped immunoblots for **c** and **d** are displayed in Supplementary Fig. 12, and are representative of $n=4$ independent experiments.

We employed BJP-DTB to assess the selectivity of BJP-06-005-3 using Covalent Inhibitor Target Site Identification (CITe-Id), which quantifies the proteome-wide binding of covalent inhibitors to individual cysteines[26]. In this competition assay, HEK293 lysates were pre-incubated with BJP-06-005-3, followed by co-incubation with BJP-DTB (Fig. 3a). Across the proteome, of 604 sites reproducibly labeled by BJP-DTB in two independent experiments, Pin1's Cys113 was the only cysteine that exhibited dose-dependent binding to BJP-06-005-3 (Fig. 3b and Supplementary Dataset 1), highlighting the remarkable selectivity of this inhibitor.

To assess the permeability and cellular target engagement of BJP-06-005-3, we performed a live cell competition assay (Supplementary Fig. 5a). Following a 5-h treatment in PATU-8988T cells, BJP-06-005-3 exhibited dose-dependent competition with BJP-DTB for binding to Pin1, with maximal engagement observed at 10 μM, while the negative control, BJP-R, did not show competition (Fig. 3c). Engagement of Pin1 by BJP-06-005-3 was evident within 4 h (Fig. 3d) and was maintained for 48 h following a single treatment. Pin1 engagement was lost by 72 h, likely due to hydrolysis of the peptide (Supplementary Fig. 5b–d). BJP-06-005-3 similarly outcompeted BJP-DTB in PC3 (prostate cancer) and MDA-MB-468 (breast cancer) cells, with slight loss of Pin1 engagement observed 48 h after treatment (Supplementary Fig. 5e). These data suggest that BJP-06-005-3 engages Pin1 for 48 h following a single 5–10 μM treatment across multiple cancer cell types. However, while BJP-06-005-3 exhibits strong cellular target engagement, it has poor mouse liver microsome stability ($T_{1/2}=1.2$ min), limiting its applications to biochemical and cell-based assays.

Covalent inhibitors can induce degradation of their target proteins, often by causing structural rearrangement[27]. Pin1 is frequently degraded following inhibitor binding, as previously demonstrated with ATRA[11], ATO[12] and KPT-6566 (ref. [13]). Similarly, BJP-06-005-3 caused partial Pin1 degradation following a 48-h treatment in two $KRAS^{G12V}$ PDAC cell lines, PATU-8988T and PATU-8902, and BJP-06-005-3 reduced Pin1 half-life in PATU-8988T cells (Supplementary Fig. 6a–d). However, this phenotype was cell line-dependent as BJP-06-005-3 did not degrade Pin1 across a panel of additional cancer cell lines. In all cell lines tested, however, binding of BJP-06-005-3 to Pin1 caused an upward shift in the Pin1 protein band by immunoblot, corresponding to the molecular weight increase when

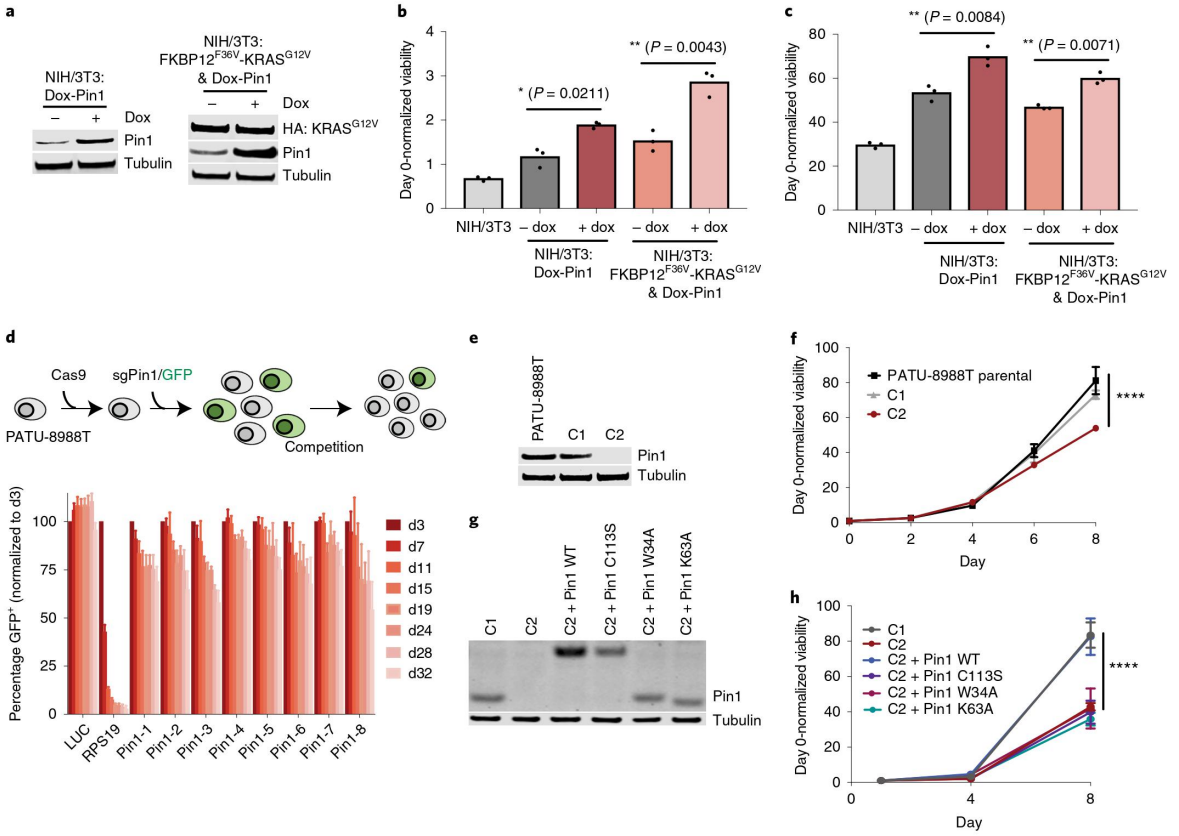


Fig. 4 | Pin1 cooperates with KRAS^{G12V} to promote cell transformation, and Pin1 loss leads to viability defects. a, Immunoblot analysis of NIH/3T3 cells transduced with Dox-inducible Pin1 alone, or in combination with FKBP12^{F36V}-KRAS^{G12V} (HA-tagged), after Dox treatment (100 ng ml^{-1}) for 48 h. **b**,**c**, Day 0-normalized proliferation of the cell lines from **a**, cultured as 3D spheroid suspensions (**b**) or as 2D monolayers (**c**), with or without Dox (100 ng ml^{-1}), for 120 h. The mean peak heights of $n = 3$ biologically independent samples are shown as dot-plots, and are representative of $n = 2$ independent experiments. *P* values were determined by Welch's two-sided *t*-test. **d**, Schematic of the CRISPR–Cas9 GFP dropout assay (above); GFP$^+$ (sgRNA$^+$) population percentage is depicted at the indicated days following lentiviral transduction in Cas9-expressing PATU-8988T cells (below). Statistics are reported in Supplementary Table 4. **e**, Immunoblot analysis of lysates from parental PATU-8988T, C1 and C2. **f**, Day 0-normalized growth rate of C1, C2 and parental PATU-8988T, with readings taken every 2 d. ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. **g**, Immunoblot analysis of lysates from C1, C2 and C2 transduced with Flag-Pin1 (WT or C113S) or Pin1 (W34A or K63A). **h**, Day 0-normalized growth rate of C1, C2 or C2 transduced with Flag-Pin1 (WT or C113S) or Pin1 (W34A or K63A), with readings taken every 4 d. ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. For **f** and **h**, data points are plotted as the average of $n = 3$ biologically independent samples ±s.e.m., and are representative of $n = 2$ independent experiments. Uncropped immunoblots for **a**, **e** and **g** are displayed in Supplementary Fig. 12, and are representative of $n = 2$ independent experiments. d, day.

Pin1 (~18 kDa) is labeled by BJP-06-005-3 (molecular weight 702) (Supplementary Fig. 6d). Taken together, these findings affirmed that BJP-06-005-3 is selective and engages cellular Pin1.

Pin1 cooperates with KRAS^{G12V} to promote transformation. Previous evidence suggests that Pin1 is required for Ras-induced transformation[6], and Pin1 overexpression in Ras-transformed human mammary epithelial cells is reported to increase tumor incidence in mice[28]. We sought to evaluate whether Pin1 cooperates with mutant KRAS to coordinate cell transformation, and to investigate the role of Pin1 in PDAC, a mutant KRAS-driven cancer. To assess the effects of Pin1 overexpression on transformation, we transduced NIH/3T3 cells with doxycycline (Dox)-inducible Pin1 alone, or in combination with FKBP12^{F36V}-KRAS^{G12V}, which is a functional oncoprotein[20]. After confirming that Dox treatment for 48 h effectively increased Pin1 levels (Fig. 4a), we assessed the ability of both cell lines to form three-dimensional (3D) spheroid suspensions in ultra-low adherent (ULA) conditions to assay cellular transformation[29,30]. Compared with parental NIH/3T3 cells, Pin1 overexpression led to a modest increase in spheroid formation, which was more significant in the context of FKBP12^{F36V}-KRAS^{G12V} expression (Fig. 4b), though this trend was less pronounced in two-dimensional (2D) adherent monolayers (Fig. 4c). These results demonstrate that Pin1 overexpression is mildly transformative by itself, and cooperates with activated KRAS to increase transformation.

Using genetic strategies, we assessed the effects of Pin1 loss on viability in a PDAC context, laying the groundwork for subsequent studies with BJP-06-005-3. We performed a competition-based CRISPR–Cas9 GFP dropout assay[31,32] in Cas9-expressing PATU-8988T cells (Fig. 4d), using eight GFP-expressing plasmids encoding single guide RNAs (sgRNAs) targeting *PIN1*, which were confirmed to induce Pin1 knockout (KO; Supplementary Fig. 6e

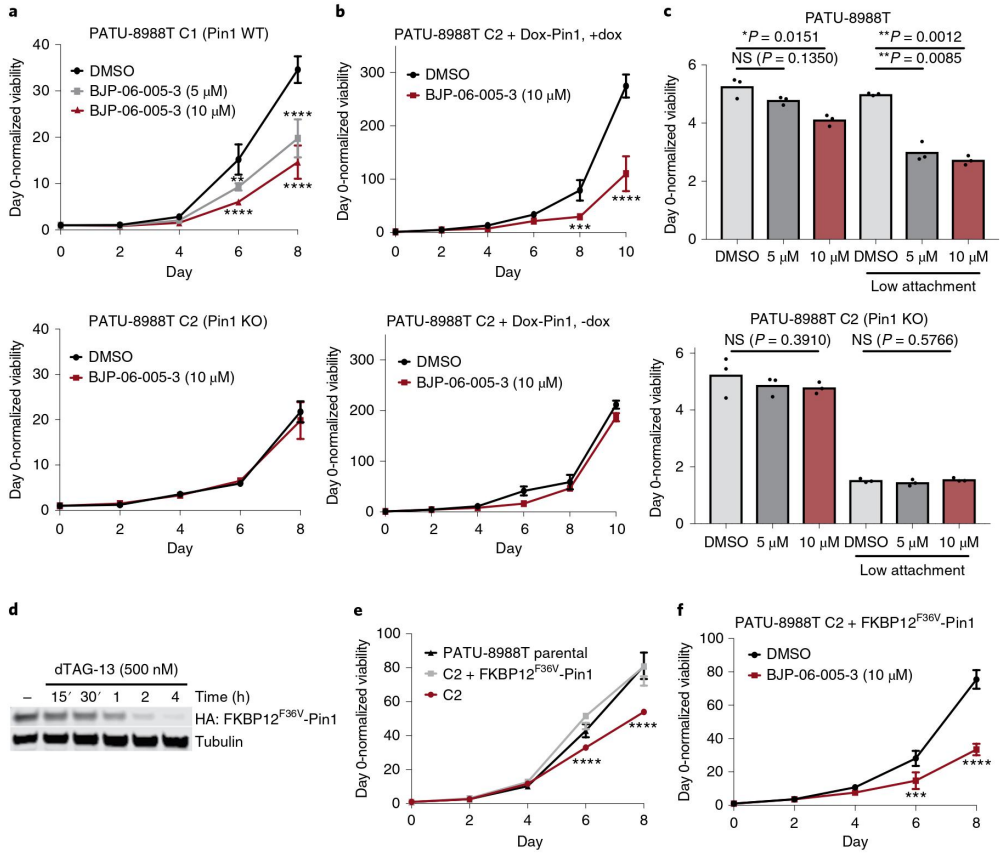


Fig. 5 | BJP-06-005-3 impacts PDAC cell viability in a Pin1-mediated manner. a, Day 0-normalized growth rate of C1 (above) or C2 (below), treated with DMSO or BJP-06-005-3 (5 or 10 μM), with readings taken every 2 d. **adj. $P = 0.0063$, ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. **b**, Day 0-normalized growth rate of C2 + Dox-inducible Pin1, treated with DMSO or BJP-06-005-3 (10 μM), in the presence (above) or absence (below) of Dox (100 ng ml^{-1}), with readings taken every 2 d. ***adj. $P = 0.0005$, ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. **c**, PATU-8988T cells (parental, above; C2, below) were cultured in 2D monolayers or as 3D spheroids (indicated by 'low attachment') and treated with DMSO or BJP-06-005-3 (5 or 10 μM) for 4 d, with compound re-treatment on day 2. The mean peak heights of $n = 3$ biologically independent samples are shown as dot-plots, and are representative of $n = 2$ independent experiments. *P* values were derived from Welch's two-sided *t*-test, with comparisons made to DMSO. NS, not significant. **d**, Immunoblot analysis of C2 + FKBP12^{F36V}-Pin1 cells treated with dTAG-13 (500 nM) for the indicated time points. Uncropped immunoblots are displayed in Supplementary Fig. 12, and are representative of $n = 3$ independent experiments. **e**, Day 0-normalized growth rate of PATU-8988T parental, C2 and C2 + FKBP12^{F36V}-Pin1, with readings taken every 2 d. ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. **f**, Day 0-normalized growth rate of C2 + FKBP12^{F36V}-Pin1 cells treated with BJP-06-005-3 (10 μM), with readings taken every 2 d. ***adj. $P = 0.0001$, ****adj. $P < 0.0001$ by Bonferroni-corrected two-way ANOVA. For **a**, **b**, **e** and **f**, data points are plotted as the average of $n = 3$ biologically independent samples ±s.e.m., and are representative of $n = 3$ independent experiments.

and Supplementary Tables 2 and 3). By monitoring the GFP signal over time, we found that Pin1 KO led to mild but statistically significant, time-dependent defects in PDAC cell growth (Fig. 4d and Supplementary Table 4).

To confirm this observation, we isolated single-cell Pin1$^{-/-}$ clones from PATU-8988T cells and compared the basal growth rate of a clone lacking Pin1 (referred to as C2) with a clone that maintained expression of endogenous Pin1 (referred to as C1; Fig. 4e). C2 exhibited a significantly slower growth rate, as compared with parental PATU-8988T cells or C1 (Fig. 4f). To evaluate whether Pin1's substrate-binding WW domain and/or its catalytic PPIase domain were required for this viability effect, we transduced C2 with WT Pin1, a Pin1 mutant lacking WW domain activity (W34A) or Pin1 mutants with disrupted PPIase activity (K63A; C113S)[33] (Fig. 4g). While expression of WT Pin1 increased the basal growth rate of C2, expression of any of the Pin1 mutants did not alter the basal growth rate (Fig. 4h), suggesting that Pin1 must have functional WW and PPIase domains to mediate enhanced cell viability in PATU-8988T cells. Collectively, these data indicate that Pin1 overexpression cooperates with KRAS^{G12V} to enhance cell transformation, and that Pin1 KO causes viability defects in PATU-8988T cells. This supports the further evaluation of BJP-06-005-3 in PDAC cellular contexts.

BJP-06-005-3 has Pin1-dependent cell viability effects. We next assessed the effects of BJP-06-005-3 on cell viability. While BJP-06-005-3 treatment was largely nontoxic following a 4-d treatment in PATU-8988T cells, increasing the duration of the experiment to 8 d led to diminished cell viability, with replenishment of BJP-06-005-3 every 48 h to ensure durable target engagement. The negative

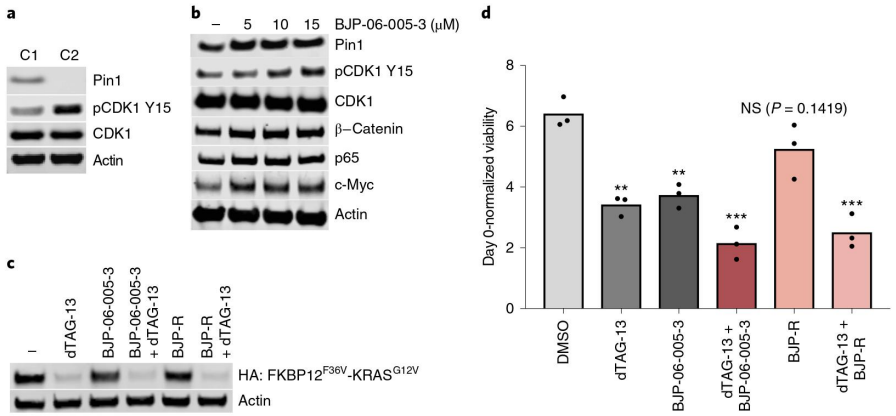


Fig. 6 | BJP-06-005-3 affects expression of downstream Pin1 substrates, and cooperates with KRAS^{G12V} degradation. a, Immunoblot analysis of lysates from C1 and C2. **b**, Immunoblot analysis following BJP-06-005-3 treatment at the indicated concentrations for 5 h in C1. **c**, Immunoblot analysis of PATU-8988T FKBP12^{F36V}-KRAS^{G12V}; *KRAS*$^{-/-}$ clone treated with dTAG-13 (100 nM), BJP-06-005-3 (5 μM), BJP-R (5 μM) or the indicated combinations for 4 h. **d**, Day 0-normalized viability of PATU-8988T FKBP12^{F36V}-KRAS^{G12V}; *KRAS*$^{-/-}$ clone treated with dTAG-13 (100 nM), BJP-06-005-3 (5 μM), BJP-R (5 μM) or the indicated combinations for 4 d, with compound replenishment on day 2. The mean peak heights of $n = 3$ biologically independent samples are shown as dot-plots, and are representative of $n = 3$ independent experiments. dTAG-13 $^{**}P = 0.0017$, BJP-06-005-3 $^{**}P = 0.0022$, dTAG-13 + BJP-06-005-3 $^{***}P = 0.0005$, dTAG-13 + BJP-R $^{***}P = 0.0009$ by Welch's two-sided *t*-test, with comparisons made with DMSO. Uncropped immunoblots for **a** are displayed in Supplementary Fig. 12, and are representative of $n = 2$ independent experiments. Uncropped immunoblots for **b** are displayed in Supplementary Fig. 13, and are representative of $n = 3$ independent experiments. Uncropped immunoblots for **c** are displayed in Supplementary Fig. 13, and are representative of $n = 2$ independent experiments.

control, BJP-R, did not affect viability (Supplementary Fig. 7a). Across a panel of cancer cell lines, BJP-06-005-3 consistently reduced cell viability in a dose- and time-dependent manner (Supplementary Fig. 7b–g).

To confirm that these antiproliferative effects were Pin1-mediated, we compared the effects of BJP-06-005-3 in C1 (Pin1 WT) versus C2 (Pin1 KO). BJP-06-005-3 led to a dose- and time-dependent decrease in cell growth in C1, while it did not affect proliferation in C2 (Fig. 5a). We next transduced C2 with Dox-inducible Pin1 (Supplementary Fig. 7h), and found the resulting cells to be re-sensitized to BJP-06-005-3 in the presence of Dox (Fig. 5b). Furthermore, BJP-06-005-3 impacted viability in 3D spheroids after a 4-d treatment in PATU-8988T parental cells, but not in C2, suggesting a Pin1-mediated effect on cell transformation (Fig. 5c). By contrast, the previously reported covalent Pin1 inhibitors, juglone[10] and KPT-6566 (ref. [13]), induced identical antiproliferative effects in C1 and C2, indicating that their viability effects are likely mediated by off-targets. We were also unable to validate that KPT-6566 engages cellular Pin1 using our BJP-DTB competition assay (Supplementary Fig. 8a–d).

To complement the characterization of BJP-06-005-3, we employed our dTAG system, in which a heterobifunctional degrader (dTAG-13) co-opts cereblon to rapidly and selectively degrade FKBP12^{F36V}-tagged fusions[20]. To first establish that N-terminally tagged Pin1 (FKBP12^{F36V}-Pin1) was functional, we performed complementation experiments in *Saccharomyces cerevisiae*, in which the *PIN1* homolog, *ESS1*, is an essential gene[8]. Overexpression of Pin1 or FKBP12^{F36V}-Pin1 successfully supported growth of an *ESS1* deletion strain, and immunoblotting confirmed expression of the fusion protein (Supplementary Fig. 9a,b).

To implement the dTAG system in PDAC cells, we exogenously expressed FKBP12^{F36V}-Pin1 in C2 (referred to as C2 + FKBP12^{F36V}-Pin1) (Supplementary Fig. 9c). Treatment with dTAG-13 caused rapid and potent degradation of FKBP12^{F36V}-Pin1, with maximal degradation at 100–500 nM, and with the hook effect[34] apparent by 1 μM, at which the binary complexes outcompete formation of a productive ternary complex (Supplementary Fig. 9d,e). FKBP12^{F36V}-Pin1 degradation occurred within 2–4 h (Fig. 5d), and degradation was rescued upon pretreatment with carfilzomib (proteasome inhibitor) or MLN4924 (NEDD8-activating enzyme inhibitor), confirming a proteasome- and neddylation-dependent mechanism (Supplementary Fig. 9f).

Introduction of FKBP12^{F36V}-Pin1 into C2 restored the growth rate to parental levels (Fig. 5e), and prolonged FKBP12^{F36V}-Pin1 degradation diminished cell proliferation after 6–8 d. This phenotype was apparent upon treatment with 100 or 500 nM dTAG-13, but not at 2.5 μM, consistent with the appearance of the hook effect at higher doses (Supplementary Fig. 9g). Furthermore, while C2 was resistant to BJP-06-005-3, treatment with BJP-06-005-3 in FKBP12^{F36V}-Pin1-expressing C2 restored the time-dependent antiproliferative effects (Fig. 5f). FKBP12^{F36V}-Pin1 could also be pulled down by BJP-DTB, which was outcompeted upon BJP-06-005-3 pretreatment (Supplementary Fig. 9h). The effects of FKBP12^{F36V}-Pin1 degradation were therefore consistent with BJP-06-005-3 treatment, further validating our inhibitor's on-target pharmacology.

Cell cycle and transcriptional effects of BJP-06-005-3. We next explored whether the time-dependent viability effects upon Pin1 loss or inhibition might be explained by subtle changes in cell cycle progression, which can translate to exponential changes in cell number over time[35]. Given that Pin1 regulates the G2/M transition through its interaction with mitotic proteins, including Cdc25C and Wee1 (ref. [3]), we hypothesized that Pin1 might impact phosphorylation of CDK1 Tyr15, a critical regulator of mitotic entry and a downstream target of Wee1. In a previous study, depletion of Pin1-binding proteins from *Xenopus* egg extracts resulted in hyperphosphorylation of Cdc2 (ref. [36]). We found that phospho-CDK1 Tyr15 was elevated in C2 as compared with C1 (Fig. 6a). Similarly, BJP-06-005-3 increased phospho-CDK1 Tyr15 in C1 (Fig. 6b), but not in C2 (Supplementary Fig. 10a). dTAG-13-mediated degradation of FKBP12^{F36V}-Pin1 also increased phospho-CDK1 Tyr15 (Supplementary Fig. 10b).

To evaluate the cell cycle effects of Pin1 inhibition, we performed propidium iodide staining in PATU-8988T cells treated with BJP-06-005-3 for 4–48 h. At all time points tested, BJP-06-005-3 modestly decreased G2/M (statistically insignificant, Bonferroni-corrected two-way analysis of variance (ANOVA)) and increased G1-phase cells (adjusted (adj.) $P<0.05$ at all time points, Bonferroni-corrected two-way ANOVA). Similarly, degradation of FKBP12^{F36V}-Pin1 moderately decreased G2/M cells (adj. $P=0.0435$, Bonferroni-corrected two-way ANOVA) and increased G1 cells (adj. $P=0.0071$, Bonferroni-corrected two-way ANOVA) (Supplementary Fig. 10c–e).

To further explore the downstream consequences of Pin1 inhibition, we evaluated the effects of BJP-06-005-3 on reported Pin1 substrates, including ß-Catenin[37], NFκB p65 (ref. [38]) and c-Myc[39]. We observed no change in the expression of ß-Catenin or p65 upon BJP-06-005-3 treatment (Fig. 6b), which could be explained by low basal activation of Wnt/ß-Catenin and NFκB signaling in PATU-8988T cells. Furthermore, following recombinant Wnt3a treatment to increase nuclear ß-Catenin, BJP-06-005-3 did not affect expression or nuclear/cytoplasmic localization of ß-Catenin (Supplementary Fig. 10f–h). However, in accordance with the literature precedent that Pin1 enhances c-Myc turnover[39], BJP-06-005-3 treatment increased c-Myc protein levels (Fig. 6b), as did dTAG-13-mediated FKBP12^{F36V}-Pin1 degradation (Supplementary Fig. 10b).

To broadly profile the downstream changes upon covalent Pin1 inhibition, we assessed the transcriptional consequences of BJP-06-005-3 in PATU-8988T cells. We identified 3,220 significantly differentially expressed transcripts (2,587 up- and 633 down-regulated at 1.25-fold; $P<0.05$) following a 4-h treatment with BJP-06-005-3, normalized relative to DMSO and BJP-R (Supplementary Dataset 2 and Supplementary Fig. 11a). Gene set enrichment analysis revealed enrichment of several pathways in the differentially expressed transcripts, including GPCR ligand signaling, extracellular matrix organization and integrin cell surface interactions (Supplementary Fig. 11b–f). EnrichR[40] analysis of the differentially expressed genes to identify the transcription factors that might be coordinating these changes identified Myc target genes as amongst the most-enriched sets, suggesting that the transcriptional responses to BJP-06-005-3 may occur through reduction of Myc transcriptional activity (Supplementary Fig. 11g). This is consistent with literature reports that Pin1 enhances c-Myc DNA binding and oncogenic activity[41], and our observations of increased c-Myc turnover.

Given the critical role of Myc in Ras-driven cancer growth[42], we hypothesized that Pin1 inhibition may cooperate with KRAS loss. We assessed whether the antiproliferative effects of FKBP12^{F36V}-KRAS^{G12V} degradation[43] could be amplified by BJP-06-005-3. We used a PATU-8988T FKBP12^{F36V}-KRAS^{G12V}; *KRAS*$^{-/-}$ clone in which dTAG-13 treatment degrades FKBP12^{F36V}-KRAS^{G12V}, leading to modest changes in cell proliferation[43]. We confirmed that dTAG-13 treatment degraded FKBP12^{F36V}-KRAS^{G12V}, which was unchanged by cotreatment with BJP-06-005-3 or BJP-R (Fig. 6c). A 4-d treatment with either dTAG-13 or BJP-06-005-3 alone led to modest viability effects, while their combination had enhanced antiproliferative activity (Fig. 6d), suggesting that the viability effects of mutant KRAS degradation are amplified by Pin1 inhibition in PATU-8988T cells.

Discussion

Despite decades of research highlighting Pin1 as an oncogenic driver, the development of selective Pin1 inhibitors has proven challenging. Achieving potency generally requires anionic compounds to interact with Pin1's phosphate-binding pocket, diminishing cell permeability[22,44]. On the other hand, reported cell-penetrant Pin1 inhibitors are nonselective. Numerous studies feature juglone, which has known off-targets such as RNA polymerase II (ref. [45]); ATRA[11], which degrades Pin1 but also targets the retinoic acid receptors; and KPT-6566 (ref. [13]), which releases a quinone-mimetic substructure after Pin1 binding that engages a host of other proteins to induce an oxidative stress response. This study describes a selective, covalent Pin1 inhibitor, BJP-06-005-3, suitable for investigating Pin1 biology. The selectivity of BJP-06-005-3 is currently unparalleled, with Pin1 Cys113 being the only competitively labeled cysteine proteome-wide.

We found Pin1 inhibition to be largely nontoxic after 4-d treatments in 2D cell culture. This result was validated using the dTAG system, highlighting that BJP-06-005-3's negligible acute cytotoxicity could not be explained by differences in Pin1 inhibition versus degradation. Given the critical role of Pro-directed phosphorylation, this lack of cytotoxicity is surprising and may be explained by upregulation of another PPIase to compensate for Pin1 inhibition, though this remains to be explored.

We also demonstrated that Pin1 cooperates with activated KRAS to drive cell transformation, and that Pin1 loss in the context of 3D spheroids induced more pronounced antiproliferative effects than in 2D monolayers. However, Pin1 inhibition did impact cancer cell viability in 2D cell culture following prolonged treatments, suggesting that Pin1-mediated growth defects become more pronounced over time. Accordingly, we observed a time-dependent growth dependence on Pin1 using a CRISPR-based GFP dropout assay. However, this growth effect was modest in comparison to that of an essential protein, 40S ribosomal protein S19 (RPS19). We hypothesize that the time-dependent cytotoxicity observed upon Pin1 loss may be explained by minor defects in cell cycle progression, though this requires further study.

Collectively, this study suggests that Pin1 inhibition could prove a viable therapeutic strategy in PDAC and highlights that 3D cell culture models and prolonged treatments best predict Pin1 dependency. Future efforts to cyclize and/or depeptidize BJP-06-005-3 could improve its cell permeability and stability. Also, studies to identify synthetic lethal interactions may reveal suitable drug combinations to enhance therapeutic efficacy. We envision that BJP-06-005-3 and future analogs, as well as the genetic and chemical-genetic tools described herein, will further our understanding of Pin1 biology in disease-relevant contexts.

Online content

Any Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-020-0550-9.

Received: 9 May 2019; Accepted: 15 April 2020;
Published online: 01 June 2020

References

1. Lu, Z. & Hunter, T. Prolyl isomerase Pin1 in cancer. *Cell Res.* **24**, 1033–1049 (2014).
2. Lu, K. P., Finn, G., Lee, T. & Nicholson, L. K. Prolyl *cis-trans* isomerization as a molecular timer. *Nat. Chem. Biol.* **3**, 619–629 (2007).
3. Yeh, E. S. & Means, A. R. Pin1, the cell cycle and cancer. *Nat. Rev.* **7**, 381–388 (2007).
4. Lu, K. P. & Zhou, X. Z. The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nat. Rev. Cancer* **16**, 463–478 (2016).
5. Liang, C. et al. PIN1 maintains redox balance via the c-Myc/NRF2 axis to counteract Kras-induced mitochondrial respiratory injury in pancreatic cancer cells. *Cancer Res.* **79**, 133–145 (2019).
6. Wulf, G., Garg, P., Liou, Y. C., Iglehart, D. & Lu, K. P. Modeling breast cancer *in vivo* and *ex vivo* reveals an essential role of Pin1 in tumorigenesis. *EMBO J.* **23**, 3397–3407 (2004).
7. Zeitouni, D., Pylayeva-Gupta, Y., Der, C. J. & Bryant, K. L. KRAS mutant pancreatic cancer: no lone path to an effective treatment. *Cancers (Basel)* **8**, 45 (2016).

8. Hanes, S. D. The Ess1 prolyl isomerase: traffic cop of the RNA polymerase II transcription cycle. *Biochim. Biophys. Acta* **4**, 316–333 (2014).
9. Liou, Y. C. et al. Loss of Pin1 function in the mouse causes phenotypes resembling cyclin D1-null phenotypes. *Proc. Natl Acad. Sci. USA* **99**, 1335–1430 (2001).
10. Hennig, L. et al. Selective inactivation of parvulin-like peptidyl-prolyl *cis*/*trans* isomerases by juglone. *Biochemistry* **37**, 5953–5960 (1998).
11. Wei, S. et al. Active Pin1 is a key target of all-*trans* retinoic acid in acute promyelocytic leukemia and breast cancer. *Nat. Med.* **21**, 457–466 (2015).
12. Kozono, S. et al. Arsenic targets Pin1 and cooperates with retinoic acid to inhibit cancer-driving pathways and tumor-initiating cells. *Nat. Commun.* **9**, 3069 (2018).
13. Campaner, E. et al. A covalent PIN1 inhibitor selectively targets cancer cells by a dual mechanism of action. *Nat. Commun.* **8**, 15772 (2017).
14. Ieda, N. et al. An irreversible inhibitor of peptidyl-prolyl *cis*/*trans* isomerase Pin1 and evaluation of cytotoxicity. *Bioorg. Med. Chem. Lett.* **29**, 353–356 (2018).
15. Moore, J. D. & Potter, A. Pin1 inhibitors: pitfalls, progress and cellular pharmacology. *Bioorg. Med. Chem. Lett.* **23**, 4283–4291 (2013).
16. Mah, R., Thomas, J. R. & Shafer, C. M. Drug discovery considerations in the development of covalent inhibitors. *Bioorg. Med. Chem. Lett.* **24**, 33–39 (2014).
17. Liu, Q. et al. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **20**, 146–159 (2013).
18. Chen, C. H. et al. Pin1 cysteine-113 oxidation inhibits its catalytic activity and cellular function in Alzheimer's disease. *Neurobiol. Dis.* **76**, 13–23 (2015).
19. Zhang, Y. et al. Structural basis for high-affinity peptide inhibition of human Pin1. *ACS Chem. Biol.* **2**, 320–328 (2007).
20. Nabet, B. et al. The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
21. Backus, K. M. et al. Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).
22. Guo, C. et al. Structure-based design of novel human Pin1 inhibitors (III): optimizing affinity beyond the phosphate recognition pocket. *Bioorg. Med. Chem. Lett.* **24**, 4187–4191 (2014).
23. Yang, N. J. & Hinner, M. J. Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods Mol. Biol.* **1266**, 29–53 (2015).
24. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
25. Valley, C. C. et al. The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem.* **287**, 34979–34991 (2012).
26. Browne, C. M. et al. A chemoproteomic strategy for direct and proteome-wide covalent inhibitor target-site identification. *J. Am. Chem. Soc.* **141**, 191–203 (2019).
27. Long, M. J., Gollapalli, D. R. & Hedstrom, L. Inhibitor mediated protein degradation. *Chem. Biol.* **19**, 629–637 (2012).
28. Luo, M. L. et al. Prolyl isomerase Pin1 acts downstream of miR-200 to promote cancer stem-like cell traits in breast cancer. *Cancer Res.* **74**, 3603–3616 (2014).
29. Rotem, A. et al. Alternative to the soft-agar assay that permits high-throughput drug and genetic screens for cellular transformation. *Proc. Natl Acad. Sci. USA* **112**, 5708–5713 (2015).
30. Janes, M. R. et al. Targeting KRAS mutant cancers with a covalent G12C-specific inhibitor. *Cell* **172**, 578–589 (2018).
31. Shi, J. et al. Discovery of cancer drug targets by CRISPR–Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
32. Erb, M. A. et al. Transcription control by the ENL YEATS domain in acute leukemia. *Nature* **543**, 270–274 (2017).
33. Behrsin, C. D. et al. Functionally important residues in the peptidyl-prolyl isomerase Pin1 revealed by unigenic evolution. *J. Mol. Biol.* **365**, 1143–1162 (2007).
34. An, S. & Liwu, F. Small-molecule PROTACs: an emerging and promising approach for the development of targeted therapy drugs. *EBioMedicine* **36**, 553–562 (2018).
35. Tzur, A., Kafri, R., LeBleu, V. S., Lahav, G. & Kirschner, M. W. Cell growth and size homeostasis in proliferating animal cells. *Science* **325**, 167–171 (2009).
36. Crenshaw, D. G., Yang, J., Means, A. R. & Kornbluth, S. The mitotic peptidyl-prolyl isomerase, Pin1, interacts with Cdc25 and Plx1. *EMBO J.* **17**, 1315–1327 (1998).
37. Ryo, A., Nakamura, M., Wulf, G., Liou, Y. C. & Lu, K. P. Pin1 regulates turnover and subcellular localization of β-catenin by inhibiting its interaction with APC. *Nat. Cell Biol.* **9**, 793–801 (2001).
38. Ryo, A. et al. Regulation of NF-κB signaling by Pin1-dependent prolyl isomerization and ubiquitin-mediated proteolysis of p65/RelA. *Mol. Cell* **12**, 1413–1426 (2003).
39. Yeh, E. et al. A signaling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat. Cell Biol.* **6**, 308–318 (2004).
40. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
41. Farrell, A. S. et al. Pin1 regulates the dynamics of c-Myc DNA binding to facilitate target gene regulation and oncogenesis. *Mol. Cell. Biol.* **33**, 2930–2949 (2013).
42. Vaseva, A. V. et al. KRAS suppression-induced degradation of MYC is antagonized by a MEK5-ERK5 compensatory mechanism. *Cancer Cell* **34**, 807–822 (2018).
43. Ferguson, F. M. et al. Discovery of a selective inhibitor of doublecortin like kinase 1. *Nat. Chem. Biol.* https://doi.org/10.1038/s41589-020-0506-0 (2020).
44. Spena, C. R. et al. Liposomal delivery of a Pin1 inhibitor complexed with cyclodextrins as new therapy for high-grade serious ovarian cancer. *J. Control. Release* **281**, 1–10 (2018).
45. Chao, S. H., Greenleaf, A. L. & Price, D. H. Juglone, an inhibitor of the peptidyl-prolyl isomerase Pin1, also directly blocks transcription. *Nucleic Acids Res.* **29**, 767–773 (2001).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

FP binding assay. Binding affinity to Pin1 was determined using an FP assay to assess competition with an N-terminal fluorescein-labeled peptide (Bth-D-phos. Thr-Pip-Nal), which was synthesized by a peptide synthesis company. The indicated concentrations of candidate compound were pre-incubated for 12 h at 4 °C with a solution containing 250 nM glutathione *S*-transferase (GST)-Pin1, 5 nM fluorescein-labeled peptide probe, 10 μg ml^{-1} BSA, 0.01% Tween-20 and 1 mM DTT in a buffer of 10 mM HEPES, 10 mM NaCl and 1% glycerol (pH 7.4). Measurements of FP were made in black 384-well plates (Corning) using an EnVision reader. K_i values obtained from the FP assay results were derived from the Kenakin K_i equation: Kenakin $K_i = (Lb)(EC_{50})(K_d)/(Lo)(Ro) + Lb(Ro - Lo + Lb - K_d)$, where K_d [M], K_d of the probe; EC_{50} [M], concentration of unlabeled compound that results in 50% inhibition of binding (obtained from FP assay); total tracer Lo [M], probe concentration in FP; bound tracer Lb [M], 85%, fraction of probe bound to Pin1; and total receptor Ro [M], Pin1 concentration in the FP assay, as described[11,46]. To determine the k_{inact}/K_i, the FP assay was performed following 0-, 5-, 10-, 15-, 20-, 25- or 30-min incubation times of recombinant Pin1 and BJP-06-005-3 (in a 9-point dose course). The observed rate constant for inhibition, k_{obs}, was determined at each concentration, and was plotted against inhibitor concentration and fitted to a hyperbolic equation (GraphPad software), according to the equation $k_{obs} = k_{inact}[I]/(K_i + [I])$, as described previously[13].

Chemical synthesis. Synthetic procedures are provided in the Supplementary Note.

PPIase isomerase inhibition assay. Inhibition of Pin1 isomerase activity was determined using the chymotrypsin-coupled PPIase assay, using GST-Pin1 and Suc-Ala-phos.Ser-Pro-Phe-pNA peptide substrate, as described previously[11,47]. GST-Pin1 was pre-incubated with the indicated concentrations of compound for 12 h at 4 °C in buffer containing 35 mM HEPES (pH 7.8), 0.2 mM DTT and 0.1 mg ml^{-1} BSA. Immediately before the assay was started, chymotrypsin (final concentration of 6 mg ml^{-1}), followed by the peptide substrate (Suc-Ala-pSer-Pro-Phe-pNA peptide substrate, final concentration 50 mM), was added. The K_i value obtained from the PPIase assay was derived from the Cheng–Prusoff equation, $K_i = IC_{50}/(1 + S/K_m)$, where K_m is the Michaelis constant for the peptide substrate, S is the initial concentration of the substrate in the assay and IC_{50} is the half-minimal inhibitory concentration of the inhibitor.

Covalent labeling by intact mass spectrometry. First, 5 μg of purified Pin1 protein in 50 μl of 20 mM HEPES pH 7.5 and 75 mM NaCl was incubated with 5 μM of respective Pin1 inhibitors for 0–3 h. A Shimadzu XR HPLC device was used to inject the entire sample onto a self-packed reverse-phase column (1/32 in outer diameter × 500-μm inner diameter, 5 cm of POROS 50R2 resin). After desalting, protein was eluted with an HPLC gradient (0% to 100% B in 4 min; A, 0.2 M acetic acid in water; B, 0.2 M acetic acid in acetonitrile (MeCN); flow rate = 10 μl min^{-1}) into an LTQ XL mass spectrometer (Thermo Fisher Scientific). LTQ XL MS spectra were acquired in centroid mode using the electron multipliers for ion detection. Mass spectra were deconvoluted using MagTran1.03b2 software[48].

Recombinant protein labeling site identification. Pin1 protein was incubated with indicated compounds for 1 h at room temperature in 20 mM HEPES, 75 mM NaCl and 1 mM tris(2-carboxyethyl)phosphine (TCEP) buffer. Protein was reduced with 10 mM TCEP at room temperature for 30 min, alkylated with 20 mM iodoacetamide at room temperature for 30 min, then digested with trypsin at 37 °C overnight. Desalted peptides were analyzed by capillary electrophoresis–mass spectrometry (CE-MS) using a ZipChip CE system and autosampler (908 Devices) interfaced to a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). Protein solutions were loaded for 30 s and separation performed at 500 V cm^{-1} for 6 min using an HR chip (908 devices) (22-cm separation channel) with a background electrolyte composed of 1% formic acid in 50% MeCN. Pressure assist was used and started at 1 min. The mass spectrometer was operated in data-dependent mode and subjected the five most abundant ions in each MS precursor scan (15,000 mass resolution, 1,000,000 target intensity threshold, lock mass enabled) to tandem MS (MS/MS) (15,000 mass resolution, 200,000 target intensity threshold, 100-ms maximum injection time, 30 normalized collision energy). Dynamic exclusion was enabled with a repeat count of 1 and an exclusion time of 6 s. MS/MS spectra of inhibitor-modified peptides were annotated using mzStudio[49].

Cell culture. PATU-8988T (DSMZ), PATU-8902 (DSMZ), HeLa (ATCC), MDA-MB-468 (ATCC), 293FT (Thermo Fisher Scientific), Phoenix-AMPHO (ATCC) and HEK293 (ATCC) cells were cultured in DMEM supplemented with 10% FBS (Sigma) and 1% penicillin/streptomycin. BxPC-3 (ATCC), PC3 (ATCC), NB4 (ATCC) and Kuramochi (Panagiotis A. Konstantinopoulos's laboratory) cells were cultured in RPMI supplemented with 10% FBS (Sigma) and 1% penicillin/streptomycin. NIH/3T3 (ATCC) cells were cultured in DMEM supplemented with 10% calf serum (Sigma) and 1% penicillin/streptomycin. All cell lines were cultured at 37 °C in a humidified chamber. Cell lines were routinely tested for the absence of *Mycoplasma* infection.

Establishment of Pin1 KO cell lines using CRISPR–Cas9. Pin1-targeting guides were annealed into the BbsI restriction site of the pX458 plasmid (Addgene cat. no. 48138). The guide RNA sequence used was: CAGTGGTGGCAAAAACGGGC. Cells were transfected using the Neon Transfection System (Life Technologies) according to manufacturer guidelines, and GFP$^+$ cells were sorted using FACS using a J AriaII SORP UV machine (BD). Single-cell clones were isolated from the bulk-sorted population using limiting dilution plating in a 96-well plate. Once clones were isolated, they were verified for Pin1 KO by western blotting and by genomic DNA PCR using the following primers: forward: GAGCCTGTGGCACATGGTG; reverse: CAGGGTCAGGTCATGCACTG; sequencing: CTGGCTTCTGGCTGTG; and sequences were analyzed using Tracking of Indels by Decomposition[50].

Generation of stable Dox-inducible and FKBP12^{F36V}-Pin1 cell lines. The Dox-inducible Pin1 construct (pCW57.1-Pin1) was generated by using gateway recombination cloning technology (Invitrogen) to insert the Pin1 coding sequence from a pDONR221-Pin1 donor vector into the pCW57.1 plasmid (Addgene cat. no. 41393) as described previously[20]. The Pin1 sequence in the donor vector was rendered insensitive to the Pin1 guide RNA using QuikChange II Mutagenesis (Agilent cat. no. 200521) according to manufacturer's instruction with the following primers: forward: GTGGCAAAAACGGGCAAGGGGAGCCTGCCAGGGTC; and reverse: GACCCTGGCAGGCTCCCCTTGCCCGTTTTTGCCAC. The N-terminal FKBP12^{F36V}-Pin1 construct (pLEX_305-N-dTAG-Pin1) was generated by using gateway recombination cloning technology (Invitrogen) to insert the Pin1 coding sequence from a pDONR221-Pin1 donor vector into pLEX_305-N-dTAG as described previously[20]. To generate lentivirus, 293FT cells were cotransfected with 2.5 μg VSV-G pMD2 (Addgene no. 12259), 3.75 μg psPAX2 (Addgene no. 12260) and 5 μg expression plasmid (either pCW57.1-Pin1 or pLEX_305-N-dTAG-Pin1) using Lipofectamine 2000 according to manufacturer protocol (Thermo Fisher Scientific). Viral particles were collected 48 h after transfection and filtered through a 0.45-μm membrane. Viral supernatants were applied to PATU-8988T Pin1$^{-/-}$ C2 or NIH/3T3 cells in the presence of 8 μg ml^{-1} polybrene, or NIH/3T3 FKBP12^{F36V}-KRAS^{G12V} cells in the presence of 4 μg ml^{-1} polybrene, and transduced cells were selected with 2 μg ml^{-1} puromycin.

Generation of stable Pin1 mutant cell lines. Cloning of pBybe Puro Pin1 WT, pBabe Hygro Pin1 W34A and pBabe Hygro Pin1 K63A was described previously[28]. pBybe Pin1 C113S was generated by using QuikChange II (Agilent) according to manufacturer protocol, and sequence was confirmed using the M13-F primer from Genewiz. Lentivirus production and concentration were performed upon transfection of 293FT cells with pBybe Puro Pin1 WT or pBybe Puro Pin1 C113S as previously described[43]. Concentrated viral supernatants were applied to PATU-8988T Pin1$^{-/-}$ C2 cells in the presence of 4 μg ml^{-1} polybrene, and transduced cell lines were selected with 2 μg ml^{-1} puromycin. Retrovirus production was performed using Phoenix-AMPHO cells, which were transfected with pBabe Hygro Pin1 W34A or pBabe Hygro Pin1 K63A using Fugene 6 (Promega), according to the manufacturer's instructions. Viral particles were collected 48 h after transfection, filtered through a 0.45-μm membrane and concentrated using 1:3 Retro-X Concentrator (Clontech), according to the manufacturer's instructions. Concentrated viral supernatants were applied to PATU-8988T Pin1$^{-/-}$ C2 cells in the presence of 4 μg ml^{-1} polybrene, and transduced cells were selected with 250 μg ml^{-1} hygromycin.

Generation of stable FKBP12^{F36V}-KRAS^{G12V} cell lines. Generation of NIH/3T3 FKBP12^{F36V}-KRAS^{G12V} cells was previously described[20]. Generation of PATU-8988T FKBP12^{F36V}-KRAS^{G12V}; *KRAS*$^{-/-}$ cells was also previously described[43].

GFP dropout assay. Plasmids were generated using oligos from IDT, annealed according to the manufacturer protocol and ligated into the BbsI-digested pLRG plasmid (Addgene no. 65656). GFP CRISPR dropout assay was conducted as described previously[31] with the following modifications. Lentivirus production and concentration were performed upon transfection of 293FT cells with lentiCas9-blast (Addgene, cat. no. 52962), as previously described[20]. Concentrated viral supernatants were applied to PATU-8988T cells in the presence of 4 μg ml^{-1} polybrene, and transduced cell lines were selected with 10 μg ml^{-1} blasticidin. PATU-8988T cells constitutively expressing Cas9 were infected with pLRG-sgPin1 (1–8) virus, generated (as described in the section pertaining to "generation of stable Dox-inducible and FKBP12^{F36V}-Pin1 cell lines") and plated in triplicate in a 96-well plate. At 3 d after plating, GFP percentages of all cells were assessed using the following protocol: cells were trypsinized using 30 μl of 0.5% trypsin after medium was aspirated from all wells. Trypsin was neutralized with the addition of 105 μl of medium. Then, 10 μl of each cell suspension was transferred to a new 96-well plate into 90 μl of medium to continue the experiment. The remaining cells were diluted once with PBS and transferred to a 96-well V-bottom plate (Sigma, cat. no. CLS3894). The percentage of GFP$^+$ cells was then assessed using a Guava

EasyCyte 8HT (Millipore). Infection was performed such that the initial percentage of GFP$^+$ cells measured on day 3 was roughly 50%. GFP percentage was assessed every 4–5 d following initial reading. GFP$^+$ cells were quantified using the Guava InCyte software.

Immunoblotting. Whole-cell lysates for immunoblotting were prepared by pelleting cells from each cell line at 4 °C (300*g*) for 5 min. The resulting cell pellets were washed once with ice-cold PBS and then resuspended in the indicated cell lysis buffer. Lysates were clarified at 14,000 r.p.m. for 15 min at 4 °C before quantification by the bicinchoninic acid (BCA) assay (Pierce, cat. no. 23225). Whole-cell lysates were loaded into Bolt 4–12% Bis-Tris Gels (Thermo Fisher, cat. no. NW04120BOX) and separated by electrophoreses at 95 V for 1.5 h. The gels were transferred to a nitrocellulose membrane using the iBlot Gel Transfer at a voltage of 10 V for 6 min (Thermo Fisher, cat. no. IB23001) and then blocked for 1 h at room temperature in Odyssey blocking buffer (LICOR Biosciences, cat. no. 927-50010). Membranes were probed using antibodies against the relevant proteins at 4 °C overnight in 20% Odyssey Blocking Buffer in 1× TBST buffer. Membranes were then washed three times with 1× TBST (at least 5 min per wash) followed by incubation with the IRDye goat anti-mouse (LICOR, cat. no. 926-32210) or goat anti-rabbit (LICOR, cat. no. 926-32211) secondary antibody (diluted 1:10,000) in 20% Odyssey Blocking Buffer in 1× TBST for 1 h at room temperature. After three washes with 1× TBST (at least 5 min per wash), the immunoblots were visualized using the Odyssey Infrared Imaging System (LICOR). Antibodies used against various proteins were as follows: Pin1 (1:1,000, Cell Signaling cat. no. 3722), α-Tubulin (1:1,000, Cell Signaling cat. no. 3873), phospho-cdc2 (Tyr15) (1:1,000, Cell Signaling cat. no. 4539), CDK1 (1:1,000, Abcam cat. no. ab131450), HA-Tag (1:1,000, Cell Signaling cat. no. 3724), ß-catenin (1:1,000, Cell Signaling cat. no. 8480), NFκB p65 (1:1,000, Cell Signaling cat. no. 6956), c-Myc (1:1,000, Cell Signaling cat. no. 5605), GAPDH (1:1,000, Cell Signaling cat. no. 2118), ß-Actin (1:1,000, Cell Signaling cat. no. 3700) and histone H3 (1:1,000, Cell Signaling cat. no. 9715).

Lysate pull-down with BJP-DTB. Cells were lysed in 50 mM HEPES pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM TCEP, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40 and protease inhibitor tablet (Roche cat. no. 4693159001) for 20 min on ice. After clarifying (14,000 r.p.m. for 15 min), lysates were incubated with the indicated concentration of BJP-DTB for the indicated time points at 4 °C, using 500 µg of protein per sample. Lysates were then incubated with streptavidin agarose resin (30 µl of 1:1 beads/lysis buffer slurry) (Thermo scientific, cat. no. 20349) for 1 h at 4 °C. Beads were washed four times with 500 µl of washing buffer (50 mM HEPES pH 7.5, 10 mM NaCl, 1 mM EDTA, 10% glycerol), then pelleted by centrifugation and dried. The beads were boiled for 5 min at 95 °C in 30 µl of 2× NuPAGE Lithium Dodecyl Sulfate (LDS) (Thermo scientific, cat. no. NP0007) + 5% β-mercaptoethanol. Lysates were probed for specified proteins by western blotting using the Bolt system (Life Technologies).

Cellular target engagement—competition with BJP-DTB. PATU-8988T, PC3 or MDA-MB-468 cells were plated in 10-cm plates with 2.5 million cells per plate in 6 ml of media. The day after plating, cells were treated with the indicated concentrations of candidate inhibitor for the indicated time points. The cells were then washed two times with cold PBS (1 ml per 10-cm plate) and collected by scraping with a cell scraper. Cells were lysed in 50 mM HEPES pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM TCEP, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40 and protease inhibitor tablet (Roche), using 210 µl of cell lysis buffer per 10-cm plate of cells. After clarifying (14,000 r.p.m. for 15 min), 5 µl of each lysate sample was combined with 5 µl of 2× LDS + 5% β-mercaptoethanol, boiled for 5 min and set aside for the input loading control (later to be loaded directly in the gel). Then, 200 µl of each lysate sample was incubated with 1 µM of BJP-DTB for 1 h at 4 °C and processed as in the section "Lysate pull-down with BJP-DTB".

Pin1 stability assays: cycloheximide (CHX) chase. PATU-8988T cells were plated in six-well plates at a density of 50,000 cells per well in 2 ml of medium. The day after plating, cells were treated with either DMSO or BJP-06-005-3 for 24 h. The medium was then aspirated and replaced with fresh medium containing cycloheximide (100 µg ml^{-1}) to block new protein synthesis. Cells were collected at the indicated time points by washing two times with ice-cold 1× PBS, and then lysing in the plate with 100 µl of RIPA lysis buffer per well (Sigma, cat. no. R0278), supplemented with protease and phosphatase inhibitor tablets (Roche cat. no. 4906845001). Samples were normalized and prepped in 4× LDS + 10% β-mercaptoethanol and boiled for 5 min at 95 °C. Lysates were probed for specified proteins by western blotting using the Bolt system (Life Technologies).

BJP-06-005-3 stability studies. PATU-8988T cells were plated at a density of 25,000 cells per well in a 24-well plate, with 1 ml of medium per well. The day after plating, cells were treated with BJP-06-005-3 (10 µM). Then, 100 µl of medium was collected off of the cells at the indicated time points. In parallel, pure medium (DMEM + 10% FBS + 1% penicillin/streptomycin) was incubated with 10 µM of BJP-06-005-3 at room temperature, with 100 µl of medium collected at the indicated time points. Then, 1 µl of each medium sample was Ziptip (Millipore) C18 desalted, mixed with α-Cyano-4-hydroxycinnamic acid matrix and spotted for MALDI MS analysis. Samples were analyzed on an Ab Sciex 4800 MALDI TOF/TOF. Maximum intensity of the BJP-06-005-3 peak was measured in triplicate and averaged for each sample. Decay curves were generated using Prism. $N = 3$ biological replicates were used for each treatment.

Cell cycle analysis. PATU-8988T cells (parental or C2 + FKBP12^{F36V}-Pin1) were plated at 100,000 cells per well in a six-well plate, with 2 ml of medium per well. The day after plating, cells were treated with the indicated concentrations of BJP-06-005-3 or dTAG-13. At the indicated time points, cells were collected using trypsin, and then washed three times with cold PBS. Cell pellets were then fixed with 1 ml of cold 80% ethanol in PBS at 4 °C overnight. The cell pellets were then washed by adding 1 ml of PBS and then centrifuging (1,500 r.p.m. for 5 min), followed by aspiration of the PBS/ethanol mix. Cells were resuspended in cell cycle staining solution (100 µg ml^{-1} RNAse A (Roche, 10109169001) and 50 µg ml^{-1} propidium iodide (Life Technologies, P1304MP), diluted in PBS) and incubated for 10 min at room temperature. Cell cycle data were acquired by flow cytometry on a Guava easyCyte flow cytometer (Millipore) using the InCyte software. Data were analyzed using FlowJo. $N = 3$ biological replicates were used for each treatment condition.

Cell viability assays: growth over time in 2D adherent monolayer cell culture. Cells were plated at a density of 100 cells per well in 100 µl of medium in a 96-well white, clear-bottom plate (Corning cat. no. 3903), with at least one plate per time point (Day 0, 2, 4, 6 and so on). Cells were treated the day after plating with 1 µl of DMSO, BJP-06-005-3, BJP-R or dTAG-13 to give the indicated concentrations, and were then incubated at 37 °C and 5% CO_2. Every 48 h, the medium was aspirated and replaced with fresh medium containing fresh compound or DMSO. For the Dox-inducible PATU-8988T Pin1$^{-/-}$ C2 cells, the medium was replenished with Dox (100 ng ml^{-1}), in addition to compound or DMSO, every 48 h. For PATU-8988T Pin1$^{-/-}$ C2 cells expressing Pin1 WT, C113S, W34A or K63A, the medium was replenished after 96 h. When the indicated time points had been reached, cell viability was evaluated using the CellTiter-Glo Luminescent Cell Viability Assay (Promega cat. no. G7570) according to the manufacturer's standards, measuring luminescence using an Envision plate-reader. To normalize for differences in plating density between cell lines, a duplicate Day 0 plate of cells was plated at the beginning of the experiment and exposed using CellTiter-Glo (Promega cat. no. G7570) the day after plating, before compound treatment. At least $n = 3$ biological replicates were used for each treatment condition.

Cell viability assays: 4-d treatment in 2D adherent monolayer cell culture. PATU-8988T cells were plated in flat-bottom 96-well plates (Corning cat. no. 3903) at a density of 1,000 cells per well in 100 µl of medium and were treated the next day with 1 µl of the indicated compounds in a threefold dilution series. The cells were incubated at 37 °C and 5% CO_2 and, when indicated, every 24 h the medium was aspirated and replaced with fresh medium/re-treated with fresh compound to ensure maximal Pin1 engagement for the duration of the experiment. At 96 h after the initial treatment, antiproliferative effects were assessed by CellTiter-Glo Luminescent Cell Viability Assay (Promega cat. no. G7570) according to the manufacturer's standards, measuring luminescence using an Envision plate-reader. $N = 3$ biological replicates were used for each treatment condition.

Inhibitor effects on Pin1 downstream signaling and Pin1 degradation. Cells were plated in six-well plates at a density of 100,000 cells per well in 2 ml of medium. The day after plating, cells were treated with DMSO or the indicated concentrations of BJP-06-005-3, BJP-R, dTAG-13 or Dox. For the Wnt3a experiment (Supplementary Fig. 10f), cells were treated for 4 h with DMSO, BJP-06-005-3 or BJP-R, followed by treatment with either DMSO or recombinant human Wnt3a (50 ng ml^{-1}) (R&D Systems, cat. no. 5036) for an additional 4 h. Cells were collected at the indicated time points by washing two times with ice-cold 1× PBS, and then lysing in the plate with 100 µl of RIPA lysis buffer per well (Sigma, cat. no. R0278) supplemented with protease and phosphatase inhibitor tablets (Roche cat. no. 4906845001). Samples were normalized and prepped in 4× LDS + 10% β-mercaptoethanol and boiled for 5 min at 95 °C. Lysates were probed for specified proteins by western blotting using the Bolt system (Life Technologies).

ULA 3D spheroid assay. Cells were plated in clear, round-bottom, ULA 96-well plates (Corning cat. no. 07201680) at a density of 1,000 cells per well in 100 µl of medium. After plating, cells were spun at 300*g* for 5 min. The day after plating, cells were treated with DMSO or the indicated concentrations of BJP-06-005-3. At 48 h after the initial treatment, cells were re-treated with DMSO or BJP-06-005-3. At 96 h after the initial treatment, cell viability was evaluated using CellTiter-Glo (Promega cat. no. G7570). To normalize for differences in plating density between cell lines, a duplicate Day 0 plate of cells was plated at the beginning of the experiment and exposed using CellTiter-Glo (Promega cat. no. G7570) the day after plating. The final absorbance was then divided by the average Day 0 reading to normalize according to initial plating density. For direct comparisons to growth in 2D adherent monolayer cell culture, plating conditions and treatments were performed identically as described for the 3D spheroid assay, except that cells

were plated in flat-bottom 96-well plates (Corning cat. no. 3903). $N=3$ biological replicates were used for each treatment condition.

Nuclear/cytoplasmic fractionation. PATU-8988T cells were plated at a density of 2 million cells per 10-cm plate in 6 ml of medium. The next day, cells were treated with DMSO, BJP-06-005-3 or BJP-R at the indicated concentrations for 5.5 h. Cells were collected with 0.25% trypsin-EDTA (Thermo Fisher Scientific) (2 ml per 10-cm plate), washed once with 1 ml of PBS and then processed according to the manufacturer protocol using the Thermo Scientific NE-PER Nuclear and Cytoplasmic Extraction Kit (cat. no. 78833). For the Wnt3a treatment and fractionation experiment (Supplementary Fig. 10h), cells were treated for 4 h with DMSO, BJP-06-005-3 or BJP-R at the indicated concentrations, followed by treatment with either DMSO or recombinant human Wnt3a (50 ng ml^{-1}) (R&D Systems, cat. no. 5036) for an additional 4 h, and then cells were collected and processed as described in the beginning of this "nuclear/cytoplasmic fractionation" section.

RNA sequencing. PATU-8988T cells were plated at a density of 100,000 cells per well in 2 ml of medium per well of a six-well plate. The next day, cells were treated for 4 h with DMSO, 10 µM BJP-06-005-3 or 10 µM BJP-R. To collect, cells were washed twice with 1 ml of PBS per well and lysed in Buffer RLT Plus (QIAGEN). RNA was extracted using the QIAGEN RNeasy Plus Mini Kit (cat. no. 74134) according to the manufacturer protocol. Samples were sequenced by Invitrogen using the Ion AmpliSeq platform. Analysis of the mapped reads was performed in R using the limma framework[51].

Protein expression and purification. A construct of full-length human Pin1 in a pET28 vector was overexpressed in *Escherichia coli* BL21 (DE3) in LB medium in the presence of 50 mg ml^{-1} kanamycin. Cells were grown at 37 °C to an optical density of 0.8, cooled to 17 °C, induced with 500 µM isopropyl-1-thio-D-galactopyranoside, incubated overnight at 17 °C, collected by centrifugation and stored at −80 °C. Cell pellets were sonicated in buffer A (50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 10 mM Imidazole and 3 mM ß-mercaptoethanol) and the resulting lysate was centrifuged at 30,000*g* for 40 min. Ni-NTA beads (Qiagen) were mixed with lysate supernatant for 30 min and washed with buffer A. Beads were transferred to a fast protein liquid chromatography (FPLC)-compatible column and the bound protein was washed with 15% buffer B (50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 300 mM Imidazole and 3 mM ß-mercaptoethanol) and eluted with 100% buffer B. Thrombin was added to the eluted protein and incubated at 4 °C overnight. The sample was concentrated and passed through a Superdex 200 10/300 column (GE Healthcare) in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 3 mM DTT and 1 mM TCEP. Fractions were pooled, concentrated to approximately 37 mg ml^{-1} and frozen at −80 °C.

Crystallization: BJP-02-118-2. Apo protein at a final concentration of 1 mM was crystallized by sitting-drop (200 nl + 200 nl) vapor diffusion at 20 °C in the following crystallization buffer: 3 M NH_4SO_4, 100 mM BisTris pH 7.0, 1% PEG400 and 1 mM DTT. A volume of 200 nl of 1 mM BJP-02-118-2 was added directly to crystals for soaking at 20 °C for 16 h. Crystals were transferred briefly into crystallization buffer containing 25% glycerol before flash-freezing in liquid nitrogen.

Crystallization: BJP-07-017-3. Apo protein at a final concentration of 1 mM was crystallized by sitting-drop (200 nl + 200 nl) vapor diffusion at 20 °C in the following crystallization buffer: 3 M NH_4SO_4, 100 mM BisTris pH 7.0, 1% PEG400 and 1 mM DTT. A volume of 1 µl of 100 mM BJP-07-017-3 was mixed with 25 µl of mother liquor, and 2 µl of that solution was added to crystal drops. Crystallization wells were resealed, plates were incubated at 20 °C for 2 h and crystals were cryoprotected/collected as usual.

Crystallization: data collection and structure determination. Diffraction data from complex crystals were collected at beamline 24ID-C/E of the Northeastern Collaborative Access Team (NE-CAT) at the Advanced Photon Source at the Argonne National Laboratory. Datasets were integrated and scaled using XDS[52]. Structures were solved by molecular replacement using the program Phaser[53] and the search model PDB entry 1PIN. Iterative manual model building and refinement using Phenix[54] and Coot[55] led to models with excellent statistics.

CITe-Id selectivity assessment. HEK293 cell pellets were lysed with lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 10% v/v glycerol, 0.5% v/v NP-40, protease inhibitors). After lysate clearance centrifugation, protein concentration was determined by bicinchoninic acid (BCA) assay. Samples were precleared with streptavidin resin for 1 h at 4 °C. After centrifugation, the supernatant was split in four and pretreated with DMSO or increasing concentrations of BJP-06-005-3 (100 nM, 500 nM, 1 µM) and incubated at room temperature for 3 h. Then, 2 µM DTB-tagged analog (BJP-DTB) was added to all samples and incubated at 4 °C for 18 h. Urea was added to 8 M final concentration followed by cleanup using Zeba desalting columns. Samples were diluted to 4 M urea with pH 8 lysis buffer and reduced with 10 mM DTT for 30 min at 56 °C. Samples were then alkylated with 20 mM iodoacetamide for 30 min at room temperature. Samples were further diluted to 2 M urea before trypsin digestion at 37 °C overnight. Streptavidin resin was added to the samples and incubated at room temperature for 1 h. The beads were sequentially washed with lysis buffer three times, PBS three times and water two times. Peptides were eluted with four sequential incubations of 50% MeCN and 0.1% trifluoroacetic acid (TFA) for 3 min at room temperature. Eluted peptides were concentrated by vacuum centrifugation and resuspended in 0.1% TFA before batch-mode C18 cleanup, as described previously[56]. Samples were vacuum dried and resuspended in 30% 0.5 M triethylammonium bicarbonate buffer (TEAB), 70% EtOH. iTRAQ stable isotope labeling reagent was added to each corresponding sample. After incubation at room temperature for 1 h, the samples were combined and acidified before vacuum drying. Samples were reconstituted in 0.1% TFA and desalted using an SOLA desalting plate before vacuum drying. Samples were reconstituted in 25% MeCN, 0.1% TFA and batch-mode strong cation exchange (SCX) cleaned, as described previously[56]. Samples were vacuum dried and reconstituted in 5% DMSO, 100 mM ammonium formate for 3D reversed-phase-strong anion exchange-reversed-phase (RP-SAX-RP) liquid chromatography–MS/MS (LC–MS/MS)[57]. Peptides were separated across 11 fractions using a 90-min 5% to 60% MeCN gradient on a Q-Exactive HF mass spectrometer. For data processing, native .RAW data files from the mass spectrometer were processed using the multiplierz Python-based framework[58] to generate .mgf files for input to Mascot (Matrix Science). MS/MS spectra were processed to account for inhibitor fragmentation, as described previously[59]. Peptide precursor masses were recalibrated on a per-scan basis by correcting all m/z values based on accurate mass recorded for the $Si(CH_3)_2O_6$ peak in each spectrum. All data were searched against a forward–reverse human database assembled from the NCBI Refseq database. For de-isotoped higher energy collisional dissociation (HCD) spectra, the precursor mass tolerance was set to 10 ppm and the MS/MS fragment ion tolerance was set to 25 milli-mass units (mmu). Search parameters included trypsin specificity, with a maximum of two missed cleavages; fixed carbamidomethylation of Cys (+57 Da); variable oxidation on Met (+16 Da, with −64-Da neutral loss possible); variable deamidation on Asn and Gln (+1 Da); fixed iTRAQ 4-plex labeling on Lys and N termini (+144 Da); and variable BJP-DTB labeling of Cys (+997 Da, with −997-Da neutral loss possible). Reported peptide sequences were filtered based on a 1% false discovery rate. Normalized reporter ion signal for labeled cysteine residues from multiple peptide-spectrum matches (PSMs) was summed, and a ratio was generated for each reporter channel by comparing it with the DMSO-treated control channel. Inhibitor concentrations and ratios were used to generate a trendline for each labeled site, with the slope being the competitive dose response for the cysteine site.

Yeast experiments. To test complementation of yeast Ess1, strain YSB1026 (derived from strain YXW-3, generously provided by Steve Hanes)[60] was transformed with HIS3-marked plasmids expressing yeast Ess1, human Pin1, the FKBP12^{F36V}-Pin1 fusion or the nonexpressing HIS3 vector. Transformants were grown at 30 °C overnight in synthetic complete medium lacking histidine. After normalizing cell density to ensure that equal numbers of cells were spotted, fourfold serial dilutions were spotted onto the appropriate selective plates. Plates containing 5-fluoro-orotic acid were used to select against pRS426-Pin1, carrying the WT Pin1 on a URA3 plasmid. Plates were incubated for 3 d at 30 °C. Although only one set is shown, four biological replicates gave identical results. To ensure that the FKBP12^{F36V}-Pin1 fusion protein was expressed and intact, whole-cell extracts were made by glass bead lysis. Then, 50 µg protein from each strain was separated by standard SDS–PAGE, blotted to nitrocellulose membrane and probed with antibodies against Pin1. Yeast strains:

YSB1026: MATa, ura3-1, leu2-3,112, trp1-1, his3-11,15, ade2-1, can1-100, ess1Δ:TRP1 (pRS426-Pin1)

Plasmids:

pTPI-Pin1 (Steve Hanes)

pTPI-FKBP^{F36V}-Pin1 (this report)

pRS413-Ess1 (Steve Hanes)

pRS426-Ess1 (this report)

Statistics and reproducibility. Information regarding error bars, numbers of replicates or samples, and statistical analyses are described in the corresponding figure legends. Experiments were not blinded or randomized, and sample sizes were not predetermined using statistical analysis.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

CITe-Id data for BJP-06-005-3 are provided as an excel file, Supplementary Dataset 1. RNA-sequencing data have been deposited to NCBI GEO (accession GSE147340), and the analyzed data file is provided as an excel file, Supplementary Dataset 2. All structural data have been deposited in the PDB (PDB codes 6O33 and 6O34).

References

46. Auld, D. S. et al. Receptor Binding Assays for HTS and Drug Discovery. *Assay Guidance Manual* (eds Sittampalam, G. S. et al.) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004).
47. Yaffe, M. B. et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **278**, 1957–1960 (1997).
48. Zhang, Z. & Marshall, A. G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225–233 (1998).
49. Ficarro, S. B., Alexander, W. M. & Marto, J. A. mzStudio: a dynamic digital canvas for user-driven interrogation of mass spectrometry data. *Proteomes* **5**, 20 (2017).
50. Brinkman, E. K., Chen, T., Amendola, M. & Steensel, B. V. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
51. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
52. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* **66**, 133–144 (2010).
53. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **66**, 213–221 (2010).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
56. Adelmant, G. O. et al. in *Sample Preparation in Biological Mass Spectrometry* (eds Lazarev, A. V. & Ivanov, A. R.) Ch. 22 (Springer, 2015).
57. Zhou, F. et al. Genome-scale proteome quantification by DEEP SEQ mass spectrometry. *Nat. Commun.* **4**, 2171 (2013).
58. Parikh, J. R. et al. Multiplierz: an extensible API based desktop environment for proteomics data analysis. *BMC Bioinformatics* **10**, 364 (2009).
59. Ficarro, S. B. et al. Leveraging gas-phase fragmentation pathways for improved identification and selective detection of targets modified by covalent probes. *Anal. Chem.* **88**, 12248–12254 (2016).
60. Wu, X. et al. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.* **19**, 3727–3738 (2000).

Acknowledgements

We thank all members of the Gray laboratory for helpful discussions, M. Kostic for providing feedback on the manuscript and K. Westover for discussion of k_{inact} and K_i calculations. This work was supported by the Ruth L. Kirschstein NRSA Individual Predoctoral Fellowship (grant no. F31 CA225066, B.J.P.), the Training Grant in Pharmacological Sciences (NIH grant no. T32 GM007306, B.J.P./Z.M.D), the Training Grant in Chemical Biology (NIH grant no. 5 T32 GM095450-04, B.J.P.), the Chleck Foundation (B.J.P./Z.M.D.), the American Cancer Society Postdoctoral Fellowship (grant no. PF-17-010-01-CDD, B.N.), the Claudia Adams Barr Program in Innovative Basic Cancer Research Award (B.N.), the Katherine L. and Steven C. Pinard Research Fund (N.S.G./B.N.), the Israel Science Foundation (grant no. 2462/19, N.L.), the Rising Tide Foundation (N.L.), the Israel Cancer Research Fund (N.L.), the Israeli Ministry of Science and Technology (grant no. 3-14763, N.L.), the Moross Integrated Cancer Center (N.L.), the Helen and Martin Kimmel Center for Molecular Design (N.L.), the Joel and Mady Dukler Fund for Cancer Research (N.L.), the Estate of Emile Mimran, Virgin JustGiving, the George Schwartzman Fund (N.L.), the National Institutes of Health R01s (grant no. R01 GM056663, S.B.), (grant no. R01 CA219850, grant no. R01 CA233800, J.A.M.), (grant no. R01 CA205153, N.S.G., K.P.L. and X.Z.Z.) and the Hale Center for Pancreatic Research (N.S.G.). BJP-02-118-2 and BJP-07-017-3 cocrystal structures were determined thanks to research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines (grant no. NIGMS P41 GM103403).

Author contributions

B.J.P. performed the chemical synthesis and biological characterization and wrote the manuscript. Z.M.D. generated cell lines and performed the GFP dropout experiment and RNA-sequencing analysis. B.N. generated cell lines and designed and supervised biological studies. C.M.B., S.B.F. and J.A.M. performed CITe-Id and mass spectrometry studies. M.L.M. assisted in Pin1 mutant cellular studies. S. Kozono, S. Kibe and X.L. performed the FP assay. D.Z., D.D. and N.L. provided computational docking support. L.T. advised on compound design and T.D.M. provided support for the chemical synthesis and biological characterization. Z.C.Y., N.E.V., E.A.G. and H.-S.S. performed the protein expression and crystallography; H.-S.S. and S.D.-P. analyzed the crystallography data. Y.C. and S.B. performed the yeast complementation experiments and analysis. K.P.L. and X.Z.Z. helped supervise the project and provided experimental support. N.S.G. conceived of and led this study.

Competing interests

N.S.G. is a Scientific Founder and member of the Scientific Advisory Board (SAB) of C4 Therapeutics, Syros, Soltego, Gatekeeper and Petra Pharmaceuticals, and has received research funding from Novartis, Astellas, Taiho and Deerfield. J.A.M. serves on the SAB of 908 Devices, Boston, MA, and receives sponsored research support from Vertex and AstraZeneca. N.L. is a member of the SAB of Trilogy Sciences and has received research support from Teva. B.J.P., N.S.G., H.-S.S., S.D.-P., Z.M.D., C.M.B., J.A.M., S. Kozono, X.L., K.P.L. and X.Z.Z. are inventors on a patent application related to the Pin1 inhibitors described in this manuscript (WO/2019/241496). B.N. is an inventor on patent applications related to the dTAG system described in this manuscript (WO/2017/024318, WO/2017/024319, WO/2018/148443, WO/2018/148440). K.P.L. and X.Z.Z. are inventors on many issued patents or pending patent applications related to Pin1 inhibitors and/or biomarkers.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-020-0550-9.

Correspondence and requests for materials should be addressed to K.P.L. or N.S.G.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Nathanael S. Gray & Kun Ping Lu

Last updated by author(s): 3/24/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	RNA sequencing data analysis was performed in R version 3.5.1 using Bioconductor version 3.8

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

CITe-Id data for BJP-06-005-3 is provided as an excel file, Supplementary Dataset 1. RNA-sequencing data has been deposited to the NCBI GEO (accession GSE147340), and the analyzed data file is provided as an excel file, Supplementary Dataset 2. All structural data has been deposited in the PDB (PDB codes 6O33, 6O34).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not predetermined using statistical analyses.
Data exclusions	Data was not excluded from analysis.
Replication	All biological experiments were successfully reproduced in at least two independent experiments. Information on the number of replicates and independent experiments that were performed for each measurement are disclosed in the manuscript.
Randomization	Experiments were not blinded or randomized.
Blinding	Experiments were not blinded or randomized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Details regarding all antibodies employed in this study can be found in the Methods, "Immunoblotting" section. Antibodies used against various proteins were as follows: Pin1 (1:1,000, Cell Signaling cat #3722), alpha-Tubulin (1:1000, Cell Signaling cat #3873), phoshpo-cdc2 (Tyr15) (1:1000, Cell Signaling cat #4539), CDK1 (1:1000, Abcam cat #ab131450), HA-Tag (1:1000, Cell Signaling cat #3724), ß-catenin (1:1000, Cell Signaling cat #8480), NFκB p65 (1:1000, Cell Signaling cat #6956), c-Myc (1:1000, Cell Signaling cat #5605), GAPDH (1:1000, Cell Signaling cat #2118), ß-Actin (1:1000, Cell Signaling cat #3700), histone H3 (1:1000, Cell Signaling cat #9715).
Validation	Antibodies were validated by the manufacturer according to standard protocols denoted on the manufacturer website.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Details regarding all cell lines employed in this study can be found in the Methods, "Cell Culture" section. Cell lines employed in this study: PATU-8988T (DSMZ), PATU-8902 (DSMZ), HeLa (ATCC), MDA-MB-468 (ATCC), 293FT (Thermo Fisher Scientific), Phoenix-AMPHO (ATCC), HEK293 (ATCC), BxPC-3 (ATCC), PC-3 (ATCC), NB4 (ATCC), and Kuramochi (Panagiotis A. Konstantinopoulos's laboratory).
Authentication	With the exception of Kuramochi cells, cell lines were originally purchased directly from the vendor. Authentication was performed by the vendor, which includes STR profiling. Additional authentication was not performed.
Mycoplasma contamination	All cell lines tested negative for mycoplasma.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were plated one day before treatment, and treated as indicated in the respective figures. Cells were harvested using trypsin, then washed 3x with cold PBS. Cell pellets were then fixed with 1 mL cold 80% ethanol in PBS at 4°C overnight. The cells were then washed by adding 1 mL of PBS and then centrifuging (1500 rpm for 5 minutes), followed by aspiration of the PBS/ ethanol mix. Cells were resuspended in cell cycle staining solution (100 ug/mL of RNAse A (Roche, 10109169001) and 50 ug/mL propidium iodide (Life Technologies, P1304MP), diluted in PBS), and incubated for 10 minutes at room temperature.
Instrument	Data was acquired by flow cytometry on a Guava easyCyte HT flow cytometer (Millipore).
Software	Data collection was facilitated using the InCyte software. Data was analyzed using FlowJo software (version 9.3.2).
Cell population abundance	Live singlet cells typically comprised 70-90% of the total cell population data was collected for.
Gating strategy	Live singlet cells were selected by gating on Forward Scatter vs Side Scatter. Cell cycle phase was assigned by assessing the histogram of propidium iodide (PI) content, with the first major peak as the G1 phase, the second peak as the G2/M phase, and the area between but distinct from the two peaks as the S phase.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

nature chemical biology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41589-020-0550-9

In the format provided by the authors and unedited.

Identification of a potent and selective covalent Pin1 inhibitor

Benika J. Pinch[1,2,3], Zainab M. Doctor[2,3,10], Behnam Nabet[2,3,10], Christopher M. Browne[2,3,9], Hyuk-Soo Seo[2,3], Mikaela L. Mohardt[2], Shingo Kozono[4], Xiaolan Lian[4], Theresa D. Manz[2,3,5], Yujin Chun[3], Shin Kibe[4], Daniel Zaidman[6], Dina Daitchman[6], Zoe C. Yeoh[2,3], Nicholas E. Vangos[2,3], Ezekiel A. Geffken[2,3], Li Tan[2], Scott B. Ficarro[2,7], Nir London[6], Jarrod A. Marto[7,8], Stephen Buratowski[3], Sirano Dhe-Paganon[2,3], Xiao Zhen Zhou[4], Kun Ping Lu[4]✉ and Nathanael S. Gray[2,3]✉

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [2]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. [4]Division of Translational Therapeutics, Department of Medicine and Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [5]Department of Pharmacological and Medicinal Chemistry, Saarland University, Saarbruecken, Germany. [6]Department of Organic Chemistry, Weizmann Institute of Science, Rehovot, Israel. [7]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA, USA. [8]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [9]Present address: Discovery Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Boston, MA, USA. [10]These authors contributed equally: Zainab M. Doctor, Behnam Nabet.
✉e-mail: klu@bidmc.harvard.edu; Nathanael_Gray@dfci.harvard.edu

SUPPLEMENTARY TABLES

Supplementary Table 1 | Crystallization conditions and statistics. Crystal structures were deposited in the Protein Data Bank (PDB codes 6O33 and 6O34).

Table 1 Crystallization conditions and data collection and refinement statistics for crystal structures

Structure Name	BJP-02-118-2	BJP-07-017-3
Ligand	Peptide	Peptide
RCSB accession code	**6O33**	**6O34**
Data collection [a]		
Space group	P 4_3 2_1 2	P 4_3 2_1 2
Cell dimensions		
a , *b* , *c* (Å)	49.01 49.01 137.18	48.99 48.99 137.84
a, b, g (°)	90 90 90	90 90 90
Resolution (Å)	33.6 - 1.74 (1.802 - 1.74) [b]	39.93 - 1.57 (1.626 - 1.57)
R_{pim}	0.045 (0.729)	0.044 (1.354)
I / sI	10.31 (0.99)	11.05 (1.83)
Completeness (%)	99.96 (100.00)	99.89 (99.15)
Redundancy	13.1 (13.8)	42.1 (24.7)
Structure solution		
PDB entries used for molecular replacement	1PIN	1PIN
Refinement		
Resolution (Å)	33.6 - 1.74	39.87 - 1.57
No. reflections, unique	17968 (1743)	24249 (2347)
R_{work} / R_{free}	0.1979/0.2379	0.2030/0.2529
No. atoms	1366	1346
Protein	1245	1217
Ligand/ion	5	36
Water	116	93
B-factors		
Protein	39.4	39.1
Ligand/ion	41.7	47.5
Water	44.6	43.8
R.m.s. deviations		
Bond lengths (Å)	0.005	0.017
Bond angles (°)	1.04	1.26
Ramachandran Plot [e]		
Preferred	98.6%	98.6%
Allowed	1.4%	0.7%
Not Allowed	0.0%	0.7%

[a] A single crystal was used to collect data for each of the structures reported here.

[b] Values in parentheses are for highest-resolution shell.

[c] Where multiple chains exist for the protein/peptide complexes, the atom numbers were listed for the individual polypeptide chains separated by "/", where "pep" stands for peptide.

[d] Where multiple chains exist for the protein/peptide complexes, the *B*-factors were listed for the individual polypeptide chains separated by "/", where "pep" stands for peptide.

[e] Numbers of residues were given with percentage given in parentheses.

Supplementary Table 2 | Sequences of primers annealed to form guides used in CRISPR/Cas9 GFP dropout assay in Cas9-expressing PATU-8988T cells.

CACCAGTCACGGCGGCCCTCGTCC	Pin1_gRNA-1_f
AAACGGACGAGGGCCGCCGTGACT	Pin1_gRNA-1_r
CACCAGGACGAGGGCCGCCGTGAC	Pin1_gRNA-2_f
AAAC GTCACGGCGGCCCTCGTCCT	Pin1_gRNA-2_r
CACCCAGTGGTGGCAAAAACGGGC	Pin1_gRNA-3_f
AAACGCCCGTTTTTGCCACCACTG	Pin1_gRNA-3_r
CACCGCATCACTAACGCCAGCCAGT	Pin1_gRNA-4_f
AAACACTGGCTGGCGTTAGTGATGC	Pin1_gRNA-4_r
CACCGTGAAGCACAGCCAGTCACGG	Pin1_gRNA-5_f
AAACCCGTGACTGGCTGTGCTTCAC	Pin1_gRNA-5_r
CACCGGAAGATCACCCGGACCAAGG	Pin1_gRNA-6_f
AAACCCTTGGTCCGGGTGATCTTCC	Pin1_gRNA-6_r
CACCGGATCTTCTGGATGTAGCCTG	Pin1_gRNA-7_f
AAACCAGGCTACATCCAGAAGATCC	Pin1_gRNA-7_r
CACCGCCCCGTCCGCAGCGCAAACG	Pin1_gRNA-8_f
AAACCGTTTGCGCTGCGGACGGGGC	Pin1_gRNA-8_r

Supplementary Table 3 | Sequences of sgRNAs used in the CRISPR-Cas9 GFP dropout assay in Cas9-expressing PATU-8988T cells.

Name	Target Strand	Sequence of gRNA	PAM
gRNA-1	sense	AGTCACGGCGGCCCTCGTCC	TGG
gRNA-2	antisense	AGGACGAGGGCCGCCGTGAC	TGG
gRNA-3	sense	CAGTGGTGGCAAAAACGGGC	AGG
gRNA-4	sense	CATCACTAACGCCAGCCAGT	GGG
gRNA-5	sense	TGAAGCACAGCCAGTCACGG	CGG
gRNA-6	sense	GAAGATCACCCGGACCAAGG	AGG
gRNA-7	antisense	GATCTTCTGGATGTAGCCTG	TGG
gRNA-8	antisense	CCCCGTCCGCAGCGCAAACG	AGG

Supplementary Table 4 | Statistics for CRISPR/Cas9 GFP dropout assay in Cas9-expressing PATU-8988T cells.

Summary

	LUC	RPS19	Pin1-1	Pin1-2	Pin1-3	Pin1-4	Pin1-5	Pin1-6	Pin1-7	Pin1-8
d7	ns	****	ns	ns	ns	ns	ns	ns	ns	ns
d11	ns	****	ns	ns	ns	ns	ns	*	ns	ns
d15	ns	****	**	ns	**	ns	ns	***	ns	**
d19	ns	****	**	**	***	ns	ns	**	ns	****
d24	ns	****	**	*	***	ns	ns	**	ns	****
d28	ns	****	***	***	****	ns	*	***	ns	****
d32	ns	****	****	****	****	*	***	****	**	****

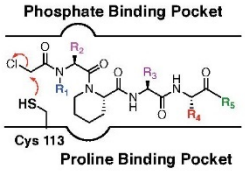
Adjusted p-value

	LUC	RPS19	Pin1-1	Pin1-2	Pin1-3	Pin1-4	Pin1-5	Pin1-6	Pin1-7	Pin1-8
d7	0.8798	0.0001	0.5522	0.9994	0.5915	0.9995	0.9996	0.3394	0.9999	0.928
d11	0.6074	0.0001	0.0529	0.8574	0.4304	0.9998	0.986	0.0373	0.9999	0.7874
d15	0.7031	0.0001	0.0018	0.1076	0.0056	0.5466	0.2493	0.0003	0.0659	0.0017
d19	0.5693	0.0001	0.0018	0.0048	0.0006	0.5139	0.1143	0.0095	0.8118	0.0001
d24	0.6809	0.0001	0.0027	0.0342	0.0007	0.6962	0.2161	0.0019	0.2459	0.0001
d28	0.378	0.0001	0.0009	0.0006	0.0001	0.1383	0.0266	0.0002	0.0899	0.0001
d32	0.9774	0.0001	0.0001	0.0001	0.0001	0.0114	0.0008	0.0001	0.002	0.0001

Two-way ANOVA analysis (significance, top and p-value, below) for normalized proportion of GFP cells over time, comparing each value to d3 for each guide. Initial measurement made on d3, three days after infection. Adjusted p-values derived from Dunnett correction of two-way ANOVA, calculated from n = 3 biologically independent samples, representative of n = 2 independent experiments. ns = not significant.

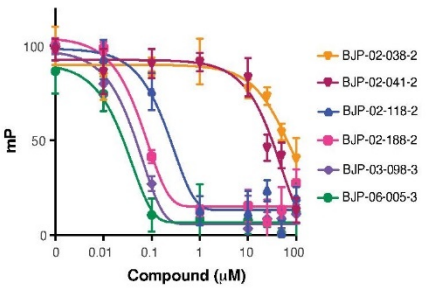
SUPPLEMENTARY FIGURES

a

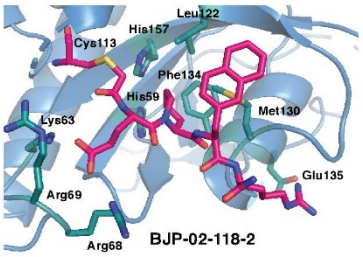


Compound	R_1	R_2	R_3	R_4	R_5	K_i (nM)	cLogP
BJP-02-038-2 (**2**)	H	Asp	Nal	Gln	NH_2	-	-0.08
BJP-02-041-2 (**3**)	H	Glu	Nal	Gln	NH_2	2000	-0.65
BJP-02-118-2 (**4**)	H	Glu	Nal	Arg	NH_2	170	-2.78
BJP-02-188-2 (**5**)	H	Glu	Nal	Cit	NH_2	70	-0.31
BJP-03-098-3 (**6**)	H	Glu	Trp	Cit	NH_2	30	-1.49
BJP-06-005-3 (**1**)	Me	Phe	Trp	Cit	OCH_2CH_3	15	2.89

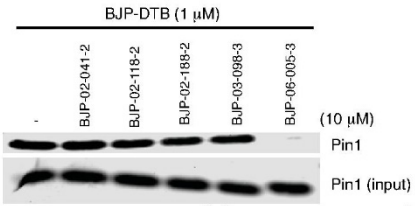
b



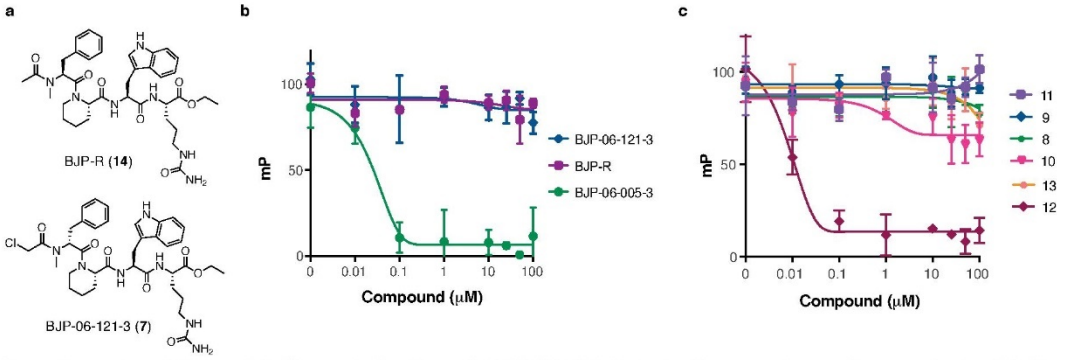
c



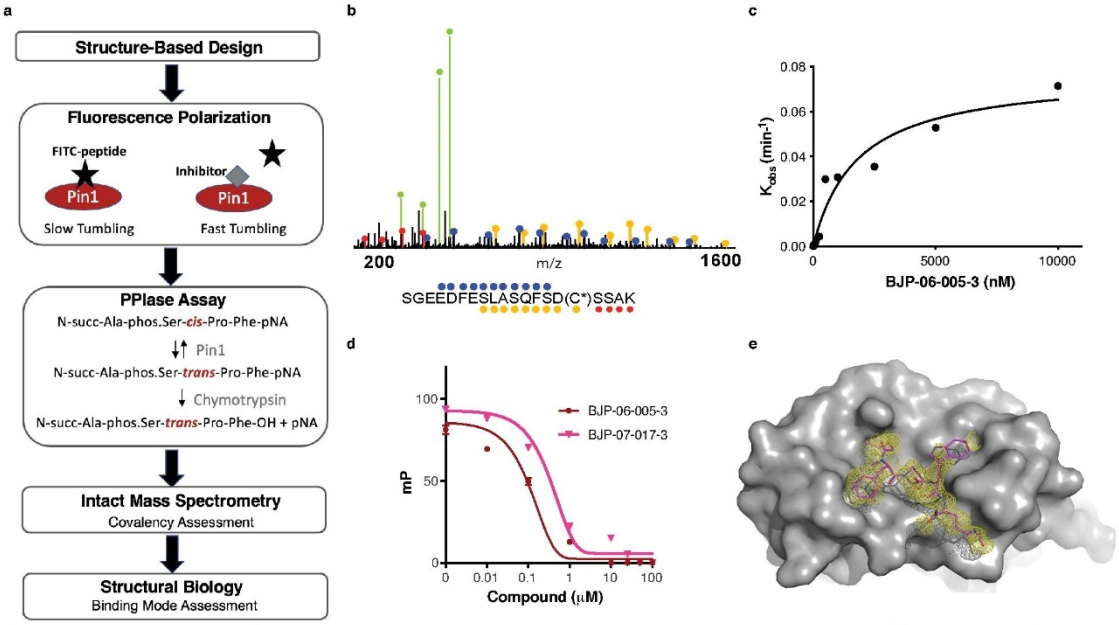
d



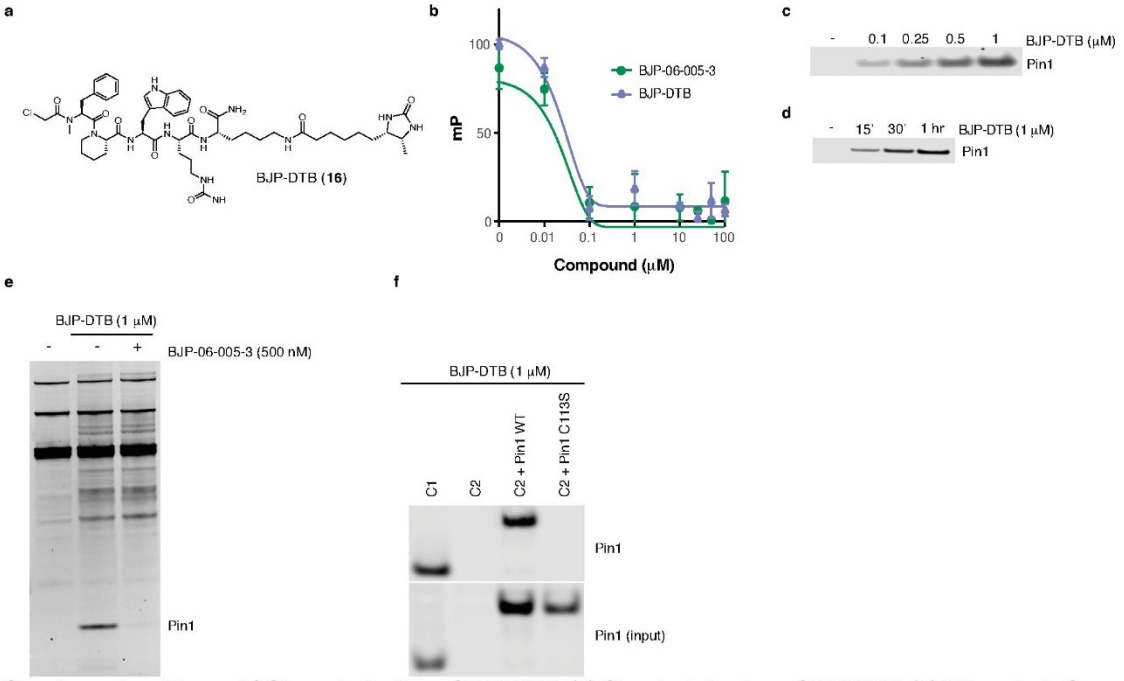
Supplementary Figure 1 | Characterization of D-PEPTIDE derivatives. (**a**) Schematic representation of Pin1 PPIase active site and table of synthesized derivatives, their apparent K_i values (FP), and cLogP values (ChemDraw); Cit = L-Citrulline; Nal = L-Naphthylalanine; Me = methyl; (**b**) FP readout for indicated compounds following a 12 h incubation with Pin1. Data points are plotted as the average of n = 3 independent samples ± SEM, and are representative of n = 2 independent experiments; (**c**) Co-crystal structure of BJP-02-118-2 (magenta) bound to Pin1 (blue) with key binding residues highlighted in green (PDB 6O33); (**d**) PATU-8988T cells were treated with 10 μM of the indicated compounds for 5 h, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down, and immunoblot analysis. Uncropped immunoblots are displayed in Supplementary Figure 13, and are representative of n = 2 independent experiments.



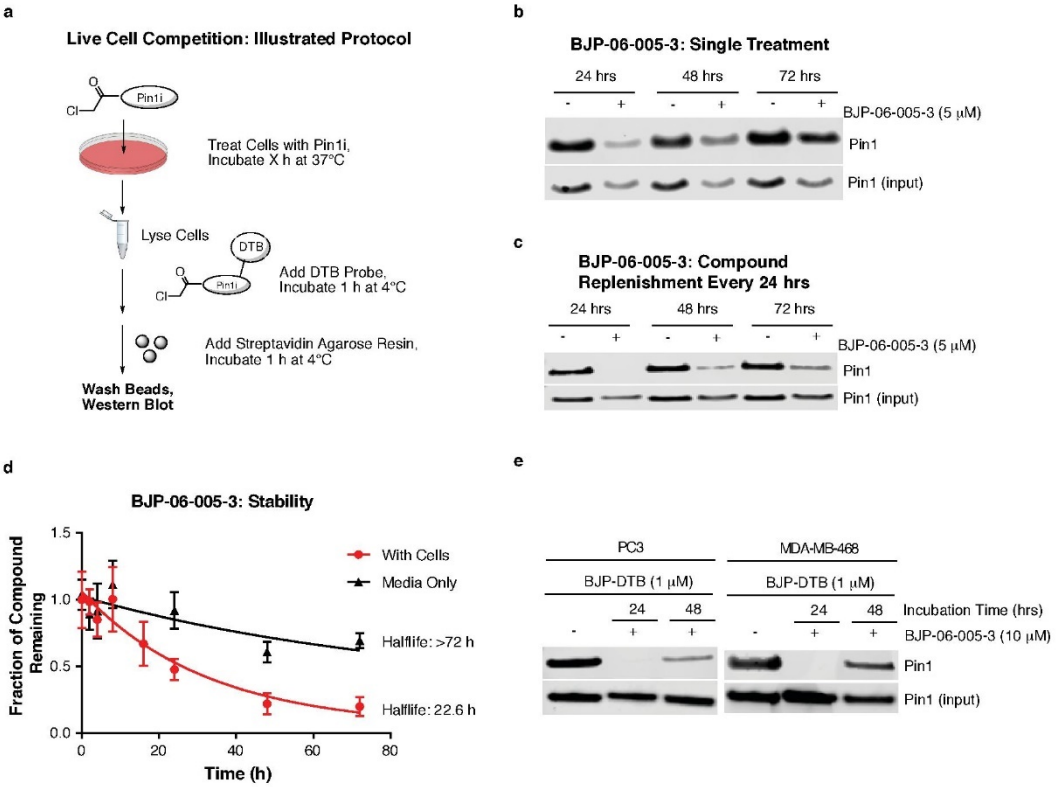
Supplementary Figure 2 | Characterization of BJP-06-005-3 derivatives and negative controls. (**a**) Chemical structures of BJP-06-005-3's negative controls, including the reversible control, BJP-R, and the inactive stereoisomer, BJP-06-121-3; (**b-c**) FP readout after a 12 h incubation with Pin1. Data points are plotted as the average of n = 3 independent samples ± SEM, and are representative of n = 2 independent experiments.



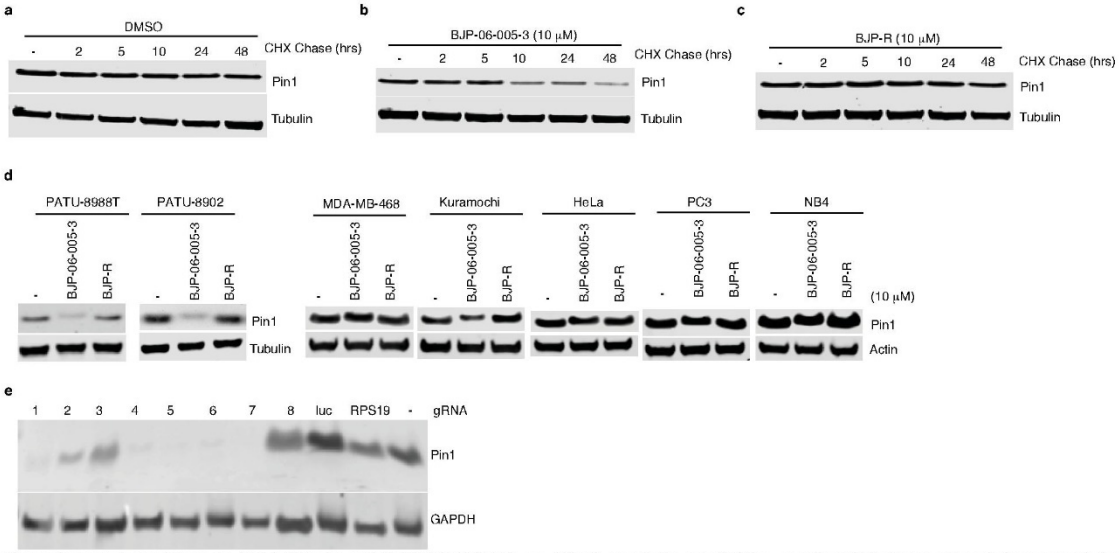
Supplementary Figure 3 | Biochemical and structural characterization of BJP-06-005-3. (**a**) Flow-chart of assays used to guide compound optimization and characterization. FITC = fluorescein isothiocyanate, succ = succinyl, pNA = para-nitroaniline, phos. = phosphorylated; (**b**) MS/MS spectra of BJP-06-005-3-modified Pin1 peptide containing Cys 113. Red: y-ions, yellow: y-ions which undergo loss of internal BJP-06-005-3 fragment, blue: b-ions, green: BJP-06-005-3 specific fragment ions; (**c**) FP readout following a 0, 5, 10, 15, 20, 25, or 30 min incubation of BJP-06-005-3 (in 9-point dose) and recombinant full-length Pin1. Plot of observed rate constants for inhibition vs BJP-06-005-3 concentration; (**d**) FP readout after a 14 h incubation with Pin1. Data points are plotted as the average of n = 3 independent samples ± SEM, and are representative of n = 2 independent experiments; (**e**) Omit map showing electron density for the co-crystal structure of Pin1 (gray) bound to BJP-07-017-3 (magenta) (PDB 6O34).



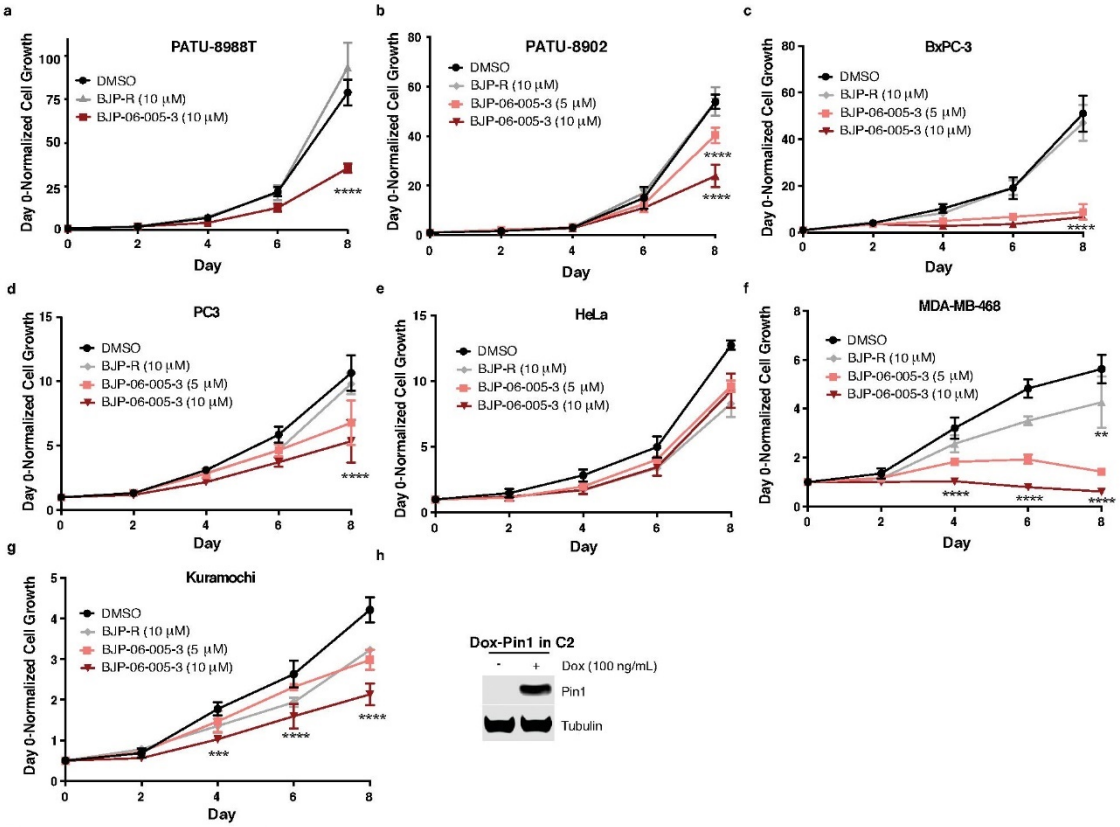
Supplementary Figure 4 | Characterization of BJP-DTB. (**a**) Chemical structure of BJP-DTB; (**b**) FP readout after a 12 h incubation with Pin1. Data points are plotted as the average of n = 3 independent samples ± SEM, and are representative of n = 2 independent experiments; (**c**) Pull-down of Pin1 from PATU-8988T lysate with the indicated concentrations of BJP-DTB after a 1 h incubation, streptavidin pull-down, and immunoblot analysis; (**d**) Pull-down of Pin1 from PATU-8988T lysate with 1 μM of BJP-DTB following incubation at the indicated time points; (**e**) PATU-8988T cell lysates were pre-treated with DMSO or BJP-06-005-3 (500 nM) for 1 h at 4°C, followed by incubation with BJP-DTB (1 μM) for 1 h at 4°C. The lysates were then run on a gel and visualized using a streptavidin secondary antibody (IRDye 800CW; cat #926-32230); (**f**) Pull-down of Pin1 from PATU-8988T lysates (C1, C2, or C2 transduced with Flag-Pin1 (WT or C113S)) with BJP-DTB (1 μM). Uncropped immunoblots for **c**, **d**, and **e** are displayed in Supplementary Figure 13, and are representative of n = 2 independent experiments. Uncropped immunoblots for **f** are displayed in Supplementary Figure 14, and are representative of n = 2 independent experiments.



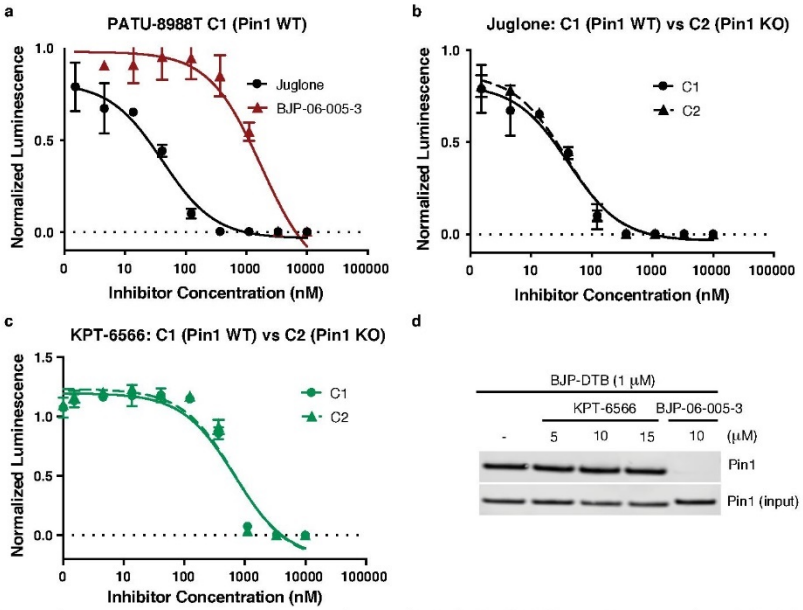
Supplementary Figure 5 | Stability and cellular target engagement of BJP-06-005-3. (**a**) Schematic representation of a live cell competition pull-down experiment; (**b**) PATU-8988T cells were treated with 5 μM BJP-06-005-3 for the indicated times, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down, and immunoblot analysis; (**c**) PATU-8988T cells were treated with 5 μM BJP-06-005-3, and every 24 hrs the media was aspirated and replaced with fresh media/BJP-06-005-3. At the indicated time points, the cells were lysed, incubated with BJP-DTB, then subject to streptavidin pull-down, and immunoblot analysis; (**d**) BJP-06-005-3 is stable in media (half-life > 72 hours) but is unstable in media containing cells (half-life 22.6 hours). PATU-8988T cells were treated with BJP-06-005-3 (10 μM) and the media off of the cells (red curve) or pure media (black curve) was analyzed by MS/MS. Data points are plotted as the average of n = 3 biologically independent samples ± SEM; (**e**) PC3 or MDA-MB-468 cells were treated with 10 μM of BJP-06-005-3 for the indicated times, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down, and immunoblot analysis. Uncropped immunoblots for **b**, **c**, and **e** are displayed in Supplementary Figure 14, and are representative of n = 2 independent experiments.



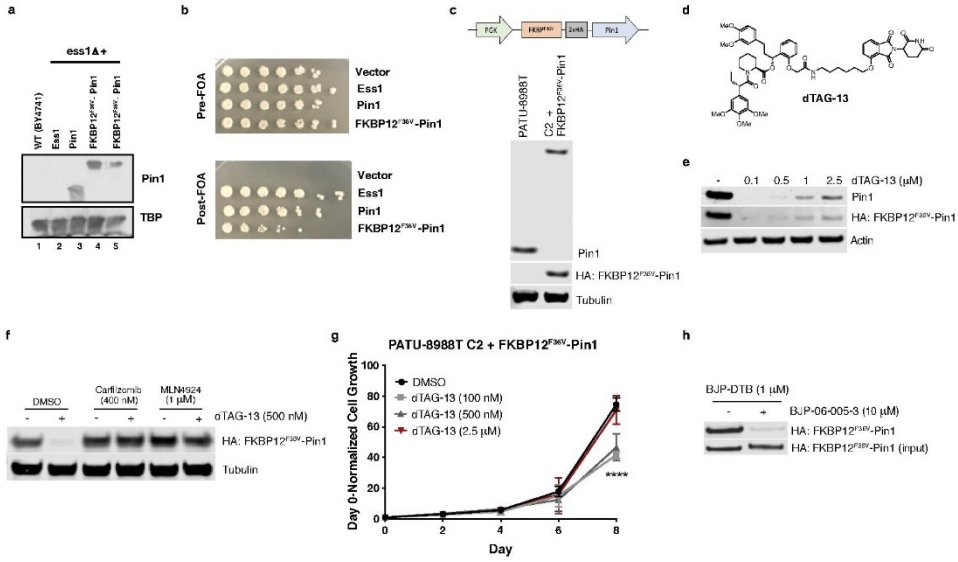
Supplementary Figure 6 | Effects of BJP-06-005-3 on Pin1 protein stability, and validation of guides used in CRISPR/Cas9 GFP dropout assay. PATU-8988T cells treated with (**a**) DMSO, (**b**) BJP-06-005-3 (10 μM), or (**c**) BJP-R (10 μM) for 24 h, followed by cycloheximide (CHX) chase for the indicated time points; (**d**) Immunoblot analysis of the indicated cell lines following a 48 h treatment with BJP-06-005-3 (10 μM) or BJP-R (10 μM), with compound/media replenishment after 24 h; (**e**) Validation of sgRNAs used in the CRISPR-Cas9 GFP dropout experiment in Cas9-expressing PATU-8988T cells. Cells were treated for 48 h with 1-8 sgRNAs (Supplementary Table 3) targeting Pin1, or guides targeting Luciferase (5'-CCCGGCGCCATTCTATCCGC-3') or RPS19 (5'-GTAGAACCAGTTCTCATCGT-3') as controls, and then Pin1 protein levels were assessed by immunoblot analysis. Uncropped immunoblots for **a-e** are displayed in Supplementary Figure 15, and are representative of n = 2 independent experiments.



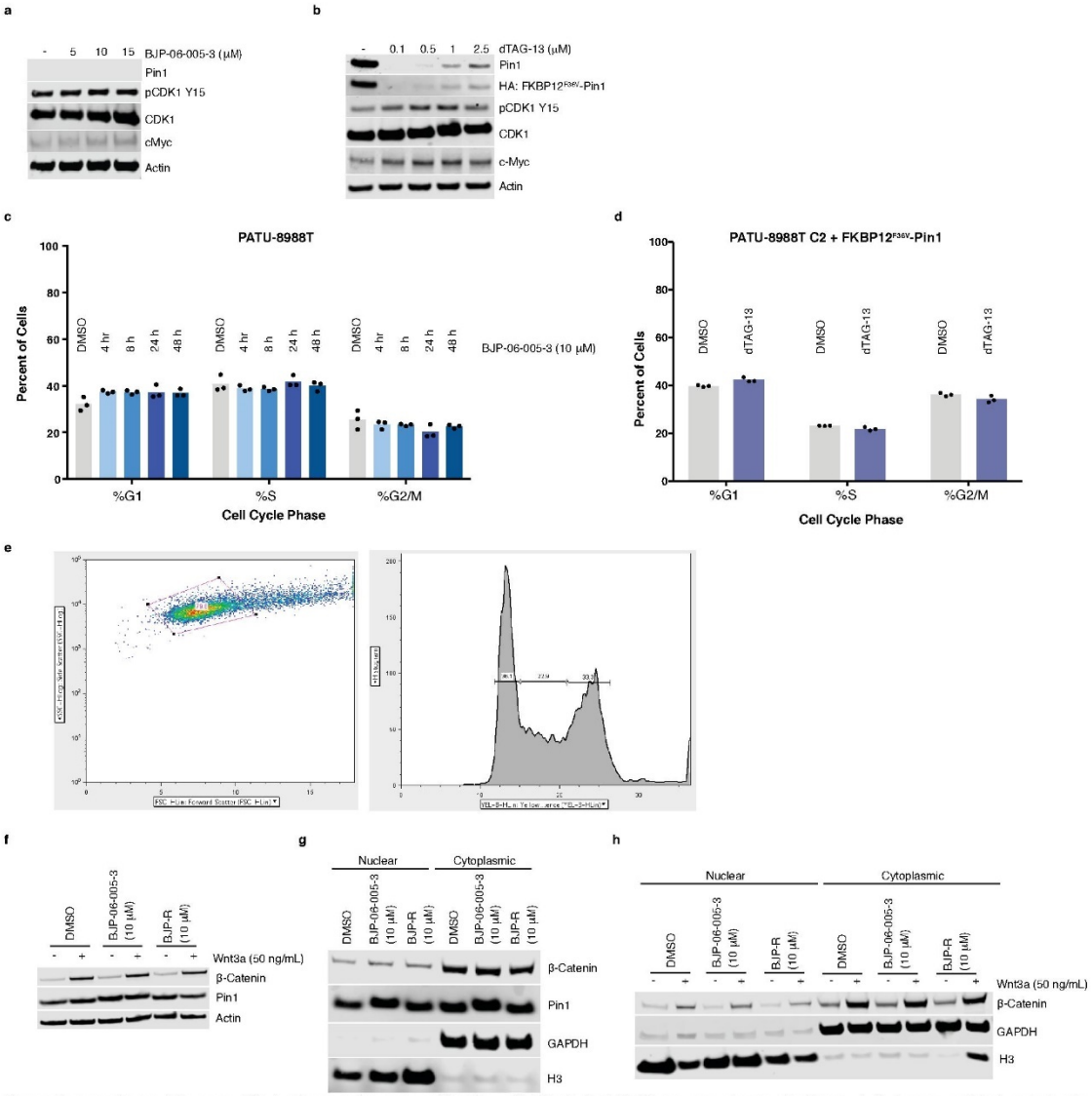
Supplementary Figure 7 | BJP-06-005-3 causes Pin1-mediated anti-proliferative effects across a panel of cancer cell lines. Growth curves of (**a**) PATU-8988T, (**b**) PATU-8902, (**c**) BxPC-3, (**d**) PC3, (**e**) HeLa, (**f**) MDA-MB-468, or (**g**) Kuramochi cells treated with DMSO or the indicated concentrations of BJP-06-005-3 or BJP-R. Readings were taken every 2 days. Data points are plotted as the average of n = 3 biologically independent samples ± SEM, and are representative of n = 2 independent experiments. **** adj. $p < 0.0001$, *** adj. $p = 0.0009$, ** adj. $p = 0.0062$ by Bonferroni-corrected two-way ANOVA; (**h**) Immunoblot analysis of C2 transduced with Dox-inducible Pin1, with or without Dox (100 ng/mL) treatment for 48 h. Uncropped immunoblots are displayed in Supplementary Figure 16, and are representative of n = 3 independent experiments.



Supplementary Figure 8 | Comparison of BJP-06-005-3 to other covalent Pin1 inhibitors. (**a**) C1 was treated with Juglone or BJP-06-005-3 for 4 days, with media/BJP-06-005-3 replaced every 24 h to ensure continual engagement of Pin1. Viability was assessed by CellTiter-Glo. Data points are plotted as the average of n = 3 biologically independent samples ± SEM, and are representative of n = 3 independent experiments; (**b**) C1 and C2 were treated with Juglone for 4 days. Viability was assessed by CellTiter-Glo. Data points are plotted as the average of n = 3 biologically independent samples ± SEM, and are representative of n = 3 independent experiments; (**c**) C1 and C2 were treated with KPT-6566 for 3 days. Viability was assessed by CellTiter-Glo. Data points are plotted as the average of n = 3 biologically independent samples ± SEM, and are representative of n = 2 independent experiments; (**d**) PATU-8988T cells treated in a competition format with the indicated concentrations of BJP-06-005-3 or KPT-6566 for 5 h, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down, and immunoblot analysis. Uncropped immunoblots are displayed in Supplementary Figure 16, and are representative of n = 2 independent experiments.

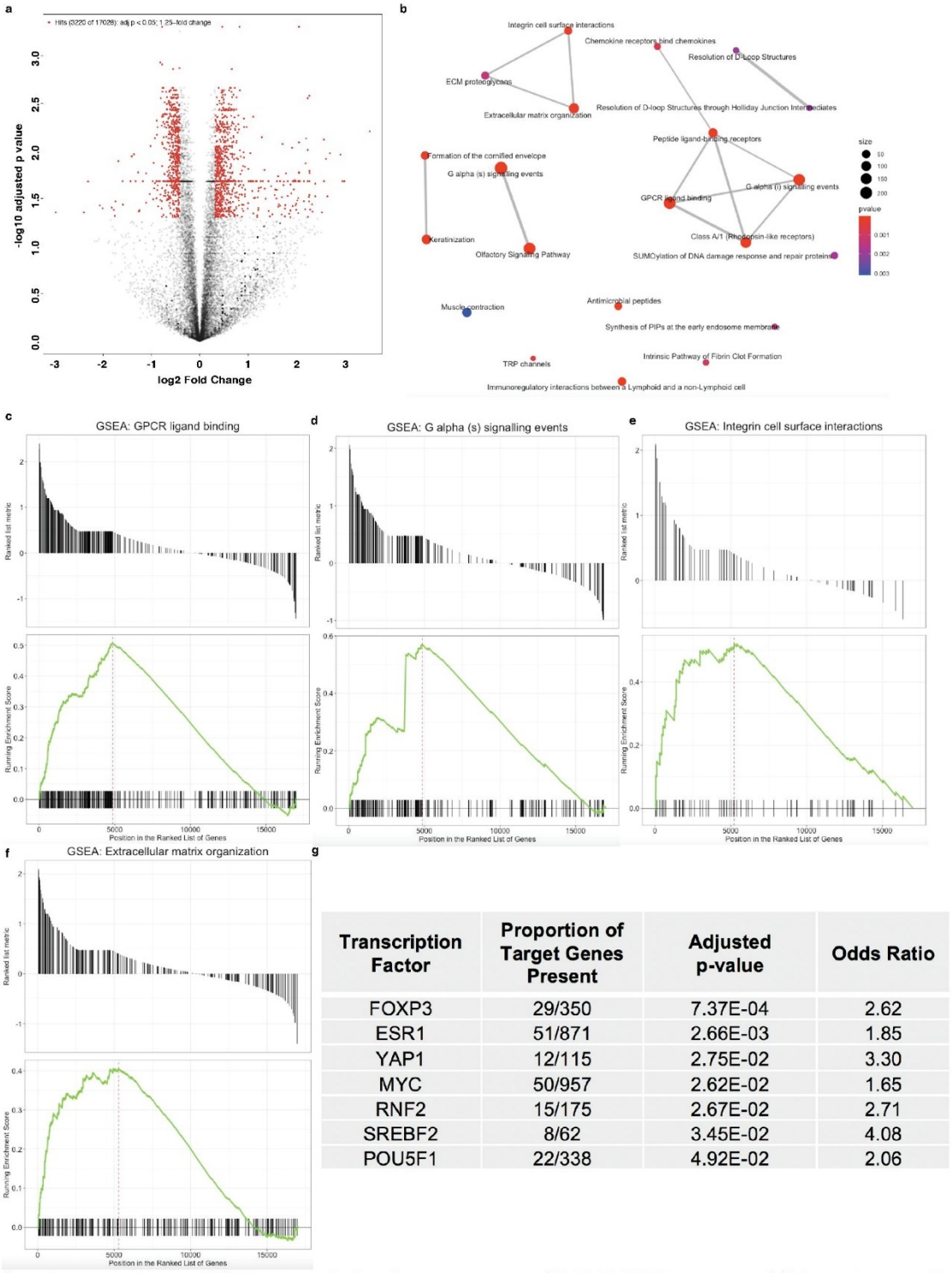


Supplementary Figure 9 | Characterization of FKBP12^{F36V}-Pin1. (**a**) Expression of Pin1 and FKBP^{F36V}-Pin1 in yeast by immunoblot analysis. Lane 1: yeast strain BY4741 with wild-type *ESS1* in the chromosome. Lanes 2-5: derivatives of YXW-3^{i}, in which the chromosomal *ESS1* deletion is complemented by plasmids expressing yeast Ess1 (lane 2), human Pin1 (lane 3), or two isolates of the FKBP12^{F36V}-Pin1 fusion (lanes 4-5). TBP = TATA-Binding Protein; (**b**) Pin1 and FKBP12^{F36V}-Pin1 complement an *ESS1* deletion. Yeast strain YSB1026 (based on YXW-3, but with covering plasmid pRS426-Pin1) was transformed with *HIS3*-marked plasmids expressing vector, Ess1, Pin1, or FKBP12^{F36V}-Pin1. Cell growth was tested before (above) and after (below) selection on media containing 5-fluoro-orotic acid (FOA), which selects against the *URA3* marker on the pRS426-Pin1 plasmid. Cells were spotted onto plates in 3-fold dilutions, from left to right; (**c**) Schematic depiction of the lentiviral expression strategy (above), immunoblot analysis of PATU-8988T parental cells versus C2 + FKBP12^{F36V}-Pin1 (below); (**d**) Chemical structure of dTAG-13; (**e**) Immunoblot analysis of C2 + FKBP12^{F36V}-Pin1 cells treated with the indicated concentrations of dTAG-13 for 5 h; (**f**) Immunoblot analysis of C2 + FKBP12^{F36V}-Pin1 cells pre-treated for 2 h with DMSO, carfilzomib (400 nM) or MLN4924 (1 μM), followed by a 5 h co-treatment with dTAG-13 (500 nM); (**g**) Day 0-normalized growth rate of C2 + FKBP12^{F36V}-Pin1 cells treated with the indicated concentrations of dTAG-13, with readings taken every 2 days and compound/media replaced every 2 days. Data points are plotted as the average of n = 3 biologically independent samples ± SEM, and are representative of n = 3 independent experiments. **** adj. $p < 0.0001$ by Bonferroni-corrected two-way ANOVA; (**h**) C2 + FKBP12^{F36V}-Pin1 cells treated with DMSO or BJP-06-005-3 (10 μM) for 5 h, followed by cell lysis, incubation with BJP-DTB, streptavidin pull-down, and immunoblot analysis. Uncropped immunoblots for **a** are displayed in Supplementary Figure 16, and are representative of n = 1 independent experiments. Uncropped immunoblots for **c** are displayed in Supplementary Figure 16, and are representative of n = 3 independent experiments. Uncropped immunoblots for **e** are displayed in Supplementary Figure 17, and are representative of n = 3 independent experiments. Uncropped immunoblots for **f** and **h** are displayed in Supplementary Figure 17, and are representative of n = 2 experiments.



Supplementary Figure 10 | Downstream effects of Pin1 inhibition or degradation. (**a**) Immunoblot analysis following BJP-06-005-3 treatment at the indicated concentrations for 5 h in C2; (**b**) Immunoblot analysis of C2 + FKBP12^{F36V}-Pin1 cells treated with the indicated concentrations of dTAG-13 for 5 h; (**c**) PATU-8988T cells treated with DMSO or BJP-06-005-3 (10 μM) for the indicated time points and stained with propidium iodide. The mean peak height of n = 3 biologically independent samples are shown as dot-plots, and are representative of n = 2 independent experiments; (**d**) C2 + FKBP12^{F36V}-Pin1 cells treated for 24 h with dTAG-13 (100 nM) and stained with propidium iodide. The mean peak height of n = 3 biologically independent samples are shown as dot-plots, and are representative of n = 2 independent experiments; (**e**) Exemplary gating strategy applied to FACS analyses in this study: representative side scatter versus forward scatter plot used to select live singlet cells (**left**), histogram of propidium iodide (YEL-B-HLin) content for singlet cells (right); (**f**) Immunoblot analysis of PATU-8988T cells treated for 4 h with DMSO, BJP-06-005-3 (10 μM) or BJP-R (10 μM), followed by a 4 h treatment with DMSO or recombinant human Wnt3a (50 ng/mL); (**g**) Immunoblot analysis of PATU-8988T cells treated for 5.5 h with DMSO, BJP-06-005-3 (10 μM) or BJP-R (10 μM), followed by fractionation of the nuclear and cytoplasmic components; (**h**) Immunoblot analysis of PATU-8988T cells treated for 4 h with DMSO, BJP-06-005-3 (10 μM) or BJP-R (10 μM), followed by a 4 h treatment with DMSO or recombinant human Wnt3a (50 ng/mL), followed by fractionation of the nuclear and cytoplasmic components.

Uncropped immunoblots for **a-b** are displayed in Supplementary Figure 17, and are representative of n = 3 independent experiments. Uncropped immunoblots for **f-h** are displayed in Supplementary Figure 18, and are representative of n = 2 independent experiments.



Transcription Factor	Proportion of Target Genes Present	Adjusted p-value	Odds Ratio
FOXP3	29/350	7.37E-04	2.62
ESR1	51/871	2.66E-03	1.85
YAP1	12/115	2.75E-02	3.30
MYC	50/957	2.62E-02	1.65
RNF2	15/175	2.67E-02	2.71
SREBF2	8/62	3.45E-02	4.08
POU5F1	22/338	4.92E-02	2.06

Supplementary Figure 11 | The transcriptional consequences of BJP-06-005-3 treatment. (**a**) Transcriptome-wide impact of 4 h BJP-06-005-3 treatment in PATU-8988T cells normalized to DMSO and BJP-R. Adjusted p-values derived

from FDR-corrected two-sided t-test, calculated from n = 3 biologically independent samples for each treatment condition; (**b**) Representation of GSEA pathway enrichments of transcripts differentially expressed upon treatment with BJP-06-005-3, determined from n = 3 biologically independent samples for each treatment condition (**c-f**) Sample GSEA plots for enriched pathways; (**g**) Transcription factors whose target gene sets were significantly enriched in the set of significantly down-regulated genes (at least 1.25-fold down-regulated with adj p-value < .05), identified using EnrichR analysis against the Transcription Factors PPI dataset. Adjusted p-values derived from the FDR-corrected Fisher exact test based on EnrichR analysis.[ii]

Figure 3c

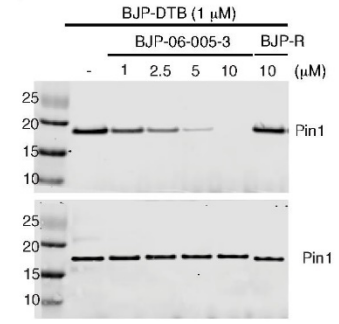


Figure 3d

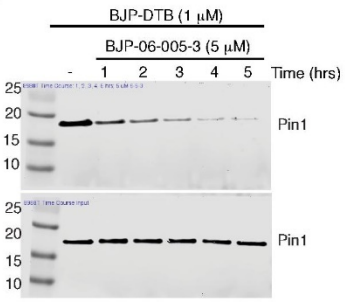


Figure 4a

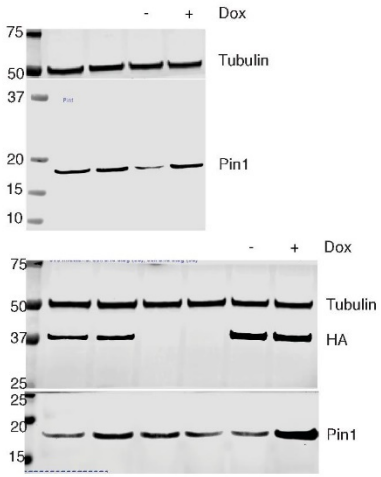


Figure 4e

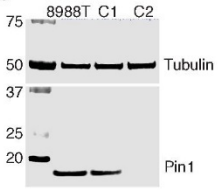


Figure 4g

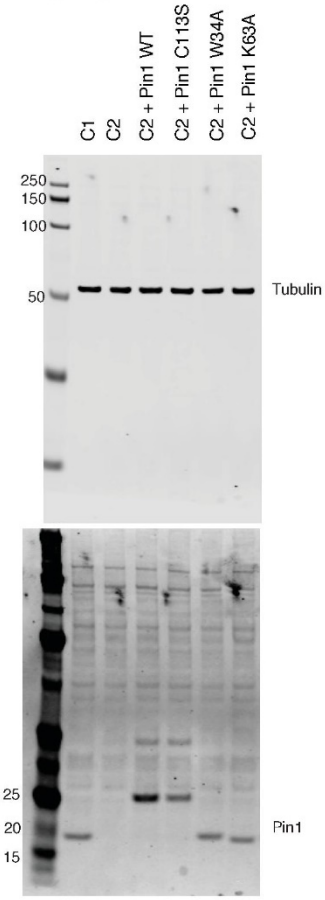


Figure 5d

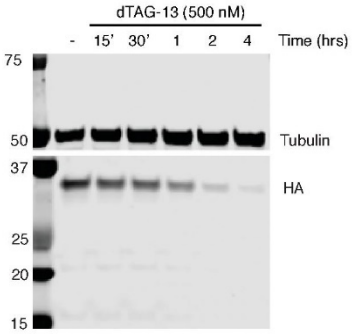
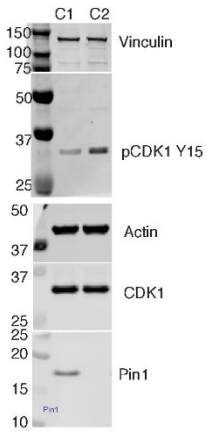
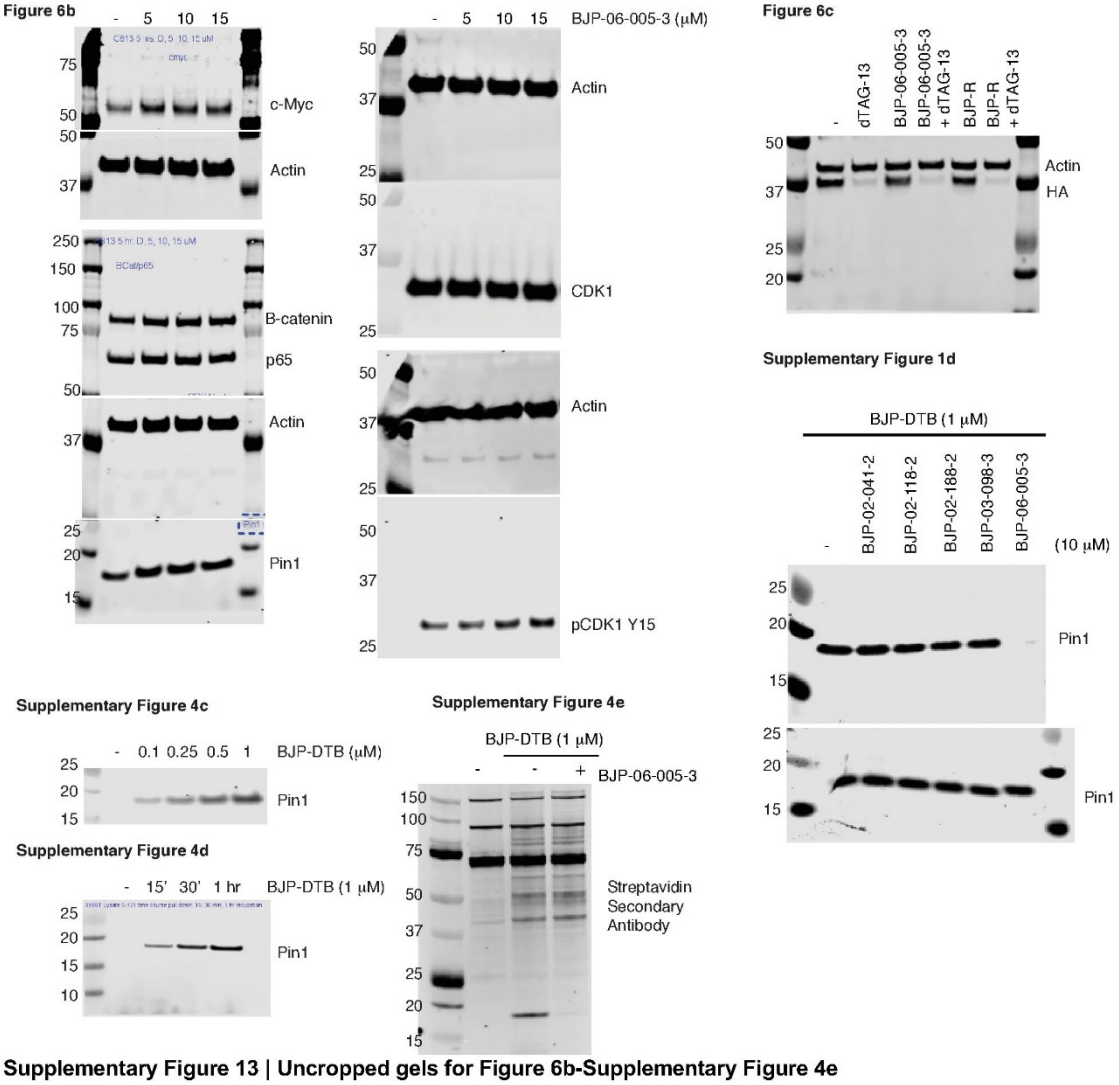


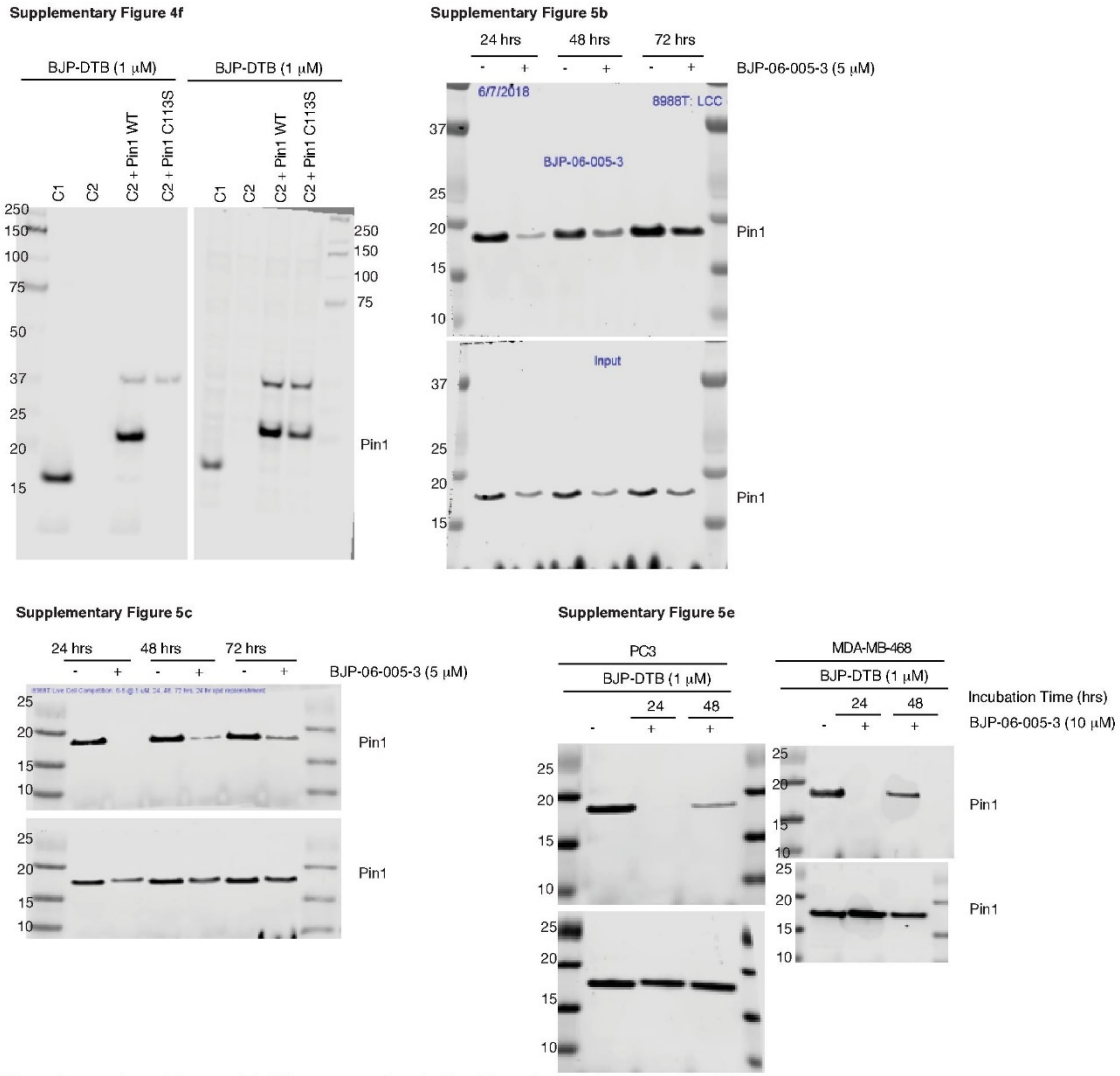
Figure 6a



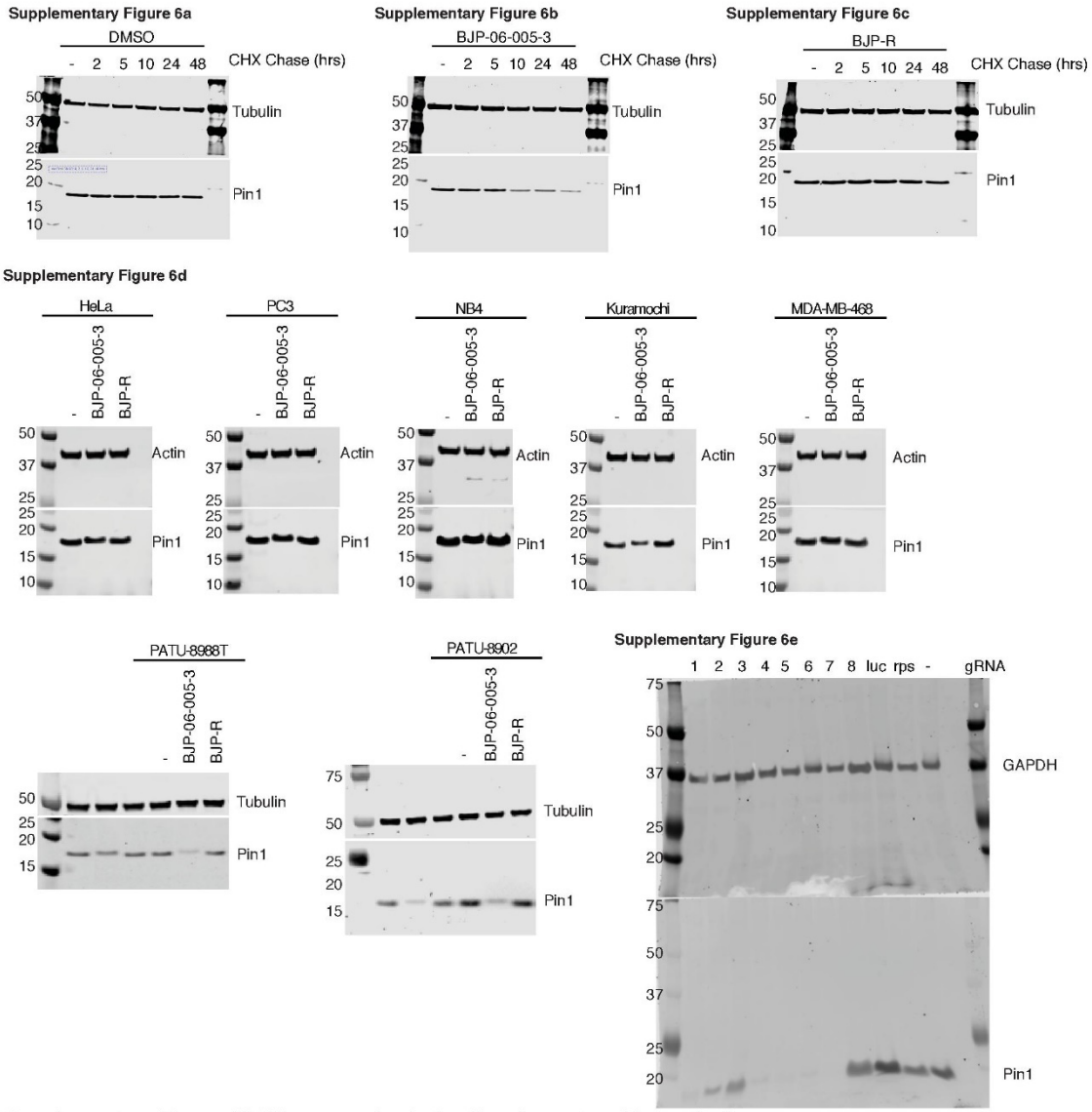
Supplementary Figure 12 | Uncropped gels for Figure 3c-6a



Supplementary Figure 13 | Uncropped gels for Figure 6b-Supplementary Figure 4e

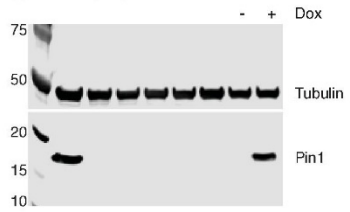


Supplementary Figure 14 | Uncropped gels for Supplementary Figure 4f-5e

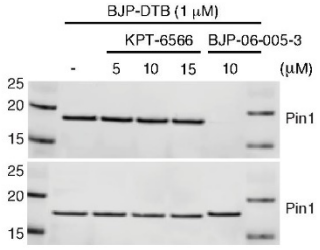


Supplementary Figure 15 | Uncropped gels for Supplementary Figure 6a-6e

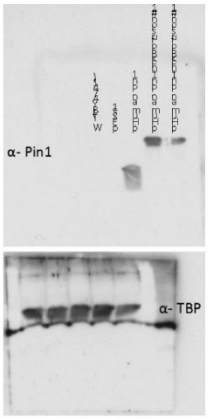
Supplementary Figure 7h



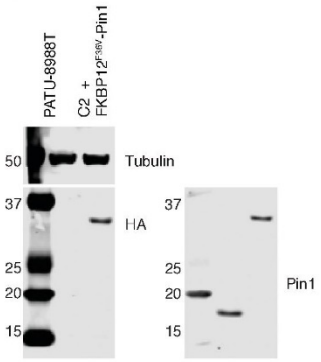
Supplementary Figure 8d



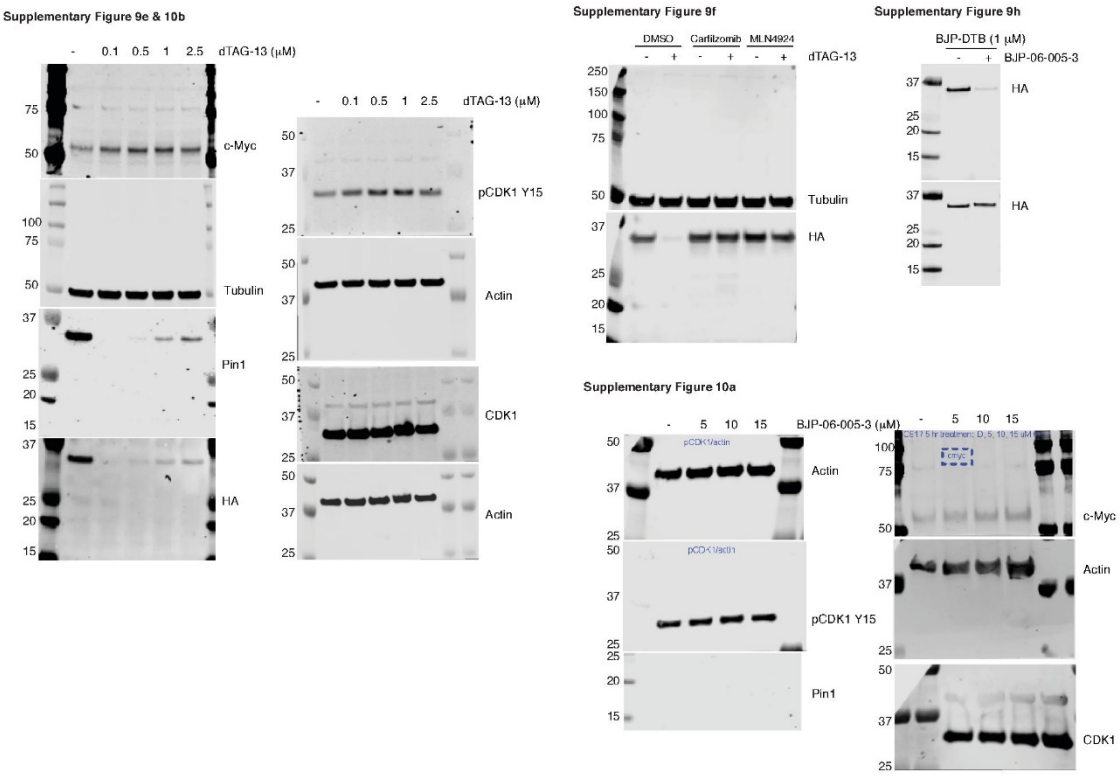
Supplementary Figure 9a



Supplementary Figure 9c

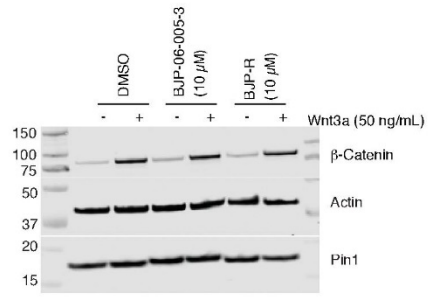


Supplementary Figure 16 | Uncropped gels for Supplementary Figure 7h-9c

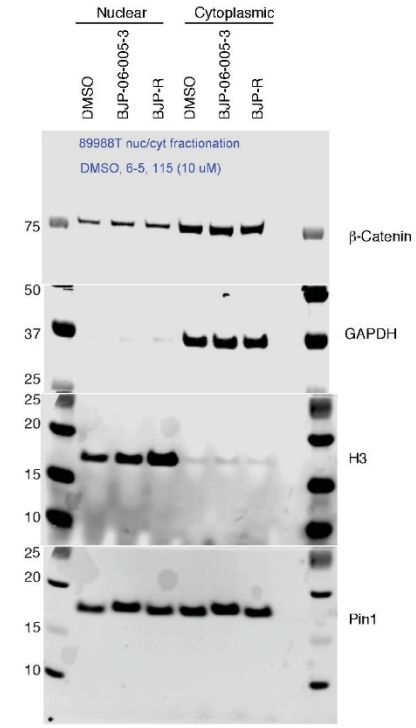


Supplementary Figure 17 | Uncropped gels for Supplementary Figure 9e-10b

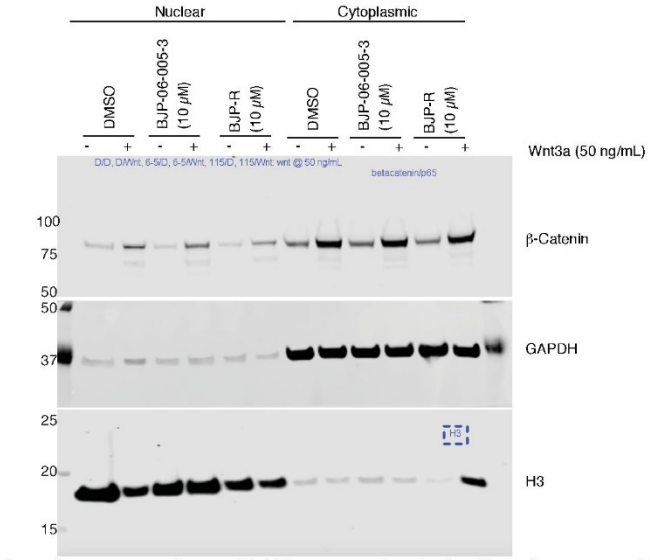
Supplementary Figure 10f



Supplementary Figure 10g



Supplementary Figure 10h



Supplementary Figure 18 | Uncropped gels for Supplementary Figure 10e-10g

SUPPLEMENTARY NOTE: SYNTHETIC METHODS

Chemical Synthesis. All solvents and reagents were used as obtained from commercial sources. ^{1}H and ^{13}C NMR spectra were recorded with a Bruker 500 MHz NMR spectrometer, and chemical shifts are reported in parts per million (ppm) downfield from tetramethylsilane (TMS). Coupling constants (J) are reported in Hz. Spin multiplicities are described as s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), and m (multiplet). Mass spectra were obtained on a Waters Acquity UPLC. Preparative HPLC was performed on a Waters Sunfire C18 column (19mm x 5mm) using a gradient of 15-95% methanol in water containing 0.05% trifluoroacetic acid (TFA) over 22 min (28 min run time) at a flow rate of 20 mL/min.

General Procedure for Synthesis of Peptidic Backbone. Peptidic backbones were prepared according to established solid-phase peptide synthesis (SPPS) protocols with Fmoc-protected amino acids. Wang or Rink Amide resin was used on a 0.12 mmol scale, and Fmoc-protected amino acids were deprotected with 20% piperidine in DMF. Activation of amino acids (0.36 mmol, 3 eq.) for coupling was performed with HATU (0.36 mmol, 3 eq.) and DIPEA (0.96 mmol, 8 eq.). Cleavage from the resin was performed with TFA/TIPS/water 95/2.5/2.5. The residual solid was dissolved in H_2O and acetonitrile, and then lyophilized to yield the peptidic backbone as a white powder.

Equipment Setup. Peptides were synthesized using a manual peptide-synthesis apparatus. A vacuum manifold (Promega, cat#A7231) was connected to a vacuum system through a trap. Three-way stopcocks were placed on the manifold and connected to a line of nitrogen gas. All unused inlets on the manifold were capped with rubber septa.

Poly-Prep chromatography columns (2 mL bed volume, Bio-Rad cat#7311550) containing the resin were placed on the manifold using the three-way stopcocks.[iii]

Reagents. Rink Amide Resin (VWR), Wang Resin (Fisher Scientific), N,N-Dimethyl formamide (DMF, Sigma Aldrich), Dichloromethane (DCM, Sigma Aldrich), Piperidine (Sigma Aldrich), N,N-Diisopropylethylamine (DIPEA, Sigma Aldrich) Trifluoroacetic acid (TFA, Fisher Scientific), Diethyl ether (Fisher Scientific), Acetic anhydride (Sigma Aldrich), Triisopropylsilane (TIPS, Sigma Aldrich), Pyridine (Sigma Aldrich), Ninhydrin Test Kit (Fisher Scientific), standard commercially available unnatural and natural Fmoc-protected amino acids (Fisher Scientific and VWR). KPT-6566 was purchased from Accela Chembio, and Juglone was purchased from Sigma Aldrich.

Removal of the FMOC. A solution of 20% piperidine in DMF was prepared and added to the resin (1 mL of solution per 0.12 mmol resin). The resin was agitated via nitrogen bubbling for 15 minutes. The solution was drained using vacuum filtration. Repeat. The resin was then washed 3x with DMF (2 mL per 0.12 mmol resin).

Coupling of Amino Acids. The resin was suspended in DMF (1 mL per 0.06 mmol resin) and agitated by nitrogen bubbling for 5 minutes at room temperature. The DMF was then removed via vacuum filtration. In a separate flask, the Fmoc-protected amino acid (3 eq. relative to resin) and HATU (3 eq. relative to resin) were dissolved in DMF (1 mL of DMF per 0.06 mmol Wang resin). To this solution was added DIPEA (8 eq. relative to resin). The resulting amino acid solution was stirred at room temperature until homogeneous, and then added to the resin in the Poly-Prep column. The reaction mixture was agitated by nitrogen bubbling for 2-3 hrs at room temperature, after which the solution was removed via vacuum filtration and the resin was washed 3x with DMF, 3x with DCM, 3x

with MeOH (for each wash step: use enough solvent to slurry the resin, bubble nitrogen for 1 min, then drain solution via vacuum filtration). The completion of the coupling was assessed using the Ninhydrin test kit. If the test indicated that there was still unreacted N-terminal amine, the coupling step was repeated.

Cleavage from the Resin. After being dried under vacuum, the resin was transferred to an 8 mL reaction vial with screw cap, equipped with a magnetic stir bar. After cooling the flask in an ice bath, a solution of TFA/TIPS/water 95/2.5/2.5 was added (1 mL per 100 mg of resin). The reaction mixture was stirred for 10 minutes at 0°C. Then, the mixture was stirred at room temperature overnight. The resin was filtered using a fine sintered glass funnel and washed 3 times with a minimal amount of TFA. To the filtrate was added 8-10 times the volume of cold diethyl ether. If necessary, the mixture was stored at 4°C overnight to precipitate the peptide. The peptide was then filtered using a fine sintered glass funnel and washed with cold diethyl ether. Purification by HPLC and subsequent lyophilization yielded the pure peptide chain in quantitative yields.

Lys(MTT) Deprotection on Rink Amide Resin (applies to Scheme 1a). The resin was suspended in DCM (1 mL per 100 mg of resin) in a Poly-Prep column, and gently agitated with nitrogen gas. The DCM was drained by vacuum filtration. The resin was then suspended in 3% TFA in DCM (1 mL per 100 mg of resin) and agitated with nitrogen gas for 10 minutes. The solution was drained by vacuum filtration. Repeat. Finally, the resin was washed 2x with DCM, 2x with MeOH, 2x with DCM, 1x with 1% DIPEA in DMF, and then 2x with DMF (for each wash step: use enough solvent to slurry the resin, bubble nitrogen for 1 min, then drain solution via vacuum filtration).

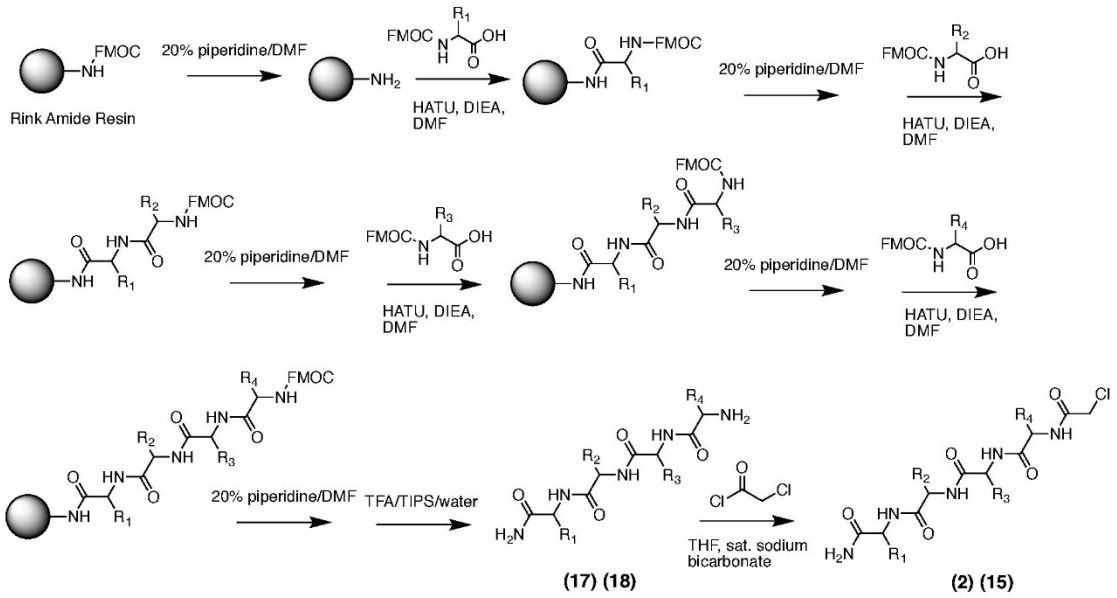
D-Desthiobiotin Coupling to Lys on Rink Amide Resin (applies to Scheme 1a). To a solution of d-desthiobiotin (10 eq. relative to resin) in 1:1 v/v DMF/DMSO (0.5 mL total volume per 100 mg resin), was added HATU (10 eq. relative to resin) and DIPEA (26 eq. relative to resin) in DMF (1 mL volume per 400 mg resin). This solution was then added to the resin, which was agitated with nitrogen gas for 6 hours. After draining the solution via vacuum filtration, the resin was washed 2x with DMF, 2x with DCM, and 2x with MeOH (for each wash step: use enough solvent to slurry the resin, bubble nitrogen for 1 min, then drain solution via vacuum filtration).

Addition of the First Amino Acid to Wang Resin (applies to Scheme 3). The resin was suspended in 9:1 v/v DCM/DMF (1 mL per 100 mg of resin) in a Poly-Prep column, and gently agitated with nitrogen gas. In a separate flask, the Fmoc-protected amino acid (3 eq. relative to resin) and HATU (3 eq. relative to resin) were dissolved in DMF (1 mL of DMF per 0.06 mmol Wang resin). To this solution was added DIPEA (8 eq. relative to resin). The resulting amino acid solution was stirred at room temperature until homogeneous, and then added to the resin in the Poly-Prep column. In a separate flask, 4-dimethylamino-pyridine (DMAP) (0.1 eq. relative to resin) was dissolved in a minimum amount of DMF. The resulting solution was also added to the resin solution in the Poly-Prep column. The reaction mixture was agitated via nitrogen bubbling for 2-3 hours at room temperature. The solution was then drained by vacuum filtration, and the resin washed 3x with DMF (for each wash step: use enough solvent to slurry the resin, bubble nitrogen for 1 min, then drain solution via vacuum filtration). Then, 2 eq. (relative to the resin) of acetic anhydride and 2 eq. (relative to the resin) of pyridine in DMF (1 mL DMF per 0.12 mmol resin) were added to the Poly-Prep column. The resin was agitated for

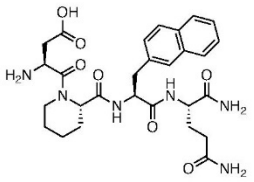
an additional 30 minutes at room temperature to cap any unreacted hydroxyls on the resin. The solution was then drained by vacuum filtration, and the resin was washed 3x with DMF, 3x with DCM, and 3x with MeOH (for each wash step: use enough solvent to slurry the resin, bubble nitrogen for 1 min, then drain solution via vacuum filtration).

SYNTHETIC SCHEMES AND REACTION METHODS

SCHEME 1 | Synthesis of BJP-02-038-2, BJP-07-017-3



Synthesis of BJP-02-038-2

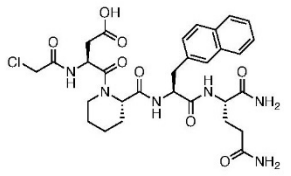


(S)-3-amino-4-((S)-2-(((S)-1-(((S)-1,5-diamino-1,5-dioxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-oxobutanoic acid

(17)

(**17**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Gln(Trt)-OH, Fmoc-L-2-Nal-OH, Fmoc-L-Pip-OH, Fmoc-L-Asp(OtBu)-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 12.78 (s, 1H), 8.14 (d, J = 5 Hz, 1H), 7.97 (dd, J = 15, 5 Hz, 1H), 7.85 (d, J = 5 Hz, 1H), 7.72 (d, J = 5 Hz, 1H), 7.51 (t, J = 5 Hz, 2H), 7.46 (t, J

= 5 Hz, 2H), 7.34 (s, 1H), 7.33-7.31 (m, 1H), 7.24-7.18 (m, 1H), 7.02 (s, 1H), 6.71 (s, 1H), 4.88 (d, J = 5 Hz, 1H), 4.68-4.62 (m, 1H), 4.54 (d, J = 5 Hz, 1H), 4.16-4.10 (m, 2H), 3.63-3.59 (m, 1H), 3.55-3.51 (m, 2H), 3.46-3.43 (m, 1H), 2.79 (t, J = 10 Hz, 1H), 2.70-2.66 (m, 1H), 2.11-2.08 (m, 1H), 2.03 (t, J = 10 Hz, 2H), 1.88-1.80 (m, 2H), 1.74-1.68 (m, 2H), 1.40-1.38 (m, 2H), 1.29-1.20 (m, 1H), 1.07-0.98 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.3, 173.4, 171.3, 170.2, 168.5, 158.3, 158.1, 134.2, 133.9, 132.1, 129.1, 127.7, 127.5, 126.6, 126.1, 125.8, 124.3, 54.1, 52.9, 52.7, 47.3, 43.4, 34.5, 31.8, 28.5, 26.7, 25.0, 20.2; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{28}H_{37}N_6O_7$, 569.27; found, 569.36.



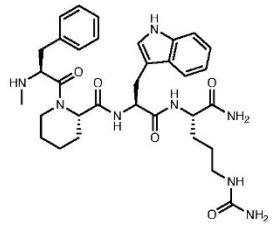
(S)-3-(2-chloroacetamido)-4-((S)-2-(((S)-1-(((S)-1,5-diamino-1,5-dioxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-oxobutanoic acid (BJP-02-038-2) (2)

To a mixture of (**17**) (45 mg, 0.08 mmol) in 0.2 mL THF and 0.2 mL saturated sodium bicarbonate was added a solution of chloroacetyl chloride (0.012 mL, 0.16 mmol) in 0.2 mL THF. The resulting mixture was stirred in an ice bath for 2 hours. The mixture was then diluted with 0.5 mL of dichloromethane and 0.5 mL of methanol and filtered through a 0.45 μm syringe filter. The solvent was evaporated, the product purified via HPLC and lyophilized to yield **BJP-02-038-2** as a white powder (35 mg, 68%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 12.41 (s, 1H), 8.70 (d, J = 5 Hz, 1H), 8.16 (d, J = 5 Hz, 1H), 7.94 (d, J = 5 Hz, 1H), 7.85 (d, J = 5 Hz, 1H), 7.73 (d, J = 5 Hz, 1H), 7.52 (t, J = 10 Hz, 2H), 7.46 (t, J = 10 Hz, 2H), 7.38 (t, J = 5 Hz, 1H), 7.34 (d, J = 10 Hz, 1H), 7.20 (d, J

= 10 Hz, 2H), 7.01 (s, 1H), 6.71 (s, 1H), 4.99-4.95 (m, 1H), 4.86 (d, J = 5 Hz, 1H), 4.70-4.65 (m, 1H), 4.63-4.58 (m, 1H), 4.14-4.09 (m, 1H), 3.97 (s, 1H), 3.59-3.56 (m, 2H), 2.73-2.67 (m, 1H), 2.04 (t, J = 5 Hz, 2H), 1.96-1.93 (m, 1H), 1.87-1.82 (m, 2H), 1.74-1.68 (m, 1H), 1.33-1.28 (m, 1H), 1.21-1.17 (m, 2H), 1.11-1.06 (m, 1H), 0.99-0.94 (m, 1H), 0.87-0.82 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.3, 173.6, 173.5, 172.7, 171.3, 170.2, 170.1, 165.7, 134.2, 133.9, 132.1, 129.1, 127.9, 127.6, 126.6, 126.1, 125.8, 124.3, 53.8, 52.8, 46.4, 42.7, 36.8, 34.6, 31.9, 28.3, 26.8, 26.1, 25.2, 20.4; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{30}H_{38}ClN_6O_8$, 645.24; found, 645.12.

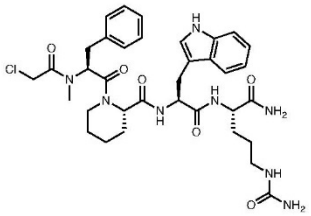
Synthesis of BJP-07-017-3 (derivative of BJP-06-005-3 used for x-ray crystallography)



(S)-N-((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)-1-(methyl-L-phenylalanyl)piperidine-2-carboxamide (18)

(**18**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-L-N-Me-Phe-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.73 (s, 1H), 7.90 (t, J = 5 Hz, 1H), 7.68 (d, J = 5 Hz, 1H), 7.53 (t, J = 5 Hz, 1H), 7.30-7.20 (m, 4H), 7.15-7.10 (m, 3H), 7.06 (d, J = 5 Hz, 1H), 7.00 (t, J = 5 Hz, 1H), 6.98-6.96 (m, 1H), 6.91 (t, J = 5 Hz, 1H), 5.87 (s, 1H), 5.33 (s, 1H), 4.98 (d, J = 5 Hz, 1H), 4.64-4.56 (m, 1H), 4.19-4.13 (m, 2H), 4.05 (d, J = 10 Hz, 1H), 3.87-3.86 (m, 1H), 3.14-3.05 (m, 2H), 2.97-2.87 (m, 4H), 2.74 (t, J = 5 Hz, 1H), 2.37 (t, J = 5 Hz, 1H), 2.10 (t, J = 5 Hz, 2H), 1.66 (d, J = 10 Hz, 1H), 1.59-1.53 (m, 1H), 1.45-1.38 (m,

3H), 1.27-1.22 (m, 4H), 0.91-0.83 (m, 1H), 0.78-0.67 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.8, 171.3, 170.4, 169.0, 167.7, 159.3, 136.5, 134.1, 130.3, 129.2, 128.9, 128.0, 127.8, 124.1, 121.4, 118.9, 118.7, 111.9, 110.5, 58.9, 57.9, 56.4, 54.2, 52.5, 37.3, 32.3, 31.8, 30.2, 27.9, 27.0, 25.7, 24.1, 19.8; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{33}H_{45}N_8O_5$, 633.35; found, 633.48.

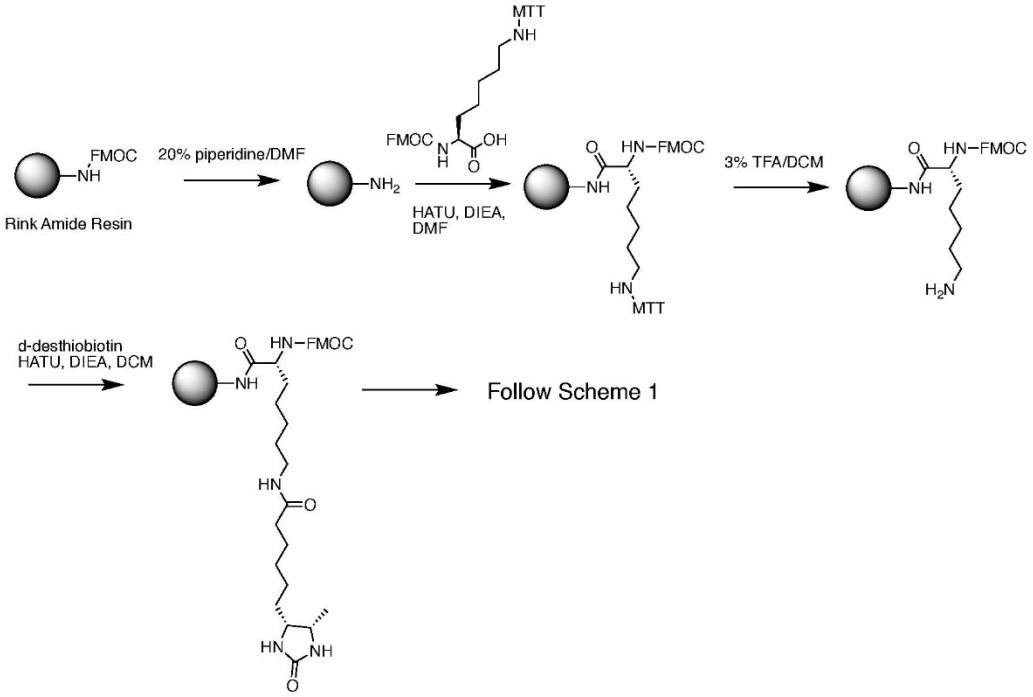


(S)-N-((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)-1-(N-(2-chloroacetyl)-N-methyl-L-phenylalanyl)piperidine-2-carboxamide (BJP-07-017-3) (15)

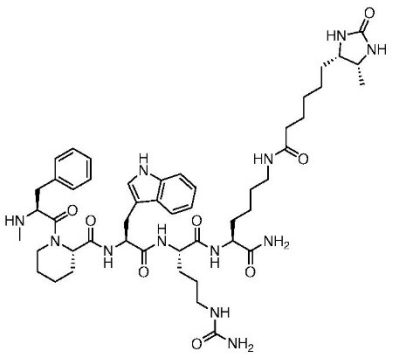
BJP-07-017-3 was synthesized using the same procedure as **BJP-02-038-2** from (**18**) (51 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. **BJP-07-017-3** was obtained as a white powder (38 mg, 66%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.82 (s, 1H), 7.96 (d, J = 10 Hz, 1H), 7.66 (d, J = 10 Hz, 1H), 7.61 (d, J = 10 Hz, 1H), 7.32-7.30 (m, 2H), 7.26-7.21 (m, 4H), 7.20-7.15 (m, 2H), 7.11-7.09 (m, 1H), 7.04 (t, J = 5 Hz, 2H), 6.97 (t, J = 5 Hz, 1H), 5.97 (s, 1H), 4.98 (d, J = 5 Hz, 1H), 4.72-4.67 (m, 1H), 4.33-4.30 (m, 1H), 4.26-4.21 (m, 1H), 4.20-4.17 (m, 1H), 3.43-3.40 (m, 1H), 3.21-3.17 (m, 1H), 3.03-2.91 (m, 3H), 2.88 (s, 3H), 2.85 (d, J = 5 Hz, 1H), 2.76 (s, 1H), 2.64-2.58 (m, 1H), 2.11-2.04 (m, 2H), 1.71-1.61 (m, 1H), 1.53-1.38 (m, 1H), 1.37-1.34 (m, 4H), 1.21-1.14 (m, 2H), 1.06-0.93 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 171.7, 170.2, 169.5, 169.1, 166.0, 159.3, 137.9, 136.6, 129.7, 129.2, 128.7, 128.6, 127.7, 126.8, 124.2, 121.3, 119.0, 118.7, 111.7, 110.5, 54.7, 53.7, 53.1, 52.6, 42.7,

42.3, 40.9, 34.6, 30.6, 30.1, 27.7, 26.9, 25.2, 20.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{35}H_{46}ClN_8O_6$, 709.32; found, 709.51.

SCHEME 1a | Synthesis of BJP-DTB



Synthesis of BJP-DTB

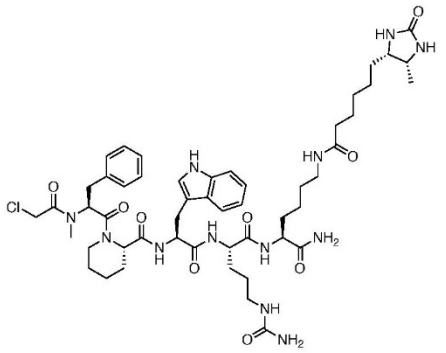


(S)-N-((S)-1-(((S)-1-(((S)-1-amino-6-(6-((4S,5R)-5-methyl-2-oxoimidazolidin-4-yl)hexanamido)-1-oxohexan-2-yl)amino)-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-

indol-3-yl)-1-oxopropan-2-yl)-1-(methyl-L-phenylalanyl)piperidine-2-carboxamide (19)

(**19**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Lys(MTT)-OH, Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-L-N-Me-Phe-OH.

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.71 (d, J = 10 Hz, 1H), 9.10 (s, 1H), 8.78-8.60 (m, 1H), 8.04 (t, J = 5 Hz, 1H), 7.81-7.74 (m, 1H), 7.72-7.65 (m, 2H), 7.56-7.52 (m, 1H), 7.30-7.21 (m, 3H), 7.11 (t, J = 10 Hz, 2H), 7.06 (s, 1H), 6.99 (t, J = 5 Hz, 1H), 6.95-6.89 (m, 2H), 6.22 (s, 1H), 6.04 (s, 1H), 5.87 (s, 1H), 5.31 (s, 1H), 4.96 (s, 1H), 4.63-4.58 (m, 1H), 4.24-4.02 (m, 4H), 3.85 (s, 1H), 3.51 (t, J = 5 Hz, 1H), 3.39 (s, 1H), 3.22-3.18 (m, 1H), 3.12-3.05 (m, 1H), 2.90 (s, 3H), 2.85 (s, 2H), 2.71 (t, J = 10 Hz, 1H), 2.35 (t, J = 5 Hz, 1H), 2.31-2.26 (m, 1H), 2.06 (t, J = 5 Hz, 2H), 2.00 (s, 2H), 1.95 (t, J = 5 Hz, 2H), 1.63-1.51 (m, 2H), 1.48-1.34 (m, 5H), 1.32-1.20 (m, 7H), 1.18-1.06 (m, 7H), 0.87 (d, J = 5 Hz, 3H), 0.78-0.67 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 172.5, 171.7, 170.4, 169.1, 163.3, 159.3, 158.6, 158.3, 136.5, 134.3, 134.1, 130.3, 129.2, 128.9, 127.9, 127.8, 124.0, 121.4, 121.3, 118.9, 118.7, 110.6, 58.9, 57.9, 56.3, 55.5, 54.1, 52.9, 52.7, 52.5, 50.7, 38.8, 37.3, 35.8, 35.4, 32.3, 31.7, 31.2, 29.9, 29.4, 29.2, 28.0, 26.9, 26.0, 25.7, 23.2, 19.9, 15.9; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{49}H_{73}N_{12}O_8$, 957.57; found, 957.25.



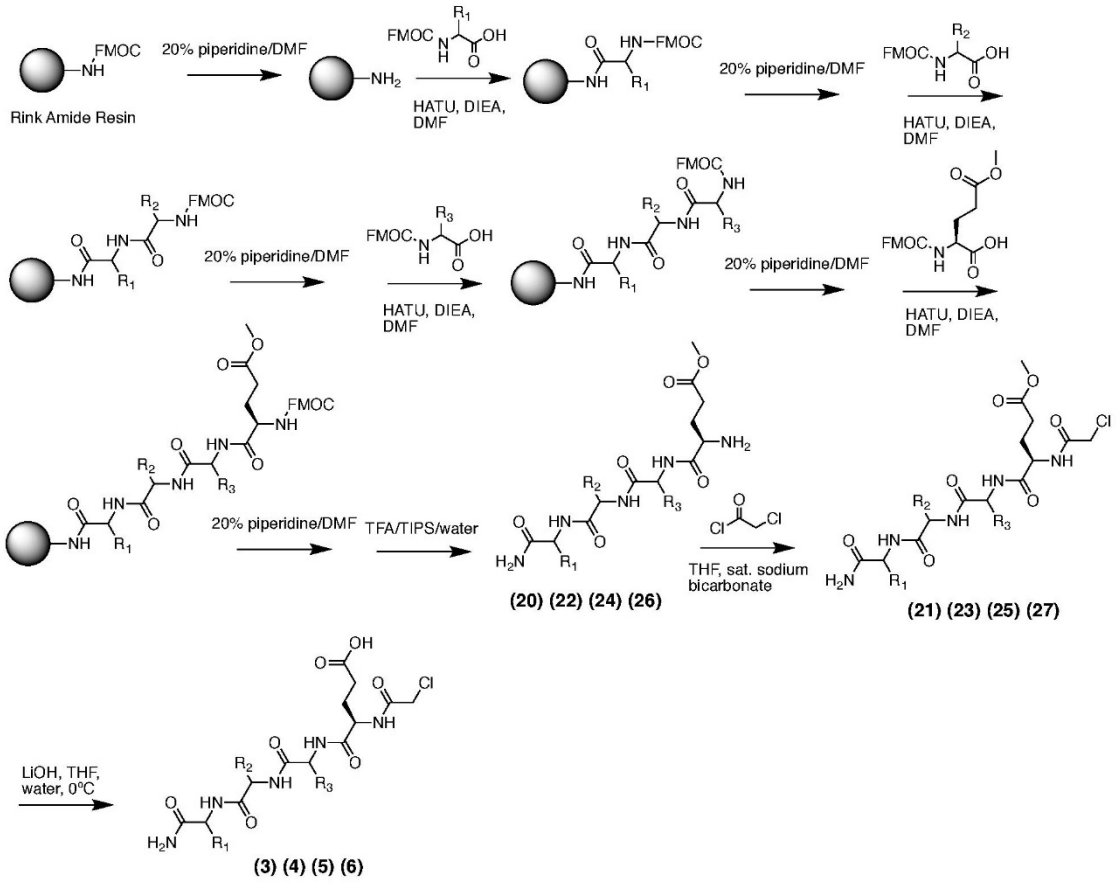
(S)-N-((S)-1-(((S)-1-(((S)-1-amino-6-(6-((4S,5R)-5-methyl-2-oxoimidazolidin-4-yl)hexanamido)-1-oxohexan-2-yl)amino)-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)-1-(N-(2-chloroacetyl)-N-methyl-L-phenylalanyl)piperidine-2-carboxamide (BJP-DTB) (16)

BJP-DTB was synthesized using the same procedure as **BJP-02-038-2** from (**19**) (76 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. **BJP-DTB** was obtained as a white powder (58 mg, 70%).

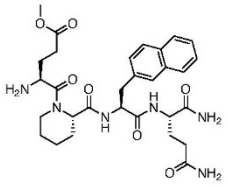
^{1}H NMR (500 MHz, DMSO-d_6): δ 10.74 (d, J = 10 Hz, 1H), 8.08 (d, J = 5 Hz, 1H), 7.75 (d, J = 5 Hz, 1H), 7.68 (t, J = 5 Hz, 1H), 7.58-7.49 (m, 2H), 7.24 (d, J = 10 Hz, 2H), 7.17-7.10 (m, 5H), 7.03-6.95 (m, 3H), 6.90 (t, J = 5 Hz, 1H), 5.92 (s, 1H), 5.43-5.40 (m, 1H), 5.37-5.34 (m, 1H), 4.89 (s, 1H), 4.66-4.61 (m, 1H), 4.56-4.51 (m, 1H), 4.31 (s, 1H), 4.24-4.18 (m, 2H), 4.11-4.07 (m, 2H), 4.0 (d, J = 10 Hz, 1H), 3.75 (d, J = 10 Hz, 1H), 3.13-3.10 (m, 2H), 2.98-2.86 (m, 7H), 2.80 (s, 3H), 2.78 (s, 1H), 1.96 (t, J = 10 Hz, 3H), 1.54-1.55 (m, 3H), 1.49-1.35 (m, 5H), 1.35-1.21 (m, 7H), 1.19-1.08 (m, 7H), 0.98-0.91 (m, 2H), 0.88 (d, J = 5 Hz, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.0, 172.6, 172.1, 171.7, 170.2, 169.5, 166.4, 166.0, 163.4, 161.4, 159.4, 137.9, 136.6, 129.7, 129.6, 128.6, 127.7, 126.8, 126.8, 124.2, 121.4, 119.0, 118.7, 117.7, 111.7, 110.4, 55.5, 54.7, 53.6, 53.1, 52.8, 50.7, 42.3,

38.8, 35.8, 34.6, 32.2, 30.9, 30.6, 29.9, 29.7, 29.4, 29.2, 27.7, 26.9, 25.9, 25.7, 25.1, 23.2, 20.5, 15.9; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{51}H_{74}ClN_{12}O_9$, 1033.53; found, 1033.40.

SCHEME 2 | Synthesis of BJP-02-041-2, BJP-02-118-2, BJP-02-188-2, BJP-03-098-3

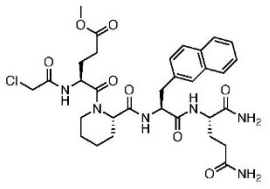


Synthesis of BJP-02-041-2



methyl (S)-4-amino-5-((S)-2-(((S)-1-(((S)-1,5-diamino-1,5-dioxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoate (20)

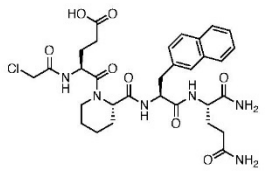
(**20**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Gln(Trt)-OH, Fmoc-L-2-Nal-OH, Fmoc-L-Pip-OH, Fmoc-L-Glu(OMe)-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.12-8.09 (m, 2H), 8.03-8.02 (m, 3H), 7.94 (d, J = 5 Hz, 1H), 7.85 (d, J = 5 Hz, 1H), 7.73-7.69 (m, 1H), 7.50 (t, J = 5 Hz, 1H), 7.45 (t, J = 5 Hz, 1H), 7.34-7.31 (m, 3H), 7.19-7.16 (m, 2H), 7.02 (s, 1H), 6.69 (s, 1H), 4.93 (d, J = 5 Hz, 1H), 4.64-4.60 (m, 1H), 4.37 (s, 1H), 4.16-4.10 (m, 1H), 3.65 (d, J = 5 Hz, 1H), 3.53 (s, 3H), 3.25-3.20 (m, 2H), 3.00 (t, J = 5 Hz, 1H), 2.13-2.08 (m, 1H), 2.02 (t, J = 5 Hz, 2H), 1.88-1.75 (m, 3H), 1.74-1.65 (m, 1H), 1.50-1.42 (m, 3H), 1.27-1.19 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.8, 174.2, 173.3, 172.9, 171.3, 170.6, 169.6, 168.8, 158.5, 158.2, 133.8, 132.1, 129.0, 127.4, 126.6, 126.0, 125.8, 124.2, 53.9, 52.6, 52.1, 49.5, 43.3, 34.6, 31.8, 28.5, 27.2, 25.5, 25.2, 20.2; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{30}H_{41}N_6O_7$, 597.30; found, 597.47.



methyl (S)-4-(2-chloroacetamido)-5-((S)-2-(((S)-1-(((S)-1,5-diamino-1,5-dioxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoate (21)

(**21**) was synthesized using the same procedure as **BJP-02-038-2** from (**20**) (48 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. (**21**) was obtained as a white powder (32 mg, 59%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.52 (d, J = 5 Hz, 1H), 8.19 (t, J = 5 Hz, 2H), 7.92 (d, J = 5 Hz, 2H), 7.79-7.78 (m, 2H), 7.58 (t, J = 5 Hz, 1H), 7.52 (t, J = 5 Hz, 1H), 7.41-7.40 (m, 2H), 7.36-7.33 (m, 1H), 7.29-7.27 (m, 1H), 7.25 (s, 1H), 7.14 (s, 1H), 7.10 (t, J = 5 Hz, 1H), 6.77 (s, 1H), 4.95 (d, J = 5 Hz, 1H), 4.79-4.75 (m, 2H), 4.72-4.68 (m, 1H), 4.23-4.14 (m, 2H), 4.10 (d, J = 5 Hz, 2H), 3.59 (s, 3H), 3.17 (s, 1H), 2.35-2.31 (m, 2H), 2.12-2.11 (m, 1H), 2.09-2.07 (m, 1H), 1.91 (s, 1H), 1.82-1.77 (m, 1H), 1.71-1.67 (m, 1H), 1.47-1.37 (m, 3H), 1.28-1.24 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.2, 173.5, 170.7, 166.2, 163.8, 158.5, 158.2, 134.2, 133.8, 132.1, 129.1, 127.7, 127.5, 126.7, 126.6, 126.0, 125.8, 124.2, 53.9, 53.4, 52.7, 52.6, 51.9, 48.5, 43.2, 42.8, 34.5, 31.8, 29.6, 27.0, 25.2, 20.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{32}H_{42}ClN_6O_8$, 673.27; found, 673.39.



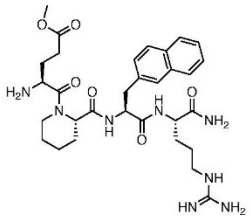
(S)-4-(2-chloroacetamido)-5-((S)-2-(((S)-1-(((S)-1,5-diamino-1,5-dioxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoic acid (BJP-02-041-2) (3)

To a mixture of (**21**) (27 mg, 0.04 mmol) in THF/water (4:1, 0.5 mL total volume), was added LiOH (1.9 mg, 0.08 mmol). The resulting mixture was stirred in an ice bath at 0°C for 3 hours. The mixture was then neutralized with 1 N HCl (1.5 eq.). The solvent was

evaporated and the product directly purified by HPLC purification and lyophilized to yield **BJP-02-041-2** as a white powder (20 mg, 75%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 12.12 (s, 1H), 8.42 (d, J = 5 Hz, 1H), 8.13 (d, J = 10 Hz, 1H), 7.95-7.90 (m, 1H), 7.85-7.80 (m, 2H), 7.71 (t, J = 5 Hz, 1H), 7.51 (t, J = 5 Hz, 1H), 7.45 (t, J = 5 Hz, 1H), 7.34 (d, J = 5 Hz, 1H), 7.31-7.26 (m, 1H), 7.20-7.09 (m, 2H), 7.07-6.96 (m, 3H), 6.70 (s, 1H), 6.48 (s, 1H), 4.88 (s, 1H), 4.72-4.62 (m, 2H), 4.15-4.08 (m, 1H), 4.03 (d, J = 5 Hz, 1H), 3.61-3.52 (m, 2H), 2.75-2.69 (m, 1H), 2.19 (t, J = 5 Hz, 1H), 2.04-1.97 (m, 3H), 1.87-1.78 (m, 2H), 1.72-1.68 (m, 1H), 1.60-1.56 (m, 1H), 1.37-1.17 (m, 4H), 0.99-0.93 (m, 1H); MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{31}H_{40}ClN_6O_8$, 659.26; found, 659.64.

Synthesis of BJP-02-118-2

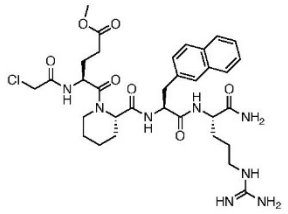


methyl (S)-4-amino-5-((S)-2-(((S)-1-(((S)-1-amino-5-guanidino-1-oxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoate (22)

(**22**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Arg(Pmc)-OH, Fmoc-L-2-Nal-OH, Fmoc-L-Pip-OH, Fmoc-L-Glu(OMe)-OH.

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.18 (d, J = 5 Hz, 2H), 8.13-8.12 (m, 3H), 8.03 (d, J = 5 Hz, 1H), 7.92 (d, J = 5 Hz, 1H), 7.79 (d, J = 5 Hz, 1H), 7.62-7.59 (m, 1H), 7.58 (d, J = 5

Hz, 1H), 7.53 (t, J = 5 Hz, 1H), 7.40-7.38 (m, 2H), 7.25 (s, 1H), 7.14 (s, 1H), 5.00 (d, J = 5 Hz, 1H), 4.74-4.70 (m, 1H), 4.48-4.45 (m, 1H), 4.25-4.21 (m, 1H), 3.73 (d, J = 5 Hz, 2H), 3.60 (s, 3H), 3.33-3.28 (m, 2H), 3.12-3.10 (m, 3H), 2.46-2.42 (m, 1H), 2.18 (d, J = 5 Hz, 1H), 1.97-1.91 (m, 2H), 1.87-1.81 (m, 1H), 1.75-1.68 (m, 1H), 1.58-1.51 (m, 4H), 1.48-1.43 (m, 3H), 1.35-1.30 (m, 1H), 1.29-1.24 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.4, 172.9, 171.4, 170.6, 168.8, 158.8, 158.5, 157.2, 133.9, 133.8, 132.1, 129.1, 127.5, 127.4, 126.6, 126.1, 125.8, 124.1, 53.9, 52.5, 52.5, 52.1, 49.5, 43.3, 34.6, 29.8, 28.5, 27.3, 25.5, 25.2, 20.2; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{31}H_{45}N_8O_6$, 625.35; found, 625.53.

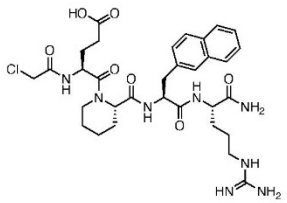


methyl (S)-5-((S)-2-(((S)-1-(((S)-1-amino-5-guanidino-1-oxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoate (23)

(**23**) was synthesized using the same procedure as **BJP-02-038-2** from (**22**) (50 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. (**23**) was obtained as a white powder (35 mg, 61%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.50 (d, J = 5 Hz, 1H), 8.20 (d, J = 5 Hz, 1H), 8.04 (d, J = 5 Hz, 1H), 7.93-7.89 (m, 3H), 7.79 (d, J = 5 Hz, 1H), 7.58 (d, J = 5 Hz, 2H), 7.53 (t, J = 5 Hz, 2H), 7.41 (d, J = 5 Hz, 2H), 7.28 (s, 2H), 7.14 (s, 2H), 4.95 (d, J = 5 Hz, 1H), 4.78-4.71 (m, 2H), 4.67-4.64 (m, 1H), 4.26-4.21 (m, 1H), 4.16 (d, J = 5 Hz, 1H), 4.09 (d, J = 5 Hz, 2H), 3.59 (s, 3H), 3.13-3.09 (m, 2H), 2.34 (t, J = 5 Hz, 2H), 2.07-2.05 (m, 1H), 1.95-

1.87 (m, 1H), 1.76-1.67 (m, 2H), 1.58-1.52 (m, 1H), 1.48-1.40 (m, 4H), 1.30-1.24 (m, 2H), 1.15-1.08 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.5, 171.4, 170.7, 170.7, 170.2, 169.8, 167.2, 166.2, 158.6, 158.3, 157.2, 134.1, 133.9, 132.1, 129.1, 127.6, 127.5, 126.6, 126.1, 125.8, 124.2, 53.8, 52.6, 52.5, 51.9, 48.5, 43.2, 42.8, 34.5, 29.6, 27.0, 25.5, 20.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{33}H_{46}ClN_8O_7$, 701.32; found, 701.56.

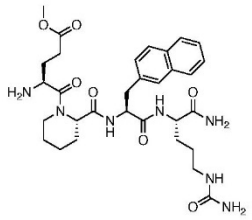


(S)-5-((S)-2-(((S)-1-(((S)-1-amino-5-guanidino-1-oxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoic acid (BJP-02-118-2) (4)

BJP-02-118-2 was synthesized using the same procedure as **BJP-02-041-2** from (**23**) (28 mg, 0.04 mmol) and LiOH (1.9 mg, 0.08 mmol) in THF/water. **BJP-02-118-2** was obtained as a white powder (20 mg, 73%).

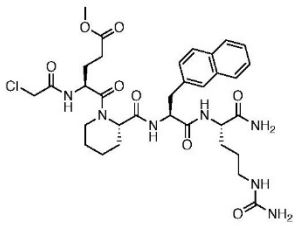
^{1}H NMR (500 MHz, DMSO-d_6): δ 12.13 (s, 1H), 8.39 (d, J = 5 Hz, 1H), 8.08 (d, J = 10 Hz, 1H), 7.79-7.74 (m, 3H), 7.70-7.67 (m, 1H), 7.46 (t, J = 5 Hz, 1H), 7.39 (t, J = 5 Hz, 3H), 7.29-7.27 (m, 1H), 7.06 (d, J = 5 Hz, 1H), 6.80 (s, 1H), 4.88 (d, J = 5 Hz, 1H), 4.69-4.65 (m, 2 H), 4.19-4.15 (m, 1H), 4.11-4.04 (m, 1H), 4.01 (s, 1H), 3.57 (d, J = 10 Hz, 1H), 3.23-3.16 (m, 1H), 3.04 (d, J = 10 Hz, 2H), 2.99-2.95 (m, 1H), 2.78-2.72 (m, 1H), 2.23 (t, J = 5 Hz, 1H), 2.16 (t, J = 5 Hz, 1H), 2.03-1.97 (m, 2H), 1.81-1.72 (m, 1H), 1.66-1.62 (m, 1H), 1.59-1.37 (m, 4H), 1.32-1.25 (m, 3H), 1.18-1.12 (m, 2H), 1.01-0.92 (m, 2H); MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{32}H_{44}ClN_8O_7$, 687.30; found, 687.14.

Synthesis of BJP-02-188-2



methyl (S)-4-amino-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoate (24)

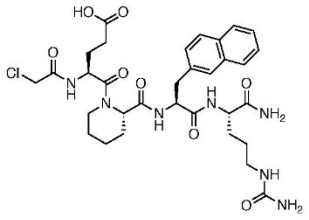
(**24**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-2-Nal-OH, Fmoc-L-Pip-OH, Fmoc-L-Glu(OMe)-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.19 (d, J = 5 Hz, 1H), 8.14-8.12 (m, 2H), 7.95 (d, J = 5 Hz, 1H), 7.91 (d, J = 5 Hz, 1H), 7.80-7.77 (m, 1H), 7.58 (t, J = 5 Hz, 1H), 7.53 (t, J = 5 Hz, 1H), 7.41-7.39 (m, 2H), 7.23 (d, J = 5 Hz, 1H), 7.06 (s, 1H), 6.55 (s, 1H), 5.98 (t, J = 5 Hz, 1H), 5.41 (s, 2H), 5.00 (d, J = 5 Hz, 1H), 4.73-4.69 (m, 1H), 4.46-4.41 (m, 1H), 4.26-4.18 (m, 2H), 3.74-3.71 (m, 1H), 3.61 (s, 3H), 3.12-3.06 (m, 1H), 3.00-2.91 (m, 2H), 2.20-2.18 (m, 1H), 1.97-1.92 (m, 2H), 1.88-1.83 (m, 1H), 1.70-1.64 (m, 1H), 1.58-1.47 (m, 4H), 1.40-1.33 (m, 3H), 1.27-1.24 (m, 2H), 1.20-1.11 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.7, 172.9, 171.2, 170.5, 168.9, 168.7, 159.3, 134.1, 133.8, 132.1, 129.0, 127.5, 126.6, 126.0, 125.8, 124.2, 53.9, 52.6, 52.5, 52.1, 49.5, 43.3, 34.5, 30.2, 28.5, 27.2, 27.1, 26.9, 25.5, 25.2, 20.2; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{31}H_{44}N_7O_7$, 626.33; found, 626.48.



methyl (S)-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoate (25)

(**25**) was synthesized using the same procedure as **BJP-02-038-2** from (**24**) (50 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. (**25**) was obtained as a white powder (30 mg, 54%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.52-8.49 (m, 1H), 8.21 (d, J = 5 Hz, 1H), 7.92 (d, J = 5 Hz, 1H), 7.78 (t, J = 5 Hz, 1H), 7.59 (t, J = 5 Hz, 1H), 7.52 (t, J = 5 Hz, 1H), 7.41 (d, J = 5 Hz, 1H), 7.26-7.24 (m, 1H), 7.07 (s, 1H), 6.00-5.97 (m, 1H), 5.40 (s, 2H), 4.95 (d, J = 5 Hz, 1H), 4.80-4.73 (m, 2H), 4.69-4.62 (m, 1H), 4.26-4.19 (m, 2H), 4.16 (d, J = 5 Hz, 1H), 4.10-4.09 (m, 2H), 3.59 (s, 3H), 2.99-2.93 (m, 2H), 2.84-2.79 (m, 1H), 2.37-2.33 (m, 1H), 2.27 (t, J = 5 Hz, 1H), 2.08-2.04 (m, 1H), 1.93-1.86 (m, 2H), 1.71-1.64 (m, 2H), 1.56-1.48 (m, 2H), 1.43-1.36 (m, 5H), 1.27-1.23 (m, 2H), 1.15-1.08 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.6, 173.8, 173.5, 171.3, 170.7, 170.6, 167.5, 166.2, 159.3, 158.4, 158.1, 134.2, 133.8, 132.1, 129.1, 127.5, 126.6, 126.0, 125.8, 124.2, 53.8, 52.6, 51.9, 48.6, 48.5, 43.2, 42.8, 34.5, 30.1, 29.6, 26.9, 25.3, 20.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{33}H_{45}ClN_7O_8$, 702.30; found, 702.46.

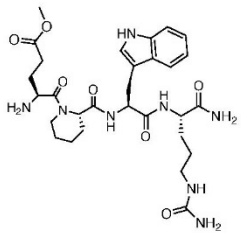


(S)-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoic acid (BJP-02-188-2) (5)

BJP-02-188-2 was synthesized using the same procedure as **BJP-02-041-2** from (**25**) (28 mg, 0.04 mmol) and LiOH (1.9 mg, 0.08 mmol) in THF/water. **BJP-02-188-2** was obtained as a white powder (20 mg, 72%).

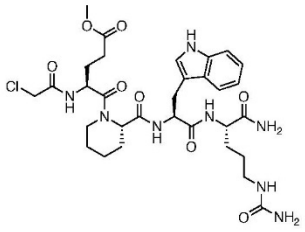
^{1}H NMR (500 MHz, DMSO-d_6): δ 12.14 (s, 1H), 8.42 (d, J = 5 Hz, 1H), 8.13 (d, J = 5 Hz, 1H), 7.87-7.80 (m, 2H), 7.71 (t, J = 5 Hz, 1H), 7.51 (t, J = 5 Hz, 1H), 7.45 (t, J = 5 Hz, 1H), 7.34-7.33 (m, 2H), 7.17 (d, J = 5 Hz, 1H), 6.99 (d, J = 5 Hz, 1H), 5.87 (s, 1H), 5.32 (s, 1H), 4.87 (d, J = 5 Hz, 1H), 4.72-4.65 (m, 2H), 4.60-4.56 (m, 1H), 4.18-4.14 (m, 1H), 4.09 (d, J = 10 Hz, 2H), 4.02 (d, J = 5 Hz, 2H), 3.61-3.58 (m, 1H), 3.54-3.50 (m, 1H), 2.93-2.85 (m, 2H), 2.74 (t, J = 10 Hz, 1H), 2.25 (t, J = 5 Hz, 1H), 2.19 (t, J = 5 Hz, 1H), 2.01-1.98 (m, 1H), 1.87-1.78 (m, 2H), 1.65-1.56 (m, 2H), 1.47-1.41 (m, 1H), 1.35-1.27 (m, 4H), 1.19-1.14 (m, 1H), 1.07-0.97 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 174.6, 173.7, 171.3, 170.9, 170.6, 166.2, 159.3, 134.1, 133.8, 132.1, 132.0, 129.0, 127.7, 127.5, 126.6, 126.0, 125.8, 124.2, 53.8, 52.6, 52.6, 48.6, 43.2, 42.9, 34.5, 30.1, 29.9, 27.1, 27.0, 26.7, 25.3, 20.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{32}H_{43}ClN_7O_8$, 688.29; found, 688.84.

Synthesis of BJP-03-098-3



methyl (S)-4-amino-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-5-oxopentanoate (26)

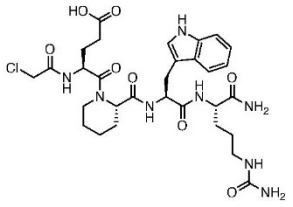
(**26**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol Rink Amide resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-L-Glu(OMe)-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.71 (s, 1H), 8.02 (s, 3H), 7.86 (d, J = 5 Hz, 1H), 7.80 (d, J = 5 Hz, 1H), 7.52 (d, J = 5 Hz, 1H), 7.24 (d, J = 10 Hz, 1 H), 7.17 (s, 1H), 7.07 (s, 1H), 6.99-6.95 (m, 2H), 6.90 (t, J = 5 Hz, 1H), 5.86 (s, 1H), 4.94 (s, 1H), 4.55-4.51 (m, 1H), 4.37 (s, 1H), 4.16-4.12 (m, 1H), 3.63 (d, J = 10 Hz, 1H), 3.54 (s, 3H), 3.52 (s, 1H), 3.10 (dd, J = 15, 5 Hz, 1H), 3.01 (t, J = 10 Hz, 1H), 2.92-2.83 (m, 3H), 2.13 (d, J = 10 Hz, 1H), 1.89-1.84 (m, 2H), 1.78-1.73 (m, 1H), 1.60-1.37 (m, 6H), 1.30-1.23 (m, 4H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.8, 172.9, 171.5, 170.3, 168.8, 159.3, 158.1, 136.5, 127.8, 124.0, 121.3, 118.9, 118.7, 111.71, 110.39, 53.9, 52.6, 52.5, 52.1, 49.6, 43.3, 30.2, 28.6, 27.7, 27.3, 26.9, 25.4, 25.2, 20.21; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{29}H_{43}N_8O_7$, 615.32; found, 615.02.



methyl (S)-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoate (27)

(**27**) was synthesized using the same procedure as **BJP-02-038-2** from (**26**) (49 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. (**27**) was obtained as a white powder (35 mg, 63%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.80 (s, 1H), 8.47 (d, J = 5 Hz, 1H), 7.93 (d, J = 5 Hz, 1H), 7.84 (t, J = 5 Hz, 1H), 7.64 (d, J = 5 Hz, 1H), 7.59 (d, J = 5 Hz, 1H), 7.32 (d, J = 5 Hz, 1H), 7.26 (s, 1H), 7.14 (s, 1H), 7.05 (d, J = 10 Hz, 2H), 6.97 (d, J = 10 Hz, 1H), 6.53 (s, 1H), 5.92 (t, J = 5 Hz, 1H), 5.39 (s, 2H), 4.97 (d, J = 5 Hz, 1H), 4.77-4.74 (m, 1H), 4.66-4.62 (m, 1H), 4.24-4.21 (m, 1H), 4.09 (d, J = 5 Hz, 2H), 3.59 (s, 3H), 3.20-3.16 (m, 2H), 3.01-2.91 (m, 3H), 2.84 (t, J = 5 Hz, 1H), 2.40 (t, J = 5 Hz, 1H), 2.33 (t, J = 5 Hz, 2H), 2.12-2.09 (m, 1H), 1.89-1.81 (m, 1H), 1.69-1.63 (m, 2H), 1.53-1.45 (m, 2H), 1.41-1.32 (m, 2H), 1.24-1.17 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 172.3, 170.9, 166.1, 159.3, 158.7, 155.5, 136.5, 132.0, 131.8, 129.6, 124.1, 121.3, 118.9, 118.7, 111.7, 110.4, 52.6, 42.8, 40.9, 30.2, 30.2, 29.9, 29.6, 27.1, 26.9, 26.9, 26.8, 25.3, 20.5, 18.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{31}H_{44}ClN_8O_8$, 691.30; found, 691.83.



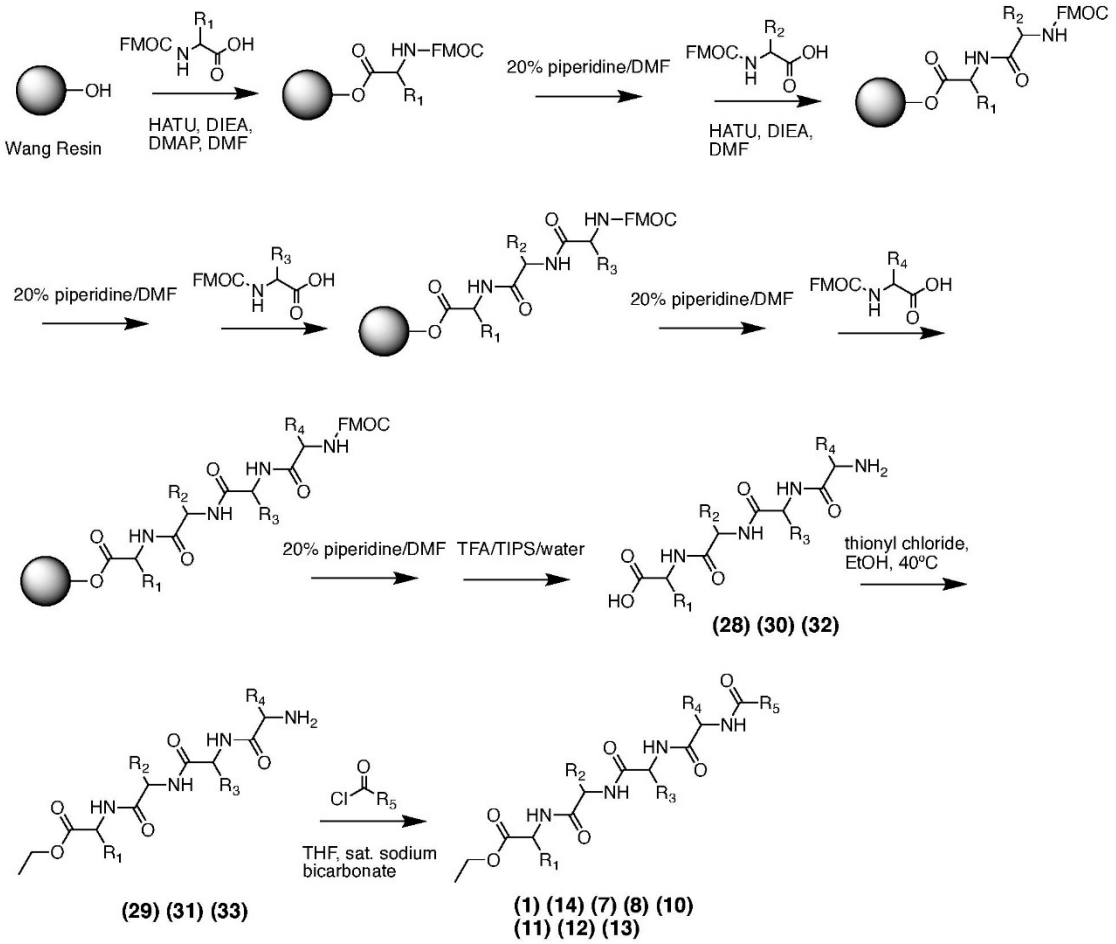
(S)-5-((S)-2-(((S)-1-(((S)-1-amino-1-oxo-5-ureidopentan-2-yl)amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)piperidin-1-yl)-4-(2-chloroacetamido)-5-oxopentanoic acid (BJP-03-098-3) (6)

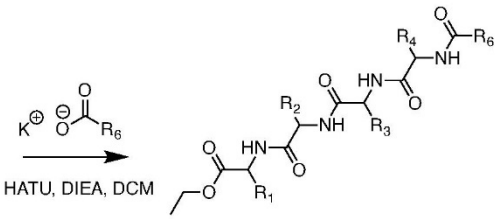
BJP-03-098-3 was synthesized using the same procedure as **BJP-02-041-2** from (**27**) (28 mg, 0.04 mmol) and LiOH (1.9 mg, 0.08 mmol) in THF/water. **BJP-03-098-3** was obtained as a white powder (21 mg, 77%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 12.22 (s, 1H), 10.81 (s, 1H), 8.48 (d, J = 5 Hz, 1H), 7.95 (d, J = 5 Hz, 1H), 7.89-7.84 (m, 1H), 7.67 (d, J = 10 Hz, 1H), 7.62-7.57 (m, 1H), 7.32-7.27 (m, 2H), 7.14 (s, 1H), 7.09-7.04 (m, 2H), 6.96 (t, J = 5 Hz, 1H), 5.98 (s, 1H), 5.40 (s, 1H),

4.97 (d, J = 5 Hz, 1H), 4.78-4.71 (m, 1H), 4.68-4.61 (m, 1H), 4.24-4.19 (m, 1H), 4.10 (d, J = 5 Hz, 1H), 3.04-2.92 (m, 3H), 2.86-2.78 (m, 1H), 2.25 (t, J = 5 Hz, 2H), 2.08 (s, 1H), 1.87-1.80 (m, 1H), 1.76-1.71 (m, 1H), 1.68-1.58 (m, 2H), 1.52-1.46 (m, 2H), 1.41-1.33 (m, 3H), 1.27-1.21 (m, 2H), 1.19-1.12 (m, 1H), 1.11-1.05 (m, 1H), 0.90-0.84 (m, 1H); MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{30}H_{42}ClN_8O_8$, 677.28; found, 677.73.

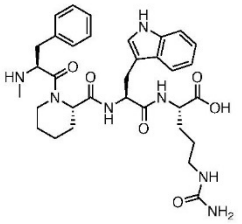
SCHEME 3 | Synthesis of BJP-06-005-3, BJP-R, BJP-06-121-3, 8, 9, 10, 11, 12, 13





(9)

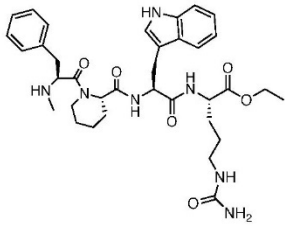
Synthesis of BJP-06-005-3 and BJP-R



(S)-2-((S)-3-(1H-indol-3-yl)-2-((S)-1-(methyl-L-phenylalanyl)piperidine-2-carboxamido)propanamido)-5-ureidopentanoic acid (28)

(**28**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol of Wang resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-L-N-Me-Phe-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 12.57 (s, 1H), 10.75 (d, J = 20 Hz, 1H), 9.26 (s, 1H), 8.85-8.75 (m, 1H), 8.61 (s, 1H), 8.26-8.22 (m, 1H), 7.72-7.65 (m, 1H), 7.56 (t, J = 5 Hz, 1H), 7.29-7.27 (m, 2H), 7.24 (d, J = 10 Hz, 1H), 7.16 (d, J = 5 Hz, 1H), 7.12 (d, J = 5 Hz, 2H), 7.08 (dd, J = 15, 5 Hz, 1H), 7.02-6.96 (m, 1H), 6.91 (dd, J = 15, 5 Hz, 1H), 5.91 (d, J = 5 Hz, 1H), 5.33 (s, 1H), 4.75-4.8 (m, 1H), 4.66-4.61 (m, 1H), 4.49 (s, 1H), 4.15-4.10 (m, 2H), 4.05-4.02 (m, 1H), 3.86 (d, J = 5 Hz, 1H), 2.97-2.93 (m, 1H), 2.88 (s, 3H), 2.35 (t, J = 5 Hz, 1H), 2.06 (t, J = 5 Hz, 2H), 1.68-1.63 (m, 1H), 1.56-1.40 (m, 4H), 1.34-1.33 (m, 3H), 1.21-1.17 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 171.7, 168.9, 167.6, 159.2, 136.5, 134.1, 130.3, 129.2, 128.9, 127.9, 124.1, 121.4, 118.9, 111.8, 110.5, 58.9,

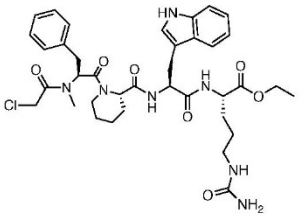
57.9, 56.4, 53.8, 52.4, 43.3, 37.4, 35.4, 32.3, 31.8, 28.9, 28.1, 27.2, 25.7, 25.4, 24.1, 19.8; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{33}H_{44}N_7O_6$, 634.33; found, 634.08.



ethyl (S)-2-((S)-3-(1H-indol-3-yl)-2-((S)-1-(methyl-L-phenylalanyl)piperidine-2-carboxamido)propanamido)-5-ureidopentanoate (29)

To a solution of (**28**) (50 mg, 0.08 mmol) in ethanol (0.5 mL), was added thionyl chloride (0.02 mL, 0.28 mmol) drop-wise in an ice bath. The resulting solution was stirred at 40°C for 1 hour. The solvent was then evaporated and the product purified by HPLC purification to yield (**29**) as a white powder (28 mg, 52%).

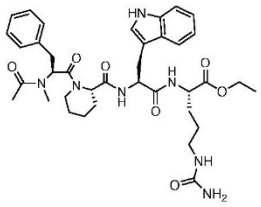
^{1}H NMR (500 MHz, DMSO-d_6): δ 10.74 (d, J = 10 Hz, 1H), 8.39 (t, J = 10 Hz, 1H), 7.70 (dd, J = 20, 10 Hz, 1H), 7.54 (t, J = 5 Hz, 1H), 7.26 (d, J = 10 Hz, 2H), 7.23 (d, J = 5 Hz, 1H), 7.15 (d, J = 5 Hz, 1H), 7.10 (d, J = 10 Hz, 2H), 7.05 (d, J = 5 Hz, 1H), 7.01-6.95 (m, 1H), 6.90 (t, J = 5 Hz, 1H), 5.89 (s, 1H), 5.33 (s, 1H), 4.69-4.60 (m, 2H), 4.14-4.09 (m, 2H), 4.04-3.97 (m, 3H), 3.84 (s, 1H), 3.10-2.99 (m, 2H), 2.93-2.86 (m, 4H), 2.71 (t, J = 10 Hz, 1H), 2.34 (t, J = 5 Hz, 2H), 2.04 (t, J = 5 Hz, 2H), 1.67-1.62 (m, 2H), 1.53-1.49 (m, 1H), 1.43-1.40 (m, 1H), 1.30-1.13 (m, 4H), 1.09 (t, J = 5 Hz, 3H), 0.90-0.85 (m, 1H), 0.76-0.67 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.3, 171.9, 170.2, 168.9, 167.6, 159.2, 136.5, 134.3, 130.3, 129.2, 127.9, 124.1, 121.4, 118.9, 111.9, 110.5, 60.9, 58.9, 57.9, 56.3, 53.7, 52.6, 43.3, 37.3, 35.4, 32.3, 31.7, 28.7, 28.2, 27.1, 25.8, 25.4, 24.1, 19.8, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{35}H_{48}N_7O_6$, 662.37; found, 662.09.



ethyl (S)-2-((S)-2-((S)-1-(N-(2-chloroacetyl)-N-methyl-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (BJP-06-005-3)

(1)

BJP-06-005-3 was synthesized using the same procedure as **BJP-02-038-2** from (**29**) (53 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. **BJP-06-005-3** was obtained as a white powder (31 mg, 53%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.75 (s, 1H), 8.40 (d, J = 5 Hz, 1H), 7.53 (dd, J = 15, 10 Hz, 1H), 7.24 (d, J = 5 Hz, 1H), 7.19-7.09 (m, 4H), 7.04-7.01 (m, 1H), 6.98 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.90 (s, 1H), 5.44-5.40 (m, 1H), 5.34 (t, J = 5 Hz, 1H), 4.90 (d, J = 5 Hz, 1H), 4.71-4.67 (m, 1H), 4.62-4.57 (m, 1H), 4.23 (d, J = 15 Hz, 1H), 4.17-4.13 (m, 1H), 4.10 (d, J = 15 Hz, 1H), 4.03 (q, J = 5 Hz, 2H), 3.76 (d, J = 15 Hz, 1H), 3.33 (d, J = 10 Hz, 2H), 3.09 (dd, J = 15, 5 Hz, 1H), 2.95-2.84 (m, 4H), 2.81 (s, 3H), 2.75 (s, 2H), 2.01 (d, J = 10 Hz, 1H), 1.69-1.63 (m, 1H), 1.57-1.53 (m, 1H), 1.38-1.25 (m, 3H), 1.21-1.17 (m, 1H), 1.12 (t, J = 5 Hz, 3H), 1.07-1.05 (m, 1H), 0.98-0.85 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.4, 172.3, 169.9, 169.9, 169.4, 165.9, 160.0, 159.2, 137.9, 137.7, 136.6, 129.7, 128.6, 127.7, 126.7, 124.2, 121.3, 119.0, 118.6, 111.7, 110.4, 60.9, 54.7, 53.3, 53.0, 52.7, 42.3, 34.6, 30.6, 28.7, 28.1, 27.1, 26.9, 25.5, 25.2, 20.6, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{49}ClN_7O_7$, 738.34; found, 738.13.

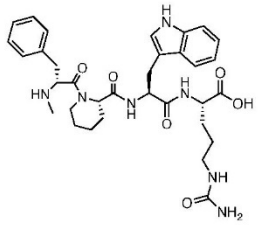


ethyl (S)-2-((S)-2-((S)-1-(N-acetyl-N-methyl-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (BJP-R) (14)

To a mixture of (**29**) (28 mg, 0.04 mmol) in 0.2 mL THF and 0.2 mL saturated sodium bicarbonate was added a solution of acetyl chloride (0.006 mL, 0.08 mmol) in 0.2 mL THF. The resulting mixture was stirred in an ice bath for 2 hours. The mixture was then diluted with 0.5 mL of dichloromethane and 0.5 mL of methanol, and filtered through a 0.45 μm syringe filter. The solvent was evaporated, the product purified via HPLC and lyophilized to yield **BJP-R** as a white powder (15 mg, 50%).

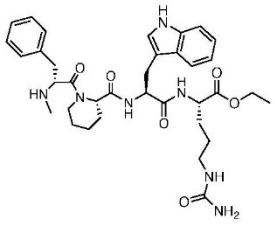
^{1}H NMR (500 MHz, DMSO-d_6): δ 10.82 (s, 1H), 8.47 (d, J = 5 Hz, 1H), 7.62 (d, J = 10 Hz, 1H), 7.58 (d, J = 10 Hz, 1H), 7.31 (d, J = 10 Hz, 2H), 7.24 (t, J = 5 Hz, 2H), 7.21 (s, 1H), 7.17 (t, J = 5 Hz, 1H), 7.12-7.09 (m, 1H), 7.05 (t, J = 5 Hz, 1H), 6.98 (t, J = 5 Hz, 1H), 5.96 (s, 1H), 5.53 (q, J = 5 Hz, 1H), 5.39 (s, 1H), 4.98 (d, J = 5 Hz, 1H), 4.79-4.74 (m, 1H), 4.67-4.65 (m, 1H), 4.27-4.21 (m, 1H), 4.11 (q, J = 5 Hz, 2H), 3.63 (d, J = 5 Hz, 1H), 3.19-3.15 (m, 2H), 3.02-2.93 (m, 4H), 2.90-2.86 (m, 1H), 2.81 (s, 3H), 2.76 (s, 1H), 2.67 (s, 1H), 2.09 (s, 3H), 1.81 (s, 1H), 1.76-1.71 (m, 1H), 1.53 (d, J = 5 Hz, 1H), 1.46-1.42 (m, 2H), 1.38-1.33 (m, 1H), 1.28-1.24 (m, 1H), 1.20 (t, J = 5 Hz, 3H), 1.00-0.94 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.4, 172.3, 172.0, 170.0, 169.9, 169.7, 169.5, 169.2, 159.2, 158.8, 158.5, 138.3, 136.6, 129.6, 128.5, 127.7, 126.6, 124.2, 121.3, 119.0, 118.6, 111.7, 110.4, 60.9, 53.8, 52.9, 52.7, 42.7, 34.8, 31.4, 28.7, 28.0, 27.1, 21.8, 20.7, 14.5, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{50}N_7O_7$, 704.38; found, 704.46.

Synthesis of BJP-06-121-3



(S)-2-((S)-3-(1H-indol-3-yl)-2-((S)-1-(methyl-D-phenylalanyl)piperidine-2-carboxamido)propanamido)-5-ureidopentanoic acid (30)

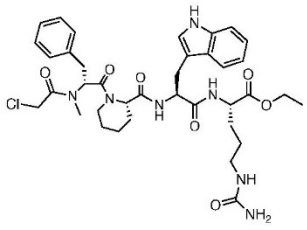
(**30**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol of Wang resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-D-N-Me-Phe-OH. ^{1}H NMR (500 MHz, DMSO-d_6): δ 12.58 (s, 1H), 10.71 (s, 1H), 9.08 (s, 1H), 8.74 (s, 1H), 8.18 (d, J = 5 Hz, 1H), 7.65 (d, J = 10 Hz, 1H), 7.54 (d, J = 10 Hz, 1H), 7.29-7.22 (m, 4H), 7.12 (d, J = 10 Hz, 2H), 7.07 (s, 1H), 6.98 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.88 (s, 1H), 5.35 (s, 2H), 4.81 (d, J = 5 Hz, 1H), 4.66-4.62 (m, 1H), 4.57-4.51 (m, 1H), 4.13-4.09 (m, 1H), 3.26-3.18 (m, 1H), 3.15-3.07 (m, 2H), 2.90-2.78 (m, 5H), 2.38 (s, 3H), 1.99 (d, J = 10 Hz, 1H), 1.67-1.61 (m, 1H), 1.53-1.43 (m, 1H), 1.34-1.25 (m, 2H), 1.19-1.14 (m, 2H), 1.09-1.02 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 171.9, 169.8, 167.9, 159.2, 136.5, 134.5, 130.0, 129.2, 127.9, 127.8, 124.1, 121.3, 118.9, 118.7, 111.7, 110.3, 71.6, 57.9, 53.7, 53.1, 52.3, 51.7, 43.7, 36.8, 31.7, 28.9, 27.9, 27.1, 26.7, 23.9, 21.1, 19.7; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{33}H_{44}N_7O_6$, 634.33; found, 634.08.



ethyl (S)-2-((S)-3-(1H-indol-3-yl)-2-((S)-1-(methyl-D-phenylalanyl)piperidine-2-carboxamido)propanamido)-5-ureidopentanoate (31)

(**31**) was synthesized using the same procedure as (**29**) from (**30**) (51 mg, 0.08 mmol) and thionyl chloride (0.02 mL, 0.28 mmol) in ethanol (0.5 mL). (**31**) was obtained as a white powder (30 mg, 56%).

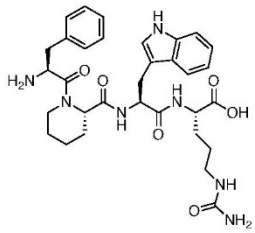
^{1}H NMR (500 MHz, DMSO-d_6): δ 10.71 (s, 1H), 9.06 (s, 1H), 8.73-8.71 (m, 1H), 8.33 (d, J = 5 Hz, 1H), 7.67 (d, J = 5 Hz, 1H), 7.52 (d, J = 5 Hz, 1H), 7.27 (d, J = 5 Hz, 1H), 7.24 (d, J = 5 Hz, 1H), 7.22 (d, J = 5 Hz, 2H), 7.11-7.09 (m, 2H), 7.04-7.02 (m, 1H), 6.96 (t, J = 5 Hz, 1H), 6.88 (t, J = 5 Hz, 1H), 5.85 (s, 1H), 5.30 (s, 1H), 4.79 (d, J = 5 Hz, 1H), 4.65-4.61 (m, 1H), 4.54-4.50 (m, 1H), 4.11 (q, J = 5 Hz, 2H), 4.00-3.95 (m, 2H), 3.50 (s, 1H), 3.18 (d, J = 10 Hz, 1H), 3.10-3.03 (m, 2H), 2.86-2.84 (m, 4H), 2.79-2.74 (m, 2H), 1.97 (d, J = 10 Hz, 1H), 1.63-1.57 (m, 1H), 1.51-1.44 (m, 2H), 1.31-1.22 (m, 2H), 1.15-1.12 (m, 2H), 1.07 (t, J = 5 Hz, 3H), 0.97-0.92 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.3, 172.0, 169.8, 167.9, 159.2, 158.6, 158.3, 158.0, 136.5, 134.5, 130.1, 129.2, 127.9, 127.8, 124.1, 121.3, 118.9, 118.7, 111.7, 110.3, 60.9, 57.9, 53.6, 53.1, 52.5, 43.7, 36.8, 31.7, 28.7, 28.0, 27.0, 26.8, 23.9, 19.7, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{35}H_{48}N_7O_6$, 662.37; found, 662.09.



ethyl (S)-2-((S)-2-((S)-1-(N-(2-chloroacetyl)-N-methyl-D-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (BJP-06-121-3) (7)

BJP-06-121-3 was synthesized using the same procedure as **BJP-02-038-2** from (**31**) (53 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. **BJP-06-121-3** was obtained as a white powder (35 mg, 59%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.75 (s, 1H), 8.32 (d, J = 5 Hz, 1H), 7.53 (d, J = 5 Hz, 1H), 7.50 (d, J = 5 Hz, 1H), 7.25 (d, J = 5 Hz, 1H), 7.28 (d, J = 5 Hz, 1H), 7.16 (d, J = 5 Hz, 2H), 7.13 (d, J = 5 Hz, 2H), 7.10-7.05 (m, 2H), 6.98 (t, J = 10 Hz, 1H), 6.90 (t, J = 10 Hz, 1H), 5.90 (s, 1H), 5.36 (t, J = 5 Hz, 1H), 4.86 (d, J = 5 Hz, 1H), 4.60-4.55 (m, 1H), 4.20 (d, J = 10 Hz, 2H), 4.16-4.11 (m, 2H), 4.02 (q, J = 5 Hz, 2H), 3.37-3.34 (m, 1H), 3.09-3.03 (m, 2H), 2.91-2.84 (m, 4H), 2.81 (s, 3H), 2.74-2.70 (m, 1H), 2.02-1.98 (m, 1H), 1.67-1.60 (m, 1H), 1.56-1.50 (m, 1H), 1.33-1.25 (m, 5H), 1.11 (t, J = 5 Hz, 3H), 0.99-0.92 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.3, 172.1, 170.5, 169.5, 166.4, 159.2, 137.9, 136.5, 129.8, 129.4, 128.6, 128.5, 127.8, 126.8, 124.0, 121.3, 118.9, 118.6, 115.6, 111.7, 110.3, 60.9, 55.0, 53.3, 52.6, 52.4, 42.5, 35.1, 30.9, 28.7, 28.1, 27.9, 27.1, 26.9, 25.0, 20.1, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{49}ClN_7O_7$, 738.34; found, 738.43.

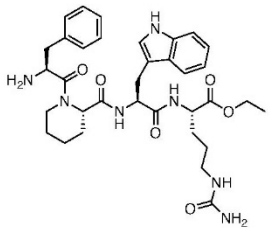
Synthesis of **8**, **9**, **10**, **11**, **12**, **13** (exploring alternate electrophiles)



(S)-2-((S)-2-((S)-1-(L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoic acid (32)

(**32**) was synthesized according to general solid phase peptide synthesis procedures starting with 0.12 mmol of Wang resin. The following amino acids were coupled to the resin: Fmoc-L-Cit-OH, Fmoc-L-Trp(BOC)-OH, Fmoc-L-Pip-OH, Fmoc-L-Phe-OH.

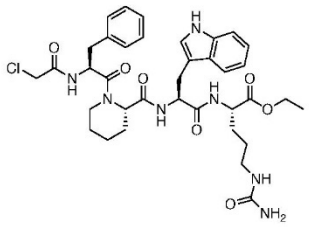
^{1}H NMR (500 MHz, DMSO-d_6): δ 12.55 (s, 1H), 10.71 (s, 1H), 8.24 (d, J = 5 Hz, 1H), 7.65 (d, J = 5 Hz, 1H), 7.56 (d, J = 5 Hz, 1H), 7.27-7.23 (m, 2H), 7.18-7.14 (m, 4H), 7.07-7.04 (m, 1H), 6.98 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.89 (s, 1H), 5.32 (s, 2H), 4.97 (d, J = 5 Hz, 1H), 4.64 (q, J = 5 Hz, 1H), 4.58 (t, J = 5 Hz, 1H), 4.26 (t, J = 5 Hz, 1H), 4.15-4.12 (m, 1H), 4.09 (d, J = 5 Hz, 1H), 3.99 (d, J = 10 Hz, 1H), 3.58 (d, J = 10 Hz, 1H), 3.11 (d, J = 5 Hz, 1H), 3.04 (t, J = 10 Hz, 1H), 2.94-2.86 (m, 3H), 2.76-2.73 (m, 1H), 2.09 (d, J = 10 Hz, 1H), 1.67-1.63 (m, 2H), 1.55-1.51 (m, 1H), 1.46-1.40 (m, 2H), 1.34-1.28 (m, 2H), 1.23-1.16 (m, 2H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 173.9, 171.9, 170.2, 169.3, 159.3, 136.5, 134.7, 130.3, 130.2, 129.2, 129.0, 127.9, 124.1, 121.3, 118.9, 118.7, 111.7, 110.3, 53.6, 52.4, 52.3, 51.3, 44.2, 35.9, 28.9, 28.1, 27.5, 27.1, 25.3, 22.7, 22.1, 20.1; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{32}H_{42}N_7O_6$, 620.32; found, 620.07.



ethyl (S)-2-((S)-2-((S)-1-(L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (33)

(**33**) was synthesized using the same procedure as (**29**) from (**32**) (49 mg, 0.08 mmol) and thionyl chloride (0.02 mL, 0.28 mmol) in ethanol (0.5 mL). (**33**) was obtained as a white powder (35 mg, 68%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.73 (s, 1H), 8.41 (d, J = 5 Hz, 1H), 8.19 (s, 1H), 7.96 (s, 2H), 7.71 (d, J = 5 Hz, 1H), 7.57 (d, J = 5 Hz, 1H), 7.28-7.23 (m, 2H), 7.18-7.15 (m, 4H), 7.05-7.03 (m, 1H), 6.98 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.88 (s, 1H), 5.32 (s, 2H), 4.98 (d, J = 5 Hz, 1H), 4.66-4.60 (m, 1H), 4.28-4.22 (m, 1H), 4.15 (t, J = 5 Hz, 1H), 4.11-4.09 (m, 1H), 4.01 (q, J = 5 Hz, 2H), 3.63-3.58 (m, 1H), 3.12 (d, J = 5 Hz, 1H), 3.09-3.04 (m, 1H), 2.93-2.86 (m, 3H), 2.77-2.72 (m, 1H), 2.09 (d, J = 10 Hz, 1H), 1.68-1.61 (m, 2H), 1.56-1.52 (m, 1H), 1.47-1.45 (m, 1H), 1.35-1.39 (m, 2H), 1.22-1.18 (m, 2H), 1.10 (t, J = 5 Hz, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.3, 172.1, 171.9, 170.3, 169.3, 168.5, 159.2, 136.5, 134.7, 130.3, 129.2, 128.9, 127.8, 127.7, 124.1, 121.3, 118.9, 118.7, 111.7, 110.3, 60.9, 53.5, 52.6, 51.3, 43.3, 35.9, 28.7, 28.2, 27.5, 27.0, 25.4, 24.3, 20.1, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{34}H_{46}N_7O_6$, 648.35; found, 648.08.

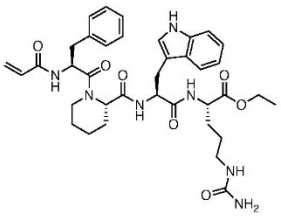


ethyl (S)-2-((S)-2-((S)-1-((2-chloroacetyl)-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (12)

12 was synthesized using the same procedure as **BJP-02-038-2** from (**33**) (52 mg, 0.08 mmol) and chloroacetyl chloride (0.012 mL, 0.16 mmol) in THF/saturated sodium bicarbonate. **12** was obtained as a white powder (38 mg, 65%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.75 (s, 1H), 8.49 (d, J = 5 Hz, 1H), 8.40 (d, J = 5 Hz, 1H), 7.55 (d, J = 5 Hz, 1H), 7.50 (t, J = 5 Hz, 1H), 7.25-7.20 (m, 2H), 7.17-7.12 (m, 4H), 7.03 (s, 1H), 6.98 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.88 (s, 1H), 5.29 (s, 2H), 4.94 (d, J = 5 Hz, 1H), 4.85-4.77 (m, 1H), 4.68-4.63 (m, 1H), 4.48-4.44 (m, 1H), 4.30-4.27 (m,

1H), 4.18-4.11 (m, 1H), 4.04-3.99 (m, 2H), 3.92 (q, J = 10 Hz, 2H), 3.55 (s, 2H), 3.13-3.08 (m, 1H), 2.93-2.85 (m, 3H), 2.76 (t, J = 5 Hz, 1H), 2.69-2.65 (m, 1H), 2.04 (d, J = 10 Hz, 1H), 1.71-1.64 (m, 2H), 1.57-1.50 (m, 1H), 1.37-1.31 (m, 2H), 1.30-1.24 (m, 2H), 1.11 (t, J = 5 Hz, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.4, 170.7, 170.3, 165.6, 159.2, 137.8, 136.5, 130.1, 129.8, 128.9, 128.6, 126.8, 124.1, 121.3, 118.9, 118.6, 111.7, 110.3, 60.9, 54.0, 53.3, 52.6, 50.8, 50.7, 43.2, 42.7, 42.7, 38.1, 37.2, 28.7, 28.1, 27.1, 26.9, 20.4, 14.5, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{36}H_{47}ClN_7O_7$, 724.32; found, 724.42.

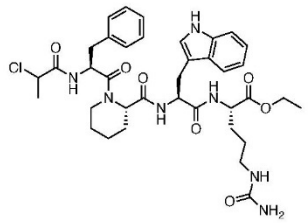


ethyl (S)-2-((S)-2-((S)-1-(acryloyl-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (8)

To a mixture of (**33**) (30 mg, 0.04 mmol) in 0.2 mL THF and 0.2 mL saturated sodium bicarbonate was added a solution of acryloyl chloride (0.0064 mL, 0.08 mmol) in 0.2 mL THF. The resulting mixture was stirred in an ice bath for 2 hours. The mixture was then diluted with 0.5 mL of dichloromethane and 0.5 mL of methanol, and filtered through a 0.45 μm syringe filter. The solvent was evaporated, the product purified via HPLC and lyophilized to yield **8** as a white powder (18 mg, 63%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.80 (s, 1H), 8.86 (d, J = 5 Hz, 1H), 8.48-8.44 (m, 1H), 8.36 (d, J = 5 Hz, 1H), 8.11 (d, J = 5 Hz, 1H), 7.59 (d, J = 5 Hz, 1H), 7.50 (dd, J = 15, 5 Hz, 1H), 7.29-7.25 (m, 2H), 7.21-7.15 (m, 4H), 7.08 (s, 1H), 7.04-6.98 (m, 1H), 6.94-6.91 (m, 1H), 6.37-6.31 (m, 1H), 6.21-6.15 (m, 1H), 6.13-6.08 (m, 1H), 5.88 (s, 1H), 5.35 (s, 2H), 4.97 (d, J = 5 Hz, 1H), 4.88-4.84 (m, 1H), 4.73-4.69 (m, 1H), 4.45-4.41 (m, 1H), 4.31

(d, J = 5 Hz, 1H), 4.20 (q, J = 5 Hz, 2H), 4.06 (t, J = 5 Hz, 2H), 3.20-3.13 (m, 1H), 3.07-3.02 (m, 1H), 2.99-2.93 (m, 2H), 2.91-2.87 (m, 1H), 2.75-2.70 (m, 1H), 2.07 (d, J = 10 Hz, 1H), 1.74-1.68 (m, 2H), 1.63-1.57 (m, 1H), 1.42-1.36 (m, 2H), 1.33-1.27 (m, 1H), 1.15 (t, J = 5 Hz, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 171.3, 164.3, 159.2, 138.3, 136.8, 136.5, 131.7, 131.0, 130.2, 129.7, 128.9, 128.6, 126.7, 124.1, 121.3, 121.3, 118.9, 118.7, 111.7, 110.6, 60.9, 56.3, 54.6, 53.2, 52.6, 52.5, 50.5, 43.2, 37.4, 28.7, 27.1, 26.9, 25.3, 24.5, 20.4, 14.5, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{48}N_7O_7$, 702.36; found, 702.51.

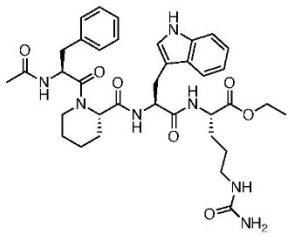


ethyl (2S)-2-((2S)-2-((2S)-1-((2-chloropropanoyl)-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (11)

To a mixture of (**33**) (30 mg, 0.04 mmol) in 0.2 mL THF and 0.2 mL saturated sodium bicarbonate was added a solution of 2-chloropropionyl chloride (0.008 mL, 0.08 mmol) in 0.2 mL THF. The resulting mixture was stirred in an ice bath for 2 hours. The mixture was then diluted with 0.5 mL of dichloromethane and 0.5 mL of methanol, and filtered through a 0.45 μm syringe filter. The solvent was evaporated, the product purified via HPLC and lyophilized to yield **11** as a white powder (15 mg, 50%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.75 (s, 1H), 8.40 (d, J = 5 Hz, 1H), 7.56 (d, J = 5 Hz, 1H), 7.50 (d, J = 5 Hz, 1H), 7.26 (d, J = 5 Hz, 1H), 7.21 (d, J = 5 Hz, 1H), 7.18 (d, J = 5 Hz, 1H), 7.13-7.11 (m, 2H), 7.05 (d, J = 5 Hz, 1H), 6.97 (t, J = 5 Hz, 1H), 6.90 (t, J = 5 Hz, 1H), 5.91 (s, 1H), 4.93 (d, J = 5 Hz, 1H), 4.84-4.77 (m, 2H), 4.67-4.64 (m, 1H), 4.53 (t, J = 5 Hz, 1H), 4.40-4.34 (m, 2H), 4.18-4.14 (m, 1H), 4.01 (q, J = 5 Hz, 2H), 3.13-3.09 (m,

1H), 2.99-2.93 (m, 1H), 2.95-2.86 (m, 4H), 2.77-2.71 (m, 2H), 2.70-2.63 (m, 1H), 2.04-2.00 (m, 1H), 1.71-1.63 (m, 2H), 1.58-1.49 (m, 2H), 1.40-1.36 (m, 1H), 1.34 (s, 3H), 1.16 (d, J = 5 Hz, 2H), 1.11 (t, J = 5 Hz, 3H), 0.97-0.88 (m, 1H), 0.83-0.75 (m, 1H); ^{13}C NMR (126 MHz, DMSO-d_6): δ 172.4, 172.2, 170.7, 170.3, 169.4, 168.6, 168.3, 159.2, 137.9, 136.5, 130.2, 129.9, 128.8, 128.5, 127.8, 126.8, 124.1, 121.3, 118.6, 111.7, 110.3, 60.9, 54.3, 53.3, 52.6, 50.7, 43.1, 37.4, 28.7, 28.1, 27.1, 25.5, 25.4, 21.9, 21.8, 20.4, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{49}ClN_7O_7$, 738.34; found, 738.48.

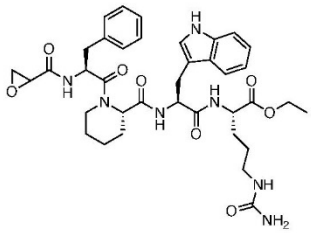


ethyl (S)-2-((S)-2-((S)-1-(acetyl-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (13)

13 was synthesized using the same procedure as **BJP-R** from (**33**) (26 mg, 0.04 mmol) and acetyl chloride (0.006 mL, 0.08 mmol) in THF/saturated sodium bicarbonate. **13** was obtained as a white powder (22 mg, 80%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.76 (s, 1H), 8.17 (d, J = 5 Hz, 1H), 8.09-8.04 (m, 1H), 7.55 (d, J = 10 Hz, 1H), 7.48 (d, J = 10 Hz, 1H), 7.25-7.21 (m, 3H), 7.19-7.12 (m, 4H), 7.03 (s, 1H), 7.00-6.96 (m, 1H), 6.92-6.87 (m, 1H), 5.86 (s, 1H), 4.93 (d, J = 5 Hz, 1H), 4.82-4.78 (m, 1H), 4.73-4.64 (m, 1H), 4.47-4.33 (m, 1H), 4.28 (d, J = 5 Hz, 1H), 4.19-4.12 (m, 2H), 4.05-3.99 (m, 3H), 3.14-3.09 (m, 1H), 2.95-2.83 (m, 4H), 2.68-2.62 (m, 1H), 1.76 (s, 1H), 1.64 (s, 3H), 1.56-1.47 (m, 1H), 1.38-1.34 (m, 1H), 1.29-1.23 (m, 2H), 1.14-1.10 (m, 3H), 1.06-1.00 (m, 2H), 0.88-0.82 (m, 1H), 0.77-0.72 (m, 1H); ^{13}C NMR (126 MHz,

DMSO-d_6): δ 172.2, 171.5, 170.5, 159.2, 138.4, 136.6, 130.1, 129.7, 128.9, 128.5, 127.6, 127.2, 126.7, 124.2, 121.3, 118.9, 118.8, 118.7, 111.7, 110.4, 60.9, 56.3, 54.3, 53.2, 52.6, 50.3, 37.9, 37.5, 28.7, 27.7, 27.1, 25.3, 22.7, 22.5, 20.4, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{36}H_{48}N_7O_7$, 690.36; found, 690.50.

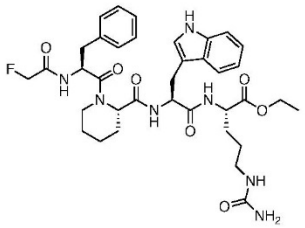


ethyl (2S)-2-((2S)-3-(1H-indol-3-yl)-2-((2S)-1-((oxirane-2-carbonyl)-L-phenylalanyl)piperidine-2-carboxamido)propanamido)-5-ureidopentanoate (10)

To a mixture of (**33**) (42 mg, 0.065 mmol), HATU (49 mg, 0.13 mmol), and potassium oxirane-2-carboxylate (9.8 mg, 0.078 mmol) in DCM (1 mL) was added N,N-Diisopropylethylamine (0.057 mL, 0.33 mmol). The reaction was stirred at room temperature for 3 hrs. The product was then extracted three times with 5 mL of 4:1 chloroform:isopropanol. After washing with water (5 mL) and then brine (5 mL), the solvent was evaporated. The product was isolated via HPLC purification and lyophilized to yield **10** as a white powder (16 mg, 34%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.82 (s, 1H), 8.47 (d, J = 5 Hz, 1H), 7.62 (d, J = 5 Hz, 1H), 7.58-7.54 (m, 1H), 7.32-7.27 (m, 2H), 7.24-7.20 (m, 4H), 7.17-7.15 (m, 1H), 7.12 (s, 1H), 7.06 (t, J = 5 Hz, 1H), 6.97 (t, J = 5 Hz, 1H), 5.97 (s, 1H), 5.01 (d, J = 5 Hz, 1H), 4.75-4.71 (m, 1H), 4.26-4.20 (m, 2H), 4.11 (q, J = 5 Hz, 2H), 3.35-3.31 (m, 1H), 3.19-3.15 (m, 1H), 3.01-2.93 (m, 3H), 2.87-2.85 (m, 1H), 2.81-2.78 (m, 2H), 2.11 (d, J = 5 Hz, 1H), 1.76-1.70 (m, 1H), 1.64-1.58 (m, 1H), 1.43-1.34 (m, 2H), 1.18 (t, J = 5 Hz, 3H); ^{13}C NMR

(126 MHz, DMSO-d_6): δ 172.4, 172.2, 170.3, 159.2, 136.5, 130.2, 129.9, 129.8, 128.8, 128.6, 127.8, 124.1, 121.3, 118.6, 111.7, 110.4, 110.3, 60.9, 53.3, 52.6, 50.5, 50.1, 48.6, 48.3, 45.9, 45.7, 40.9, 36.9, 28.7, 28.1, 27.1, 25.3, 21.9, 20.4, 20.3, 14.5, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{37}H_{48}N_7O_8$, 718.36; found, 718.47.



ethyl (S)-2-((S)-2-((S)-1-((2-fluoroacetyl)-L-phenylalanyl)piperidine-2-carboxamido)-3-(1H-indol-3-yl)propanamido)-5-ureidopentanoate (9)

To a solution of (**33**) (52 mg, 0.08 mmol) in DCM (0.5 mL), was added HATU (59 mg, 0.16 mmol) and potassium fluoroacetate (10 mg, 0.09 mmol), followed by N,N-Diisopropylethylamine (0.06 mL, 0.4 mmol). The resulting reaction mixture was stirred at room temperature for 4 hours. The product was then extracted three times with 5 mL of 4:1 chloroform:isopropanol. After washing with water (5 mL) and then brine (5 mL), the solvent was evaporated. The product was isolated via HPLC purification and lyophilized to yield **9** as a white powder (35 mg, 62%).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.82 (s, 1H), 8.47 (d, J = 5 Hz, 1H), 8.32 (d, J = 5 Hz, 1H), 7.62-7.58 (m, 2H), 7.32-7.27 (m, 2H), 7.23-7.17 (m, 4H), 7.11 (s, 1H), 7.05 (t, J = 5 Hz, 1H), 6.97 (t, J = 5 Hz, 1H), 5.98 (s, 1H), 5.00 (d, J = 5 Hz, 1H), 4.95-4.91 (m, 1H), 4.89-4.84 (m, 1H), 4.74 (s, 2H), 4.64 (s, 1H), 4.58-4.53 (m, 1H), 4.38 (d, J = 5 Hz, 1H), 4.24-4.19 (m, 2H), 4.10 (q, J = 5 Hz, 2H), 3.19-3.17 (m, 1H), 3.02-2.93 (m, 3H), 2.87-2.81 (m, 2H), 2.10 (d, J = 5 Hz, 1H), 1.85-1.80 (m, 1H), 1.76-1.69 (m, 1H), 1.65-1.57 (m, 2H), 1.46-1.41 (m, 2H), 1.38-1.34 (m, 2H), 1.18 (t, J = 5 Hz, 3H), 0.99-0.87 (m, 1H); ^{13}C NMR

(126 MHz, DMSO-d_6): δ 172.4, 172.2, 170.7, 170.4, 159.3, 137.9, 136.5, 130.1, 129.8, 128.8, 128.6, 127.8, 126.8, 124.1, 121.3, 118.9, 118.7, 111.7, 110.3, 80.7, 79.3, 60.9, 53.3, 52.6, 50.2, 50.0, 40.8, 39.2, 36.9, 28.7, 28.1, 27.0, 25.3, 20.4, 14.5, 14.5; MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{36}H_{47}FN_7O_7$, 708.35; found, 708.41.

---

[i] Wu, X. et al. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.* **19**, 3727-3738 (2000).
[ii] Kuleshov, M.V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44** (Web Server issue), W90-W97. https://doi.org/10.1093/nar/gkw377.
[iii] Kim, Y.W., Grossmann, T.N., & Verdine, G.L. Synthesis of all-hydrocarbon stapled alpha-helix peptides by ring-closing olefin metathesis. *Nature Protocols* **6** (6): 761-771 (2011).

7.3 References

[1] Browne, C. M.; Jiang, B.; Ficarro, S. B.; Doctor, Z. M.; Johnson, J. L.; Card, J. D.; Sivakumaren, S. C.; Alexander, W. M.; Yaron, T. M.; Murphy, C. J.; Kwiatkowski, N. P.; Zhang, T.; Cantley, L. C.; Gray, N. S.; Marto, J. A. A Chemoproteomic Strategy for Direct and Proteome-Wide Covalent Inhibitor Target-Site Identification. *J. Am. Chem. Soc.* **2019**, *141* (1), 191-203.

[2] Schiene-Fischer, C. Multidomain Peptidyl Prolyl *cis*/*trans* Isomerases. *Biochim. Biophys. Acta* **2015**, *1850* (10), 2005-2016.

[3] Matena, A.; Rehic, E.; Hönig, D.; Kamba, B.; Bayer, P. Structure and function of the human parvulins Pin1 and Par14/17. *Biol. Chem.* **2018**, *399* (2), 101-125.

[4] Lu, P. K.; Hanes, S. D.; Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **1996**, *380* (6574), 544-547.

[5] Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell* **1997**, *89* (6), 875-886.

[6] Olsson, S.; Strotz, D.; Vögeli, B.; Riek, R.; Cavalli, A. The Dynamic Basis for Signal Propagation in Human Pin1-WW. *Structure* **2016**, *24* (9), 1464-1475.

[7] Chen, D.; Wang, L.; Lee, T. H. Post-translational Modifications of the Peptidyl-Prolyl Isomerase Pin1. *Front. Cell Dev. Biol.* **2020**, *8*, 129.

[8] Lu, K. P.; Zhou, X. Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (11), 904-916.

[9] Litchfield, D. W.; Shilton, B. H.; Brandl, C. J.; Gyenis, L. Pin1: Intimate involvement with the regulatory protein kinase networks in the global phosphorylation landscape. *Biochim. Biophys. Acta* **2015**, *1850* (10), 2077-2086.

[10] Lin, C.-H.; Li, H.-Y.; Lee, Y.-C.; Calkins, M. J.; Lee, K.-H.; Yang, C.-N.; Lu, P.-J. Landscape of Pin1 in the cell cycle. *Exp. Biol. Med.* **2015**, *240* (3), 403-408.

[11] Hanes, S. D. Prolyl isomerases in gene transcription. *Biochim. Biophys. Acta* **2015**, *1850* (10), 2017-2034.

[12] Hu, X.; Chen, L.-F. Pinning Down the Transcription: A Role for Peptidyl-Prolyl *cis-trans* Isomerase Pin1 in Gene Expression. *Front. Cell Dev. Biol.* **2020**, *8*, 179.

[13] Thapar, R. Roles of Prolyl Isomerases in RNA-Mediated Gene Expression. *Biomolecules* **2015**, *5* (2), 974-999.

[14] Nakatsu, Y.; Matsunaga, Y.; Yamamotoya, T.; Ueda, K.; Inoue, Y.; Mori, K.; Sakoda, H.; Fujishiro, M.; Ono, H.; Kushiyama, A.; Asano, T. Physiological and Pathogenic Roles of Prolyl isomerase Pin1 in Metabolic Regulations via Multiple Signal Transduction Pathway Modulations. *Int. J. Mol. Sci.* **2016**, *17* (9), 1495.

[15] Nakatsu, Y.; Mori, K.; Matsunaga, Y.; Yamamotoya, T.; Ueda, K.; Inoue, Y.; Mitsuzaki-Miyoshi, K.; Sakoda, H.; Fujishiro, M.; Yamaguchi, S.; Kushiyama, A.; Ono, H.; Ishihara, H.; Asano, T. The prolyl isomerase Pin1 increases β-cell proliferation and enhances insulin secretion. *J. Biol. Chem.* **2017**, *292* (28), 11886-11895.

[16] Nakatsu, Y.; Matsunaga, Y.; Yamamotoya, T.; Ueda, K.; Inoue, M.-k.; Mizuno, Y.; Nakanishi, M.; Sano, T.; Yamawaki, Y.; Kushiyama, A.; Sakoda, H.; Fujishiro, M.; Ryo, A.; Ono, H.; Minamino, T.; Takahashi, S.-I.; Ohno, H.; Yoneda, M.; Takahashi, K.; Ishihara, H.; Katagiri, H.; Nishimura, F.; Kanematsu, T.; Yamada, T.; Asano, T. Prolyl Isomerase Pin1 Suppresses Thermogenic Programs in Adipocytes by Promoting Degradation of Transcriptional Co-activator PRDM16. *Cell Rep.* **2019**, *26* (12), 3221-3230.e3.

[17] Islam, R.; Yoon, H.; Shin, H.-R.; Bae, H.-S.; Kim, B.-S.; Yoon, W.-J.; Woo, K.-M.; Baek, J.-H.; Lee, Y.-S.; Ryoo, H.-M. Peptidyl-prolyl cis-trans isomerase NIMA interacting 1 regulates skeletal muscle fusion through structural modification of Smad3 in the linker region. *J. Cell. Physiol.* **2018**, *233* (12), 9390-9403.

[18] Islam, R.; Yoon, W.-J.; Ryoo, H.-M. Pin1, the Master Orchestrator of Bone Cell Differentiation. *J. Cell. Physiol.* **2017**, *232* (9), 2339-2347.

[19] Kurita-Suzuki, A.; Kamo, Y.; Uchida, C.; Tanemura, K.; Hara, K.; Uchida, T. Prolyl isomerase Pin1 is required sperm production by promoting mitosis progression of spermatogonial stem cells. *Biochem. Biophys. Res. Commun.* **2018**, *497* (1), 388-393.

[20] Hariharan, N.; Sussman, M. A. Pin1: A Molecular Orchestrator in the Heart. *Trends Cardiovasc. Med.* **2014**, *24* (6), 256-262.

[21] Wang, J.-Z.; Liu, G.-J.; Li, Z.-Y.; Wang, X.-H. Pin1 in cardiovascular dysfunction: A potential double-edge role. *Int. J. Cardiol.* **2016**, *212*, 280-283.

[22] Liu, M.; Yu, P.; Jiang, H.; Yang, X.; Zhao, J.; Zou, Y.; Ge, J. The Essential Role of Pin1 via NF-κB Signaling in Vascular Inflammation and Atherosclerosis in ApoE$^{-/-}$ Mice. *Int. J. Mol. Sci.* **2017**, *18* (3), 644.

[23] Inoue, M.-K.; Nakatsu, Y.; Yamamotoya, T.; Hasei, S.; Kanamoto, M.; Naitou, M.; Matsunaga, Y.; Sakoda, H.; Fujishiro, M.; Ono, H.; Kushiyama, A.; Asano, T. Pin1 Plays Essential Roles in NASH Development by Modulating Multiple Target Proteins. *Cells* **2019**, *8* (12) 1545.

[24] Wang, L.; Zhou, Y.; Chen, D.; Lee, T. H. Peptidyl-Prolyl *Cis*/*Trans* Isomerase Pin1 an Alzheimer's Disease. *Front. Cell Dev. Biol.* **2020**, *8*, 355.

[25] Barberi, T. J.; Dunkle, A.; He, Y.-W.; Racioppi, L.; Means, A. R. The Prolyl Isomerase Pin1 Modulates Development of CD8+ cDC in Mice. *PLoS One* **2012**, *7* (1), e29808.

[26] Wei, S.; Yoshida, N.; Finn, G.; Kozono, S.; Nechama, M.; Kyttaris, V. C.; Zhou, X. Z.; Tsokos, G. C.; Lu, K. P. Pin1-Targeted Therapy for Systemic Lupus Erythematosus. *Arthritis Rheumatol.* **2016**, *68* (10), 2503-2513.

[27] Shen, Z.-J.; Malter, J. S. Eosinophils, Pin1 and the Response to Respiratory Viral Infection and Allergic Stimuli. *Crit. Rev. Immunol.* **2019**, *39* (2), 135-149.

[28] Tanaka, Y.; Amano, A.; Morisaki, M.; Sato, Y.; Sasaki, T. Cellular peptidyl-prolyl *cis*/*trans* isomerase Pin1 facilitates replication of feline coronavirus. *Antiviral Res.* **2016**, *126*, 1-7.

[29] Schütz, M.; Thomas, M.; Wangen, C.; Wagner, S.; Rauschert, L.; Errerd, T.; Kießling, M.; Sticht, H.; Milbradt, J.; Marschall, M. The peptidyl-prolyl cis/trans isomerase Pin1 interacts with three early regulatory proteins of human cytomegalovirus. *Virus Res.* **2020**, *285*, 198023.

[30] Nishi, M.; Miyakawa, K.; Matsunaga, S.; Khatun, H.; Yamaoka, Y.; Watashi, K.; Sugiyama, M.; Kimura, H.; Wakita, T.; Ryo, A. Prolyl Isomerase Pin1 Regulates the Stability of Hepatitis B Virus Core Protein. *Front. Cell Dev. Biol.* **2020**, *8*, 26.

[31] Liu, M.; Bedouhene, S.; Hurtado-Nedelec, M.; Pintard, C.; Dang, P. M-c.; Yu, S.; El-Benna, J. The Prolyl Isomerase Pin1 Controls Lipopolysaccharide-Induced Priming of NADPH Oxidase in Human Neutrophils. *Front. Immunol.* **2019**, *10*, 2567.

[32] Knowlson, C.; Haddock, P.; Bingham, V.; McQuaid, S.; Mullan, P. B.; Buckley, N. E. Pin1 plays a key role in the response to treatment and clinical outcome in triple negative breast cancer. *Ther. Adv. Med. Oncol.* **2020**, *12*, 1758835920906047.

[33] Kim, G.; Bhattarai, P. Y.; Choi, H. S. Peptidyl-prolyl cis/trans isomerase NIMA-interacting 1 as a molecular target in breast cancer: a therapeutic perspective of gynecological cancer. *Arch. Pharm. Res.* **2019**, *42* (2), 128-139.

[34] Rustighi, A.; Zannini, A.; Campaner, E.; Ciani, Y.; Piazza, S.; Del Sal, G. PIN1 in breast development and cancer: a clinical perspective. *Cell Death Differ.* **2017**, *24* (2), 200-211.

[35] Ma, J.-Q.; Yang, Y.; Juan, J.; Guo, C.-F.; Tuerxun, M.; Ting, W.; Hasim, A. Over-expression of prolyl isomerase Pin1 promotes cervical tumorigenesis and metastasis. *Int. J. Clin. Exp. Pathol.* **2018**, *11* (2), 664-674.

[36] Spena, C. R.; De Stefano, L.; Poli, G.; Granchi, C.; El Boustani, M.; Ecca, F.; Grassi, G.; Grassi, M.; Canzonieri, V.; Giordano, A.; Tuccinardi, T.; Caligiuri, I.; Rizzolio, F. Virtual screening identifies a PIN1 inhibitor with possible antiovarian cancer effects. *J. Cell. Physiol.* **2019**; DOI: 10.1002/jcp.28224.

[37] Wu, K.-J.; Zhong, H.-J.; Yang, G.; Wu, C.; Huang, J.-M.; Li, G.; Ma, D.-L.; Leung, C.-H. Small Molecule Pin1 Inhibitor Blocking NF-κB Signaling in Prostate Cancer Cells. *Chem. Asian J.* **2018**, *13* (3), 275-279.

[38] Xu, T.; Zhang, H.; Park, S.-S.; Venneti, S.; Kuick, R.; Ha, K.; Michael, L. E.; Santi, M.; Uchida, C.; Uchida, T.; Srinivasan, A.; Olson, J. M.; Dlugosz, A. A.; Camelo-Piragua, S.; Rual, J.-F. Loss of *Pin1* Suppresses Hedgehog-Driven Medulloblastoma Tumorigenesis. *Neoplasia* **2017**, *19* (3), 216-225.

[39] Stifani, S. The Multiple Roles of Peptidyl Prolyl Isomerases in Brain Cancer. *Biomolecules* **2018**, *8* (4), 112.

[40] Zhang, Z.-Z.; Yu, W.-X.; Zheng, M.; Liao, X.-H.; Wang, J.-C.; Yang, D.-Y.; Lu, W.-X.; Wang, L.; Zhang, S.; Liu, H.-K.; Zhou, X. Z.; Lu, K. P. PIN1 Inhibition Sensitizes Chemotherapy in Gastric Cancer Cells by Targeting Stem Cell-like Traits and Multiple Biomarkers. *Mol. Cancer Ther.* **2020**, *19* (3), 906-919.

[41] Pyo, J.-S.; Son, B. K.; Oh, I. H. Cytoplasmic Pin1 expression is correlated with poor prognosis in colorectal cancer. *Pathol. Res. Pract.* **2018**, *214* (11), 1848-1853.

[42] Liang, C.; Shi, S.; Liu, M.; Qin, Y.; Meng, Q.; Hua, J.; Ji, S.; Zhang, Y.; Yang, J.; Xu, J.; Ni, Q.; Li, M.; Yu, X. PIN1 Maintains Redox Balance via the c-Myc/NRF2 Axis to Counteract Kras-Induced Mitochondrial Respiratory Injury in Pancreatic Cancer Cells. *Cancer Res.* **2019**, *79* (1), 133-145.

[43] Sun, Q.; Fan, G.; Zhuo, Q.; Dai, W.; Ye, Z.; Ji, S.; Xu, W.; Liu, W.; Hu, Q.; Zhang, Z.; Liu, M.; Yu, X.; Xu, X.; Qin, Y. Pin1 promotes pancreatic cancer progression and metastasis by activation of NF-κB-IL-18 feedback loop. *Cell Prolif.* **2020**, *53* (5), e12816.

[44] Kim, G.; Kim, J. Y.; Choi, H. S. Peptidyl-Prolyl *cis*/*trans* Isomerase NIMA-Interacting 1 as a Therapeutic Target in Hepatocellular Carcinoma. *Biol. Pharm. Bull.* **2015**, *38* (7), 975-979.

[45] Bae, J. S.; Noh, S. J.; Kim, K. M.; Jang, K. Y.; Park, H. S.; Chung, M. J.; Park, B.-H.; Moon, W. S. PIN1 in hepatocellular carcinoma is associated with *TP53* gene status. *Oncol. Rep.* **2016**, *36* (4), 2405-2411.

[46] D'Artista, L.; Bisso, A.; Piontini, A.; Doni, M.; Verrecchia, A.; Kress, T. R.; Morelli, M. J.; Del Sal, G.; Amati, B.; Campaner, S. Pin1 is required for sustained B cell proliferation upon oncogenic activation of Myc. *Oncotarget* **2016**, *7* (16), 21786-21798.

[47] Chen, X.; Liu, X.; Deng, B.; Martinka, M.; Zhou, Y.; Lan, X.; Cheng, Y. Cytoplasmic Pin1 expression is increased in human cutaneous melanoma and predicts poor prognosis. *Sci. Rep.* **2018**, *8* (1), 16867.

[48] Braun, C.; Schneider, N.; Pei, D.; Bosserhoff, A.; Kuphal, S. Inhibition of peptidyl-prolyl isomerase (Pin1) and BRAF signaling to target melanoma. *Am. J. Transl. Res.* **2019**, *11* (7), 4425-4437.

[49] El Boustani, M.; De Stefano, L.; Caligiuri, I.; Mouawad, N.; Granchi, C.; Canzonieri, V.; Tuccinardi, T.; Giordano, A.; Rizzolio, F. A Guide to PIN1 Function and Mutations Across Cancers. *Front. Pharmacol.* **2019**, *9*, 1477.

[50] Lu, Z.; Hunter, T. Prolyl isomerase Pin1 in cancer. *Cell Res.* **2014**, *24* (9), 1033-1049.

[51] Zhou, X. Z.; Lu, K. P. The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nat. Rev. Cancer* **2016**, *16* (7), 463-478.

[52] Pu, W.; Zheng, Y.; Peng, Y. Prolyl Isomerase Pin1 in Human Cancer: Function, Mechanism, and Significance. *Front. Cell Dev. Biol.* **2020**, *8*, 168.

[53] Zannini, A.; Rustighi, A.; Campaner, E.; Del Sal, G. Oncogenic Hijacking of the PIN1 Signaling Network. *Front. Oncol.* **2019**, *9*, 94.

[54] Yu, J. H.; Im, C. Y.; Min, S.-H. Function of PIN1 in Cancer Development and Its Inhibitors as Cancer Therapeutics. *Front. Cell Dev. Biol.* **2020**, *8*, 120.

[55] Hennig, L.; Christner, C.; Kipping, M.; Schelbert, B.; Rücknagel, K. P.; Grabley, S.; Küllertz, G.; Fischer, G. Selective Inactivation of Parvulin-Like Peptidyl-Prolyl *cis*/*trans* Isomerases by Juglone. *Biochemistry* **1998**, *37* (17), 5953-5960.

[56] Guo, C.; Hou, X.; Dong, L.; Dagostino, E.; Greasley, S.; Ferre, R.; Marakovits, J.; Johnson, M. C.; Matthews, D.; Mroczkowski, B.; Parge, H.; VanArsdale, T.; Popoff, I.; Piraino, J.; Margosiak, S.; Thomson, J.; Los, G.; Murray, B. W. Structure-based design of novel human Pin1 inhibitors (I). *Bioorg. Med. Chem. Lett.* **2009**, *19* (19), 5613-5616.

[57] Wildemann, D.; Erdmann, F.; Alvarez, B. H.; Stoller, G.; Zhou, X. Z.; Fanghänel, J.; Schutkowski, M.; Lu, K. P.; Fischer G. Nanomolar Inhibitors of the Peptidyl Prolyl Cis/Trans Isomerase Pin1 from Combinatorial Peptide Libraries. *J. Med. Chem.* **2006**, *49* (7), 2147-2150.

[58] Liu, T.; Liu, Y.; Kao, H.-Y.; Pei, D. Cyclic Peptidyl Inhibitors against Human Peptidyl-Prolyl Isomerase Pin1. *J. Med. Chem.* **2010**, *53* (6), 2494-2501.

[59] Moore, J. D.; Potter, A. Pin1 inhibitors: Pitfalls, progress and cellular pharmacology. *Bioorg. Med. Chem. Lett.* **2013**, *23* (15), 4283-4291.

[60] Wei, S.; Kozono, S.; Kats, L.; Nechama, M.; Li, W.; Guarnerio, J.; Luo, M.; You, M.-H.; Yao, Y.; Kondo, A.; Hu, H.; Bozkurt, G.; Moerke, N. J.; Cao, S.; Reschke, M.; Chen, C.-H.; Rego, E. M.; LoCoco, F.; Cantley, L.; Lee, T. H.; Wu, H.; Zhang, Y.; Pandolfi, P. P.; Zhou, X. Z.; Lu, K. P. Active Pin1 is a key target of all-*trans* retinoic acid in acute promyelocytic leukemia and breast cancer. *Nat. Med.* **2015**, *21* (5), 457-466.

[61] Kozono, S.; Lin, Y.-M.; Seo, H.-S.; Pinch, B.; Lian, X.; Qiu, C.; Herbert, M. K.; Chen, C.-H.; Tan, L.; Gao, Z. J.; Massefski, W.; Doctor, Z. M.; Jackson, B. P.; Chen, Y.; Dhe-Paganon, S.; Lu, K. P.; Zhou, X. Z. Arsenic targets Pin1 and cooperates with retinoic acid to inhibit cancer-driving pathways and tumor-initiating cells. *Nat. Commun.* **2018**, *9* (1), 3069.

[62] Campaner, E.; Rustighi, A.; Zannini, A.; Cristiani, A.; Piazza, S.; Ciani, Y.; Kalid, O.; Golan, G.; Baloglu, E.; Shacham, S.; Valsasina, B.; Cucchi, U.; Pippione, A. C.; Lolli, M. L.; Giabbai, B.; Storici, P.; Carloni, P.; Rossetti, G.; Benvenuti, F.; Bello, E.; D'Incalci, M.; Cappuzzello, E.; Rosato, A.; Del Sal, G. A covalent PIN1 inhibitor selectively targets cancer cells by a dual mechanism of action. *Nat. Commun.* **2017**, *8*, 15772.

[63] Fan, X.; He, H.; Li, J.; Luo, G.; Zheng, Y.; Zhou, J.-K.; He, J.; Pu, W.; Zhao, Y. Discovery of 4,6-bis(benzyloxy)-3-phenylbenzofuran as a novel Pin1 inhibitor to suppress hepatocellular carcinoma via upregulating microRNA biogenesis. *Bioorg. Med. Chem.* **2019**, *27* (11), 2235-2244.

[64] Wu, K.-J.; Liu, X.; Wong, S.-Y.; Zhou, Y.; Ma, D.-L.; Leung, C.-H. Synthesis and Evaluation of Dibenzothiophene Analogues as Pin1 Inhibitors for Cervical Cancer Therapy. *ACS Omega* **2019**, *4* (5), 9228-9234.

8 Conclusion and Perspectives

In this thesis, the main goal was the development and characterization of chemical probes for the study of biological targets or novel protein inhibitor scaffolds. Generally, chemical probes that were investigated in the presented projects have been confirmed to potently and selectively engage their targets in biochemical and cellular assays. This ensured a high quality of the developed probes, which is an essential feature to produce reliable data.

In Chapter 5, two different scaffolds of PI5P4K inhibitors were investigated. With the development of THZ-P1-2 and its desthiobiotin analog, THZ-P1-2-DTB, a decently selective and potent PI5P4K inhibitor was presented. The desthiobiotin-labeled analog provided a valuable tool in investigating target engagement as well as selectivity within the cellular context, providing important information on the quality of THZ-P1-2 as a PI5P4K-targeting probe. While this probe was thus able to demonstrate that small molecule-based inhibition of PI5P4K can lead to autophagy defects in cells, further medicinal chemistry efforts led to increased proteome-wide selectivity of this inhibitor scaffold, resulting in the reported compound **30**. These novel compounds are thus the first potent and selective, irreversible PI5P4K inhibitors that have been reported up to this point and could be applied to further investigate the still poorly understood mode of action of this lipid kinase family and their potential as therapeutic targets in diseases such as cancer.

While the THZ-P1-2 compound covalently engages all three PI5P4K isoforms by targeting a remote cysteine on a disordered loop outside of the ATP-binding pocket, another promising, reversible PI5P4K-binding scaffold was discovered in a high-throughput screen of an in-house kinase inhibitor library. Since covalent kinase inhibitors might have off-target effects due to the reactivity of their electrophilic warheads as well as the potential to induce resistance as the result of mutations in the targeted cysteines of the protein binding site, developing a reversible PI5P4K inhibitor might pose another interesting opportunity to study the PI5P4K kinases. After identification of CVM-05-002 as the initial hit compound of the screen, thorough medicinal chemistry efforts led to improved lipophilicity of the compound, thus enhancing cellular target engagement of this scaffold. This was confirmed using an isothermal CETSA to evaluate cellular target engagement of CVM-based analogs. While this novel PI5P4K inhibitor scaffold is another promising tool to probe PI5P4K activity in cells in further studies, this SAR study provides an intriguing case of a PAINS-like screening hit, which, with careful evaluation, turned out to be a valuable starting point for the development of a small molecule protein inhibitor. Future investigations with this reversible kinase inhibitor scaffold will potentially yield important information on the effects of PI5P4K function in cells and could be used to compare with the effects seen with the covalent THZ-based scaffold.

In Chapter 6, CDK12/13-selective, covalent inhibitors were characterized. As a result of optimization efforts of the in-house CDK7/12/13 inhibitor THZ1, THZ531 was developed. With the synthesis of a biotin analog of this novel CDK12/13-selective inhibitor, it was possible to characterize the inhibitor's target engagement and selectivity in the cellular context, which again proved to be valuable in the evaluation of THZ531 as a chemical probe to target CDK12 and CDK13. After the characterization of the probe, important cellular consequences of CDK12 and CDK13 inhibition were investigated and could confirm that CDK12 and CDK13 are involved in transcription elongation of target genes of the DDR. These results provide essential information for CDK12/13 function in health

as well as in disease models of cancer. This could furthermore lead to the investigation of CDK12/13 inhibitors as potential therapeutic targets for a variety of cancerous conditions. Importantly, the presented study introduces the first CDK12/13-selective, potent inhibitors reported to that date and highlights the importance of high-quality chemical probes for the study of biological targets.

Finally, Pin1-targeting inhibitor scaffolds are investigated in Chapters 7 and 9 (Appendix). In the first part, a peptide-based, covalent Pin1 inhibitor, BJP-06-005-3, was developed and thoroughly characterized, which included the development of a desthiobiotin-labeled analog for further cellular characterization. Similarly, the second part describes the development of a fragment-based, covalent Pin1 inhibitor, sulfopin, with a respective desthiobiotin-labeled probe being used to provide valuable information regarding compound potency and selectivity in the cellular context. In addition to the previously described applications of biotin/desthiobiotin-labeled probes of covalent protein inhibitors, another application, the CITe-Id technique, was introduced, which is able to employ these labeled probes to test for proteome-wide compound selectivity of cysteine-targeting compounds. Both approaches led to the development of compounds that can be used to study Pin1 in the cellular context, while sulfopin further showed promising efficacy *in vivo*. Since Pin1 has been widely discussed in the literature, with an enormous amount of reported target proteins, probes for the study of Pin1 historically were either not potent or highly promiscuous, making it difficult to generate reliable data. The presented studies are thus introducing the first highly selective and potent Pin1 inhibitors reported to this point and provide the option to thoroughly analyze Pin1 activity *in cellulo* and *in vivo*. These studies are essential to investigate Pin1's potential use as a therapeutic target in diseases including PDAC or other cancers.

While the described studies are either already published or have been accepted for publication during review of this thesis (Chapter 9), further approaches of chemical probes in drug discovery and development have been investigated as part of this work and are presented in additional chapters of the Appendix. For instance, in Chapter 10, reversible MraY binders, the muraymycins, were used as the starting point to develop photoactive chemical probes for investigations on these natural product antibiotics. Extensive organic chemistry efforts yielded a first reference compound, which was used to confirm MraY-engagement of the designed probe, providing a valuable *proof of concept* for the design of this project. By simplifying the synthetic route to access the main core of this natural product, it would be possible to produce enough material to further investigate human and bacterial targets as well as the binding mode of muraymycins. This data would provide essential information for the elevation of these promising antibiotic agents to potential novel drug candidates.

Lastly, final Chapter 11 summarizes first efforts in the development of targeted protein degraders based on the protein inhibitor scaffolds presented in Chapters 5 and 9. While the work presented in this thesis does not include compounds that are able to degrade PI5P4K or Pin1, the described chemistry efforts and first characterizations of these compounds provide valuable information for further design strategies. For example, for PI5P4K-targeting degraders, it turned out that the disordered loop, covalently targeted by the THZ-P1-2-based scaffold, might interrupt ternary complex formation to facilitate target degradation. It would thus be interesting to design compounds with a different exit point for the thalidomide-carrying linker. Moreover, in context of fragment-based degraders

derived from sulfopin, the issue of cellular uptake of these heterobifunctional small molecules became obvious. Desired linker length and composition to increase cell-permeability of the compounds and optimal linker length to facilitate ideal ternary complex formation turned out to be incompatible. Consequently, further optimization of the linker composition or a change in the targeted E3 ligase could potentially lead to compounds with a better balance in linker properties and overall cell permeability of the degraders.

Interestingly, some of the discussed degraders turned out to have surprising targets (data not provided, ongoing projects), which are currently followed-up on. These findings highlight some advantages of targeted protein degraders, which, in contrast to conventional protein inhibitors, only require low affinity to induce effects on their target.

In summary, the presented work describes the value of chemical probes for target validation studies as well as their use in the characterization of novel chemical scaffolds, especially in the cellular context. By combining medicinal chemistry and chemical biology approaches, different kinds of chemical probes were developed and characterized. These probes further provided or have the potential to provide valuable information for future studies of various disease biology, including cancer and infectious diseases, as well as to determine the quality of developed protein modulators. While other approaches, such as genetic studies, are nonetheless powerful techniques for drug discovery, chemical probes and the steady growth of their applications provide essential tools, which complement target validation and identification studies and furnish important starting points for drug discovery and development efforts.

Part III

APPENDIX

.....

Unpublished Work

.....

9 Novel Chemical Probes for the Study of the Peptidyl-prolyl Isomerase Pin1 (continued)

9.1 Introduction

In this chapter, another covalent Pin1 inhibitor series is introduced. Since the general topic of this chapter has been already discussed in Chapter 7 of this thesis, no further introduction is provided.

9.2 Pre-Print: *Sulfopin, a selective covalent inhibitor of Pin1, blocks Myc-driven tumor initiation and growth in vivo*

9.2.1 Specific Aims

BJP-06-005-3 (Chapter 7.2) has poor mouse microsome stability and therefore, its applications are limited to biochemical and cell-based assays. Consequently, in a parallel Pin1 project, the aim was to identify novel Pin1 inhibitor scaffolds that could potentially be applied *in vivo*. As a result of this effort, sulfopin (originally named Pin1-3), was identified as a hit in a screen of an electrophilic fragment-based library. This low molecular weight compound is equipped with a chloroacetamide warhead and potently inhibits Pin1 catalytic activity *via* covalent labeling of its active site Cys113. When joining this effort, the main goal was to design, synthesize, and characterize a desthiobiotin-labeled sulfopin derivative, which should serve to further interrogate sulfopin's properties as a Pin1 inhibitor.

To this end, a desthiobiotin linker was attached to the sulfopin scaffold using a structure-guided design approach, which included optimization of the linker length, to give sulfopin-DTB. Furthermore, optimal incubation conditions for Pin1 pulldown were investigated, and ultimately, sulfopin-DTB was applied to study the selectivity of sulfopin as well as its cellular target engagement.

9.2.2 Main Outcomes and Context

In the presented study, a fragment-based, covalent Pin1 modulator, sulfopin, was discovered and characterized in the context of cancer cell line models. Mass spectrometry analysis as well as a co-crystal structure of Pin1 in complex with sulfopin confirmed covalent binding of sulfopin to Pin1 Cys113. In competitive pulldown assays using sulfopin's desthiobiotin analog (sulfopin-DTB), sulfopin demonstrated potent cellular engagement of Pin1 in several different cancer cell lines, including PATU 8988T (pancreatic cancer) and HCT116 (colon cancer) cells, as well as in spleen samples from an *in vivo* mouse model. Despite its low molecular weight, sulfopin exhibited excellent proteome-wide selectivity in two different experimental set-ups (CITe-Id with sulfopin-DTB and TOP-ABPP).

After thorough characterization of sulfopin as a Pin1 modulator, the probe was used to interrogate the phenotype of Pin1 inhibition *in cellulo* and *in vivo*. Sulfopin treatment led to phenotypes that resembled $Pin1^{-/-}$ cells, such as time-dependent antiproliferative effects in cancer cell lines, with more pronounced effects observed in 3D cell culture models. Moreover, the transcriptional effects of Pin1 inhibition were investigated, suggesting that Pin1 inhibition leads to the downregulation of MYC-regulated genes. Consequently, sulfopin treatment was interrogated in *in vivo* models (neuroblastoma models in zebrafish and mice) of MYC-driven cancers, in which it successfully blocked Pin1-mediated tumor initiation and progression.

In summary, this novel fragment-based Pin1 inhibitor is the first reported selective and potent Pin1 inhibitor that can be applied *in vivo*. Application of the desthiobiotin label to sulfopin once again elegantly demonstrated the compound's selectivity and cellular target engagement in cells and *in vivo*, highlighting the advantages and applicability of this probe-

based technique. Furthermore, the presented study provides another example of the importance of using a well-characterized probe to interrogate biological phenotypes.

The synthesis of the sulfopin-DTB probe is described in the Supplementary Information, and its characterization is shown in the main article as well as the Supplementary Information.

9.2.3 Pre-print Article

The following pre-print was originally published (not peer-reviewed) as:

Sulfopin, a selective covalent inhibitor of Pin1, blocks Myc-driven tumor initiation and growth *in vivo*

Christian Dubiella, Benika J. Pinch, Daniel Zaidman, Theresa D. Manz, Evon Poon, Shuning He, Efrat Resnick, Ellen M. Langer, Colin J. Daniel, Hyuk-Soo Seo, Ying Chen, Scott B. Ficarro, Yann Jamin, Xiaolan Lian, Shin Kibe, Shingo Kozono, Kazuhiro Koikawa, Zainab M. Doctor, Behnam Nabet, Christopher M. Browne, Annan Yang, Liat Stoler-Barak, Richa B. Shah, Nick E. Vangos, Ezekiel A. Geffken, Roni Oren, Samuel Sidi, Ziv Shulman, Chu Wang, Jarrod A. Marto, Sirano Dhe-Paganon, Thomas Look, Xiao Zhen Zhou, Kun Ping Lu, Rosalie C. Sears, Louis Chesler, Nathanael S. Gray, Nir London.

bioRxiv 2020.03.20.998443.

DOI: https://doi.org/10.1101/2020.03.20.998443

This article, including Supporting Information, was reproduced with kind permission from Dubiella and Pinch *et al.*© 2020, and was made available under a CC-BY 4.0 International license.

Note: This pre-print was submitted and has been accepted at *Nature Chemical Biology* while this thesis was reviewed.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Sulfopin, a selective covalent inhibitor of Pin1, blocks Myc-driven tumor initiation and growth *in vivo*

Christian Dubiella[1,*], Benika J. Pinch[2,3,4,*], Daniel Zaidman[1], Theresa D. Manz[2,3,5], Evon Poon[6], Shuning He[7], Efrat Resnick[1], Ellen M. Langer[8,9], Colin J. Daniel[8,9], Hyuk-Soo Seo[2], Ying Chen[10], Scott B. Ficarro[2,11,12,13], Yann Jamin[14], Xiaolan Lian[15,16,17], Shin Kibe[15,16,17], Shingo Kozono[15,16,17], Kazuhiro Koikawa[15,16,17], Zainab M. Doctor[2,3], Behnam Nabet[2,3], Christopher M. Browne[2,3,18], Annan Yang[2,19], Liat Stoler-Barak[20], Richa B. Shah[21,22], Nick E. Vangos[2], Ezekiel A. Geffken[2], Roni Oren[23], Samuel Sidi[21,22], Ziv Shulman[20], Chu Wang[10], Jarrod A. Marto[2,11,12,13], Sirano Dhe-Paganon[2], Thomas Look[7,24], Xiao Zhen Zhou[15,16,17], Kun Ping Lu[15,16,17], Rosalie C. Sears[8,9,25], Louis Chesler[6], Nathanael S. Gray[2,3,#], Nir London[1,#]

[1] Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[2] Department of Cancer Biology, Dana–Farber Cancer Institute, Boston, MA 02115, USA.
[3] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA.
[4] Department of Chemistry and Chemical Biology; Department of Chemical Biology, Harvard University, Cambridge, MA, 02138, USA.
[5] Department of Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbruecken, Germany.
[6] Division of Clinical Studies, The Institute of Cancer Research, London, UK.
[7] Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[8] Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon.
[9] Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon.
[10] College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.
[11] Department of Oncologic Pathology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[12] Blais Proteomics Center, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[13] Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA.
[14] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK.
[15] Department of Medicine, Division of Translational Therapeutics, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA.
[16] Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA.
[17] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
[18] Currently at: Discovery Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Boston, MA USA
[19] Department of Medical Oncology, Dana–Farber Cancer Institute, Boston, MA 02115, USA.
[20] Department of Immunology, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[21] Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[22] Department of Cell, Developmental and Regenerative Biology, The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[23] Department of Veterinary Resources, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[24] Division of Pediatric Hematology/Oncology Boston Children's Hospital, Boston, MA, 02115, USA.
[25] Brenden-Colson Center for Pancreatic Care, Oregon Health & Science University, Portland, Oregon.

[*] equal contribution.
[#] e-mail: nir.london@weizmann.ac.il; nathanael_gray@dfci.harvard.edu

Keywords: covalent fragments; covalent probes; Pin1; neuroblastoma; Myc

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Abstract

The peptidyl-prolyl cis-trans isomerase, Pin1, acts as a unified signaling hub that is exploited in cancer to activate oncogenes and inactivate tumor suppressors, in particular through up-regulation of c-Myc target genes. However, despite considerable efforts, Pin1 has remained an elusive drug target. Here, we screened an electrophilic fragment library to discover covalent inhibitors targeting Pin1's active site nucleophile - Cys113, leading to the development of Sulfopin, a double-digit nanomolar Pin1 inhibitor. Sulfopin is highly selective for Pin1, as validated by two independent chemoproteomics methods, achieves potent cellular and *in vivo* target engagement, and phenocopies genetic knockout of Pin1. Although Pin1 inhibition had a modest effect on viability in cancer cell cultures, Sulfopin induced downregulation of c-Myc target genes and reduced tumor initiation and tumor progression in murine and zebrafish models of MYCN-driven neuroblastoma. Our results suggest that Sulfopin is a suitable chemical probe for assessing Pin1-dependent pharmacology in cells and *in vivo*. Moreover, these studies indicate that Pin1 should be further investigated as a potential cancer target.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Introduction

Cancer relies on multiple signaling pathways to sustain proliferation and downregulate apoptotic signals [1], including the phosphorylation of serine/threonine - proline motif (pSer/Thr-Pro) [2,3] found in many cellular proteins. This motif is specifically recognized and isomerized by the peptidyl-prolyl isomerase NIMA-interacting-1 (Pin1), which is the only known phosphorylation-dependent isomerase amongst the ~30 peptidyl-prolyl cis-trans isomerases (PPIases) in the human proteome [4]. Pin1-mediated isomerization was shown to impact substrate stability [5–9], activity [10,11], subcellular localization [8], and binding to interaction partners, including Proline-directed kinases and phosphatases, which are mostly *trans*-specific [12–14]. Thus, Pin1 represents a unified signaling hub that is exploited by cancer to activate oncogenes and inactivate tumor suppressors [15,16].

Several lines of evidence suggest that aberrant Pin1 activation drives oncogenesis. Pin1 is over-expressed and/or -activated in at least 38 tumor types [17]. While elevated Pin1 expression correlates with poor clinical prognosis [18,19], polymorphisms that lower Pin1 expression are associated with reduced cancer risk [20]. Pin1 sustains proliferative signaling in cancer cells by upregulating over 50 oncogenes or growth-promoting factors [15,16] , including NF-κB [9], Notch1 [21] and c-Myc [22], while suppressing over 20 tumor suppressors or growth-inhibiting factors, such as FOXOs [23], Bcl2 [24] and RARα [25]. Furthermore, Pin1-null mice are resistant to tumorigenesis induced by mutant p53 [26], activated HER2/RAS [27], or constitutively expressed c-Myc [28]. In addition, Pin1 inhibition sensitizes cancer cells to chemotherapeutics [25,29,30], radiation therapy [31] and blocks the tumorigenesis of cancer stem cells [16,32,33]. However, as evidenced by the Cancer Dependency Map [34], Pin1 is not essential to cellular viability and Pin1-null mice are viable, though they develop premature aging phenotypes [7,35].

Taken together, these studies suggest that pharmacological inhibition of Pin1 has the potential to block multiple cancer-driving pathways simultaneously [16] and with limited toxicity. Indeed, compounds that inhibit Pin1, such as juglone [36], all-*trans* retinoic acid (ATRA)[37], arsenic trioxide (ATO)[38] and KPT-6566 [39], exhibit anti-cancer activity and have been used to investigate the role of Pin1 in oncogenesis. Nevertheless, these compounds have been shown to lack specificity and/or cell permeability thus making them unreliable as tools to interrogate the consequences of pharmacological inhibition of Pin1 *in vivo* [40–42]. We have recently developed a selective Pin1 covalent peptide inhibitor [43]. However, it too, was unsuitable for *in vivo* applications.

The active-site of Pin1 contains a nucleophilic cysteine residue (Cys113) which is suitable for the development of targeted covalent inhibitors [43–45]. Such covalent inhibitors have several

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

advantages over non-covalent inhibitors [46–48] and have been extremely successful against both traditional [49–51] and challenging targets [52–55].

To explore this strategy, we undertook a covalent Fragment-Based Drug Discovery (FBDD) screening campaign targeting Cys113 in Pin1's PPIase active site. FBDD, which focuses on low molecular weight compounds (typically < 300 Da), is a successful hit discovery approach [56–58] that has led to several drugs and chemical probes [57,59]. It offers good coverage of chemical space and a high probability of binding due to lower molecular complexity [60,61]. Covalent FBDD, which increasingly comes to the fore [62–72], combines the advantages of FBDD with the improved potency conferred by covalent bond formation.

Optimization of screening hits from this campaign, led to the development of Sulfopin, a double-digit nanomolar, highly selective Pin1 inhibitor that engages Pin1 in cells and *in vivo*. We found that Pin1 inhibition induced modest viability effects in 2D cancer cell culture only after prolonged exposure, and resulted in the downregulation of Myc-dependent target genes. In MYCN-driven models of neuroblastoma, both in zebrafish and in mice, Sulfopin significantly reduced tumor initiation as well as tumor progression. Sulfopin is therefore the first selective Pin1 inhibitor suitable for the evaluation of Pin1 biology *in vivo*, and provides evidence that Pin1 warrants further exploration as a potential anti-cancer target.

Results

A covalent fragment screen identifies Pin1 binders

We previously compiled a library of 993 electrophilic fragments featuring mildly reactive 'warheads' that can react covalently with cysteines in target proteins [62]. We screened our library against Pin1 to identify fragment leads for covalent inhibitor development. Purified catalytic domain of Pin1 was incubated with the fragment library (2 µM protein, 200 µM compound; 24 h at 4 °C), followed by intact protein liquid chromatography/mass-spectrometry (LC/MS) to identify and quantify Pin1 labeling (Fig. 1A). In total, 111 fragments irreversibly labeled Pin1 by > 50% under the assay conditions (Fig. 1B; Supp. Dataset 1). Among the 48 top hits (labeling > 75%) nine chloroacetamides shared a common cyclic sulfone core, suggesting a structure-activity relationship (SAR; Fig. 1B, see Supp. Fig. 1 & Supp. Table 1 for a full list of hits containing this motif). Given that the identified sulfone-containing hits were non-promiscuous in previous fragment screens against a diverse panel of proteins [62], we selected them for further development.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

To avoid undesired reactivity arising from the presence of an additional Michael acceptor in the 2-sulfolene fragments, we focused exclusively on the sulfolane analogs.

Fragment optimization yields potent Pin1 binding that is not driven by high warhead reactivity

To guide compound optimization, we used the covalent docking program, DOCKovalent [73]. Docking the sulfolane hits into various Pin1 structures and inspecting highly ranked poses, suggested two plausible binding modes, in which the sulfolane or a lipophilic moiety ('R' in Fig. 1B) either protruded into the hydrophobic proline-binding pocket, or interacted with a hydrophobic patch adjacent to Cys113 (Supp. Fig. 2). Both poses suggested maintaining the sulfolane, while diversifying the lipophilic moiety.

To optimize these original hits, we synthesized or purchased a total of 26 sulfolane-containing compounds, featuring a range of aliphatic, arylic, biphenylic or heterocyclic side-chains (Supp. Fig. 3). To identify high-affinity binders, we assessed their irreversible labeling of Pin1 in a second screen under more stringent conditions, using a 1:1 ratio of protein:compound, and a shorter incubation time (2 µM compound; 1 h at RT). Remarkably, 25 out of 26 of these second-generation compounds showed better labeling than the original screening hits, which showed no labeling under these new conditions (Supp. Table 2). Overall, the hits from this second screen revealed that a wide range of lipophilic moieties were tolerated. We concluded that an additional methylene group between the amide and the lipophilic side chain was crucial for Pin1 labeling, as exemplified by four matched molecular pairs that lacked this group and showed no labeling (Supp. Fig. 4). The top ten binders from the second screen (Fig. 1C,F) showed 35-65% Pin1 labeling. We next evaluated these analogs in a competitive fluorescence polarization (FP) assay using a FITC-labeled substrate mimetic peptide inhibitor [74]. Following a 14 h incubation with recombinant full-length Pin1, all analogs competed in the FP to a greater extent than juglone, an often cited Pin1 inhibitor (Fig. 1D).

Identifying and excluding overly reactive and potentially promiscuous compounds is critical in the development of covalent probes. Accordingly, we assessed the thiol reactivity of the top binders using a high-throughput assay that we previously applied to the entire fragment library [62]. We found that there was no correlation between labeling efficiency and reactivity (Fig. 1C; Pearson R = 0.003; Supp. Fig. 5). This was particularly evident when comparing Pin1-3, which has a tert-butyl side chain, with the structurally similar Pin-1-13, which has a cyclopropyl side chain. Both compounds showed similar Pin1 labeling (48% and 46%), but their reactivity varied by an order of magnitude, with Pin1-3 being dramatically less reactive. Furthermore, the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

compound with the highest degree of Pin1 labeling, Pin1-2-3, showed only median reactivity relative to the entire panel.

We have previously shown [62] that electrophiles with reactivity rate constants higher than $10^{-7}M^{-1}s^{-1}$ may exhibit non-selective cytotoxicity. We evaluated selected compounds from the top Pin1 labelers in a viability assay against IMR90 lung fibroblasts, and Pin1-3 was the only compound that did not show any toxicity up to 25 µM (Supp. Table 3). Pin1-3 has the lowest inherent reactivity of the top identified Pin1 labelers, and does not exhibit non-selective cytotoxicity, therefore showing the best balance of potency and selectivity. For this reason, we selected Pin1-3, henceforth **Sulfopin**, for further evaluation.

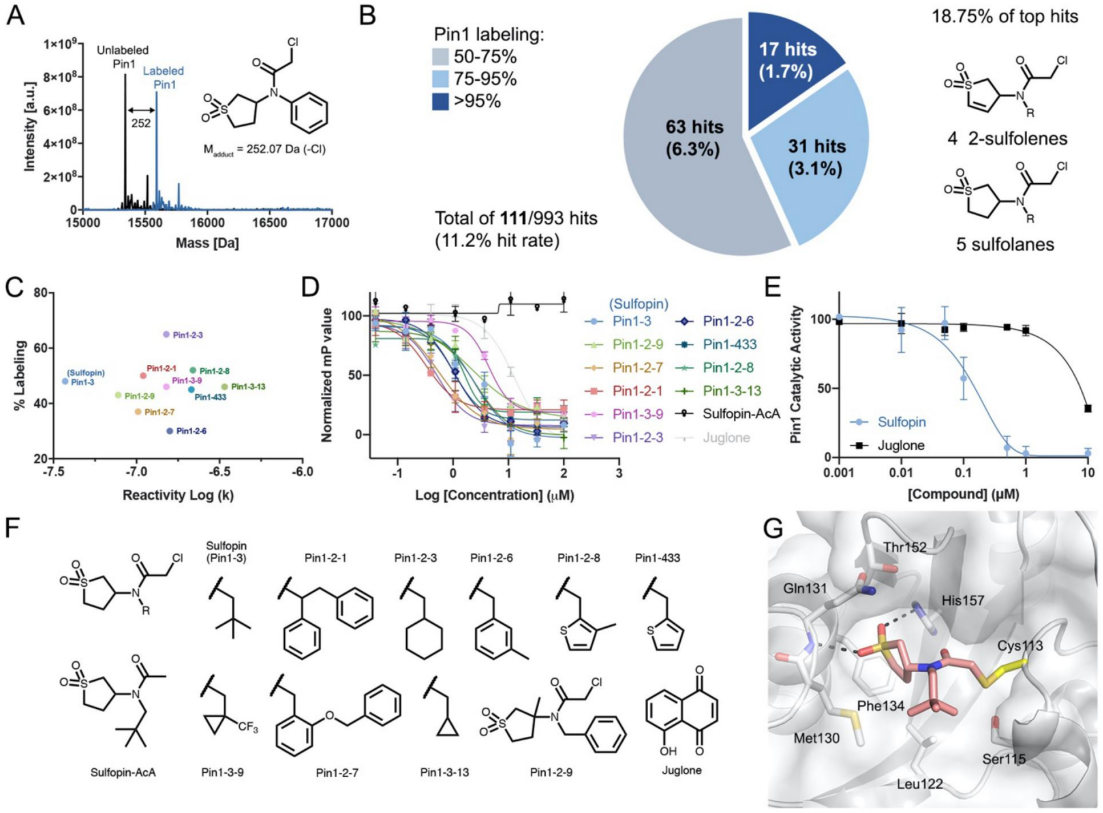


Figure 1. The discovery of a covalent Pin1 binding fragment. A. Intact protein LC/MS spectra of Pin1 (black) directly identifies covalent binders (blue) in the electrophilic library screen (200 µM compound for 24 h) **B.** Distribution of hits in the Pin1 screening campaign and their corresponding labeling in percent: Nine hits (18.75%) out of 48 top hits that labeled Pin1 > 75% (dark and light blue) share sulfolene or sulfolane moieties. **C.** 2D-analysis of the top ten optimized binders (structures in F.): Labeling % in the LC/MS-assay is plotted against reactivity (log (k)) suggests Sulfopin for further biological evaluation. **D.** Fluorescence polarization assay with Pin1 and the top ten binders including juglone and a non-reactive control (Sulfopin-AcA), after 14 h pre-incubation. See Supp. Table 3 for apparent K_i **E.** Substrate activity

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

assay of Pin1 with Sulfopin and juglone. **F.** Structures of the top ten binders in the Pin1 labeling LC/MS-assay, the non-reactive control Sulfopin-AcA and juglone. **G.** X-ray crystal structure of Pin1 in complex with Sulfopin (1.4 Å resolution, PDB code 6VAJ). Pin1 (white) with relevant side-chains in stick representation. Sulfopin is shown in pink. Hydrogen-bonds are depicted as dashed lines.

Sulfopin potently binds and inhibits Pin1

Sulfopin displayed potent Pin1 binding in the FP assay [37] with an apparent K_i = 17 nM (after 14 h; Fig. 2B). A corresponding reversible compound (Sulfopin-AcA; Fig. 1F), which lacks the chloride leaving group, was inactive in the FP assay, suggesting that the binding affinity of Sulfopin is dependent on its electrophile (Fig. 1D,2B). Sulfopin-AcA was subsequently used as a negative control compound. We then performed the FP assay in a dose- and time-dependent manner to determine the K_{inact} of Sulfopin to be 0.03 min^{-1}, with a second order rate constant (K_{inact}/K_i) of 84 $M^{-1}s^{-1}$ (Supp. Fig. 6). Sulfopin also inhibited the catalytic activity of Pin1 with an apparent K_i of 211 nM measured at 12 h, as determined using a chymotrypsin-coupled peptidyl-prolyl isomerization assay [75] (PPIase assay; Fig. 1E).

To evaluate the binding mode of Sulfopin, we determined the co-crystal structure of Pin1 in complex with Sulfopin at 1.4 Å resolution (Supp. Table 4). The complex structure shows clear electron density to Cys113 in the $2F_O$-F_C omit map, which confirmed a covalent interaction (Supp. Fig. 7). In this structure, the sulfolane ring occupies the hydrophobic proline-binding pocket that is formed by Met130, Gln131, Phe134, Thr152 and His157 (Fig. 1G). Furthermore, the sulfonyl oxygens mediate hydrogen bonds to the backbone amide of Gln131, and the imidazole NH of His157. These hydrogen bonds are analogous to those formed between Pin1 and arsenic trioxide [38] (Supp. Fig. 8). The tert-butyl group of Sulfopin covers a shallow hydrophobic patch formed by Ser114, Leu122 and Met130, but is mostly solvent-exposed, which explains the broad range of hydrophobic moieties that were tolerated at this position during the optimization campaign.

Despite being a very small ligand (heavy atom count: 17, cLogP: 0.36), Sulfopin efficiently exploits interactions with the active site of Pin1, even in the absence of a negatively charged moiety to interact with the phosphate-binding pocket, thus overcoming the cell permeability issues of previous Pin1 inhibitors, which are often highly anionic [74,76–78].

Sulfopin engages Pin1 in cells and in vivo

To evaluate the target engagement of Sulfopin in cells, we developed a desthiobiotin (DTB)-labeled probe for competition pull-down experiments. Based on the co-crystal structure of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Sulfopin bound to Pin1, we derivatized the mostly solvent-exposed tert-butyl group of Sulfopin with a PEG-linked DTB (Sulfopin-DTB; Fig. 2A). Sulfopin-DTB showed similar potency (apparent K_i = 38 nM in FP assay; Fig. 2B), and successfully engaged Pin1 in PATU-8988T cell lysates, achieving robust pull-down at 1 µM following a 1 h incubation period (Supp. Fig. 9A).

To assess the cellular target engagement and cell permeability of Sulfopin, we performed a live cell competition assay in PATU-8988T and HCT116 cells. A 1 µM treatment of Sulfopin achieved complete Pin1 engagement within 4 h (with about 50% after 2 h; Fig. 2C), and maintained significant engagement up to 72 h (Fig. 2E). Sulfopin exhibited dose-dependent competition for Pin1 binding, with maximal competition evident at 0.5-1 µM, while the negative control Sulfopin-AcA showed no competition (Fig. 2D). We found that this cellular engagement was extensible to other cell lines, such as IMR32, and MDA-MB-231 (Supp. Fig. 9C,D).

Ultimately, we were interested whether Sulfopin could be used *in vivo*. Sulfopin had encouraging metabolic stability in mouse hepatic microsomes ($T_{1/2}$ = 41 min), prompting us to submit it for pharmacokinetic (PK) profiling. In three mice, oral administration of 10 mg/kg Sulfopin achieved an average C_{max} of 11.5 µM and oral bioavailability (F%) of 30% (Supp. Table 5), suggesting that Sulfopin is suitable for oral *in vivo* dosing. We next evaluated the toxicity of Sulfopin in an acute toxicity model in which mice were administered daily doses of 10, 20 or 40 mg/kg Sulfopin by intraperitoneal (IP) injection for two weeks. No adverse effects or weight loss were recorded, and a post-mortem examination found no readily detectable pathologies.

Using Sulfopin-DTB, we were also able to assess the *in vivo* engagement of Sulfopin. To do this, mice were treated with three doses (over the course of two days) of vehicle, 10, or 20 mg/kg Sulfopin by oral gavage, followed by lysis of the spleens for a competition pull-down experiment. Effective Pin1 engagement was observed in 1 out of 3 mice treated with 10 mg/kg Sulfopin, and in 3 out of 3 mice treated with 20 mg/kg Sulfopin, with target engagement monitored by loss of Sulfopin-DTB-mediated pull-down (Fig. 2F). Based on these results, we chose a 40 mg/kg dose for further mouse experiments to ensure complete Pin1 engagement.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

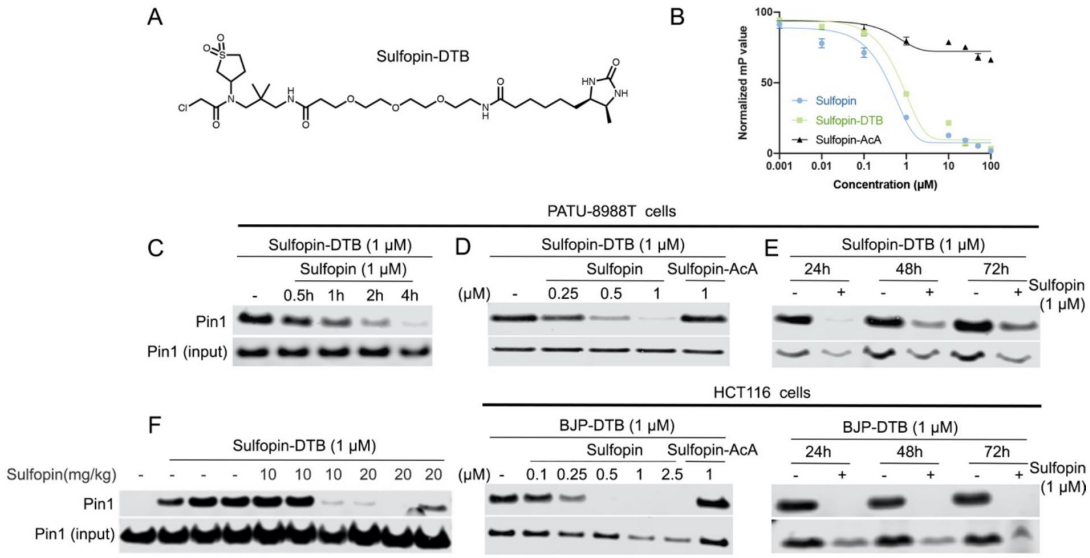


Figure 2. Sulfopin engages Pin1 in cells and *in vivo*.
A. Chemical structure of the desthiobiotin probe Sulfopin-DTB. **B.** Competitive FP assay shows that Sulfopin-DTB binds Pin1 with similar potency to Sulfopin. **C.** Sulfopin shows time dependent engagement in PATU-8988T cells. Cells were incubated for the indicated times with Sulfopin (1 μM), then lysed and incubated with Sulfopin-DTB (1 μM). Following enrichment, Pin1 levels were analysed by Western blot. **D.** Sulfopin fully engages Pin1 in PATU-8988T cells at 1 μM and in HCT116 cells at 0.5 μM (See Supp. Fig. 9B for a structure of BJP-DTB). Cells were incubated with Sulfopin at the indicated concentration for 5 h, followed by lysis and DTB probe incubation (1 h, 1 μM). The non-covalent control Sulfopin-AcA is not able to engage Pin1. **E.** Sulfopin shows long-term engagement with Pin1. PATU-8988T and HCT116 cells were incubated with or without Sulfopin for the indicated times, followed by lysis and incubation with DTB probes. Significant engagement (> 50-100%) is still evident after 72 h. **F.** Sulfopin engages Pin1 *in vivo*. Mice were treated by oral gavage with the indicated amounts of Pin1, over two days for a total of three doses. Following this treatment, their spleens were lysed for a competition pull-down experiment with Sulfopin-DTB.

Sulfopin is highly selective

To evaluate the selectivity of Sulfopin in cells, we assessed its target profile using Covalent Inhibitor Target-site Identification (CITe-Id [79], Fig. 3A). This chemoproteomic platform enables the identification and quantification of the dose-dependent binding of covalent inhibitors to cysteine residues on a proteome-wide scale. In this competition experiment, live PATU-8988T cells were incubated with Sulfopin (100, 500, 1000 nM) for 5 h, followed by cell lysis and co-incubation with Sulfopin-DTB (2 μM) for 18 h. Following trypsin digestion and avidin enrichment, the DTB-modified peptides were analyzed by multidimensional LC-MS/MS. Out of 162 cysteine residues labeled by Sulfopin-DTB, only Pin1 Cys113 (Fig. 3B; >2 s.d. from the median; Supp. Dataset 2A) exhibited dose-dependent competition (Fig. 3C), indicating the pronounced selectivity of Sulfopin.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

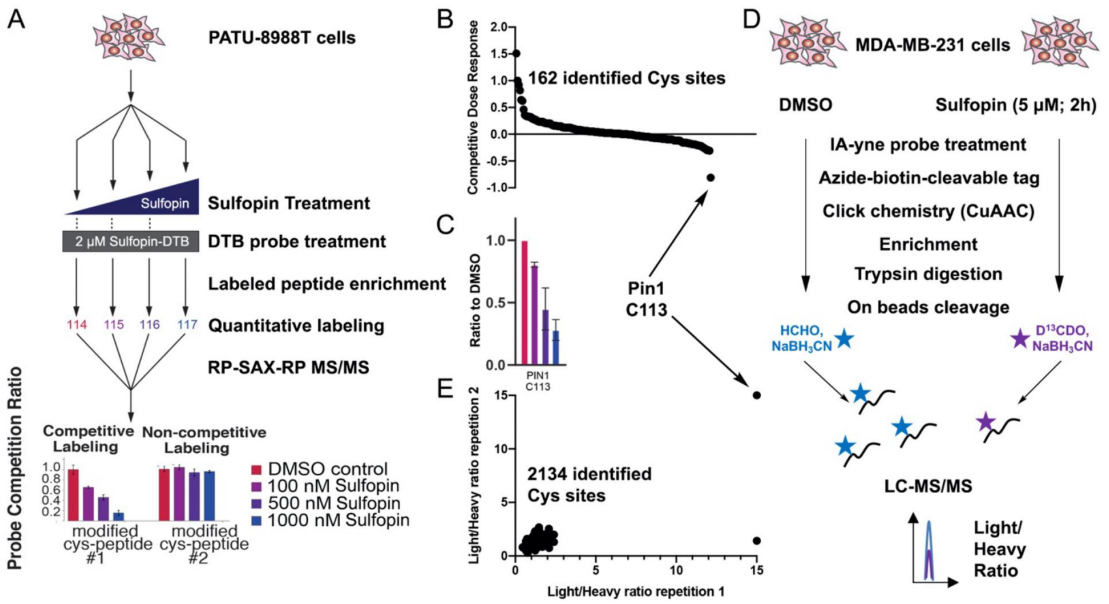


Figure 3. Sulfopin is highly selective for Pin1 in cells.
A. Scheme for Covalent Inhibitor Target-site Identification (CITe-Id) profiling of competitively labeled cysteine residues across the proteome following dose response treatment with Sulfopin. **B.** CITe-Id results. Out of 162 identified labeled cysteine residues, only C113 in Pin1 is labeled in a dose-dependent manner (see Supp. Dataset 2A for full list of identified peptides). **C.** Dose dependency of C113 labeling in CITe-Id experiment. **D.** Scheme for independent rdTOP-ABPP experiment for assessing Sulfopin proteomic selectivity **E.** Out of 2134 identified cysteines in the experiment, only two cysteines showed light/heavy ratio > 2.5, of these, one cysteine did not replicate, and only Pin1 C113 showed the maximal ratio of 15 in both replicates.

To further investigate Sulfopin's selectivity, we used a complementary chemoproteomic method. We performed an rdTOP-ABPP experiment to profile its cysteine targets throughout the proteome (Fig. 3D)[80]. This variant of the isoTOP-ABPP technique enables the site-specific quantification of cysteine binding by label-free covalent inhibitors. In brief, MDA-MB-231 cells were treated with Sulfopin (5 μM, 2 h), lysed and labeled with a bioorthogonal Iodoacetamide-alkyne (IA-yne) probe that was then conjugated to a cleavable biotin tag by copper-catalyzed azide−alkyne cycloaddition (CuAAC). After enrichment on beads, the peptides were isotopically derivatized by duplex reductive dimethylation, cleaved and analyzed via LC-MS/MS analysis. We identified Cys113 of Pin1 as the top ranked cysteine that was labeled by Sulfopin at biologically relevant concentration (5 μM), with a competition ratio R = 15 across two biological replicates (Fig. 3E; Supp. Dataset 2B). All other identified cysteines showed R values < 2.5. Overall, we used two independent chemoproteomic techniques in two different cell lines to demonstrate that

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Sulfopin has exquisite selectivity for Pin1 Cys113 and is therefore a suitable probe with which to interrogate Pin1's function in cells and in vivo.

Treatment with Sulfopin phenocopies known Pin1 knockout phenotypes

We next assessed whether pharmacological inhibition of Pin1 by Sulfopin could recapitulate two previously reported phenotypes associated with Pin1 genetic knockout. First, Pin1 knockout was reported to abrogate phosphorylation of IRAK1 Thr209 and resensitize radioresistant cancer cells to irradiation [31]. Accordingly, we found that treating radioresistant HeLa cells with Sulfopin significantly resensitized them to irradiation in a dose dependent manner (Fig. 4A), and also decreased IRAK1 Thr209 phosphorylation at concentrations as low as 100nM (Fig. 4B,C).

Second, germinal centers (GCs) are sites where B cells proliferate and undergo somatic hypermutation in a BCL6- and Myc-dependent manner. Pin1-deficient mice are reported to display a significant increase in the frequency of GC B cells in response to immunization [81]. To induce GCs and examine Sulfopin's effect on GC B cells, we immunized the hind foot pads of 12 wild-type (WT) mice with ovalbumin coupled to the 4-hydroxy-3-nitrophenylacetyl (NP-OVA). The mice were injected with two doses of Sulfopin (IP; 40mg/kg) or vehicle on days 7 and 9 post immunization, at the peak of the GC response, and on day 11 the mice were sacrificed and frequency of GC B cells in lymph nodes was assessed by flow cytometry (Fig. 4D). In accordance with the previous report [81], Sulfopin treated mice exhibited 1.34-fold higher proportion of GC B cells compared to mice treated with control vehicle (Fig. 4E). Taken together, these data demonstrate that Sulfopin phenocopies the effects of Pin1 genetic deletion.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

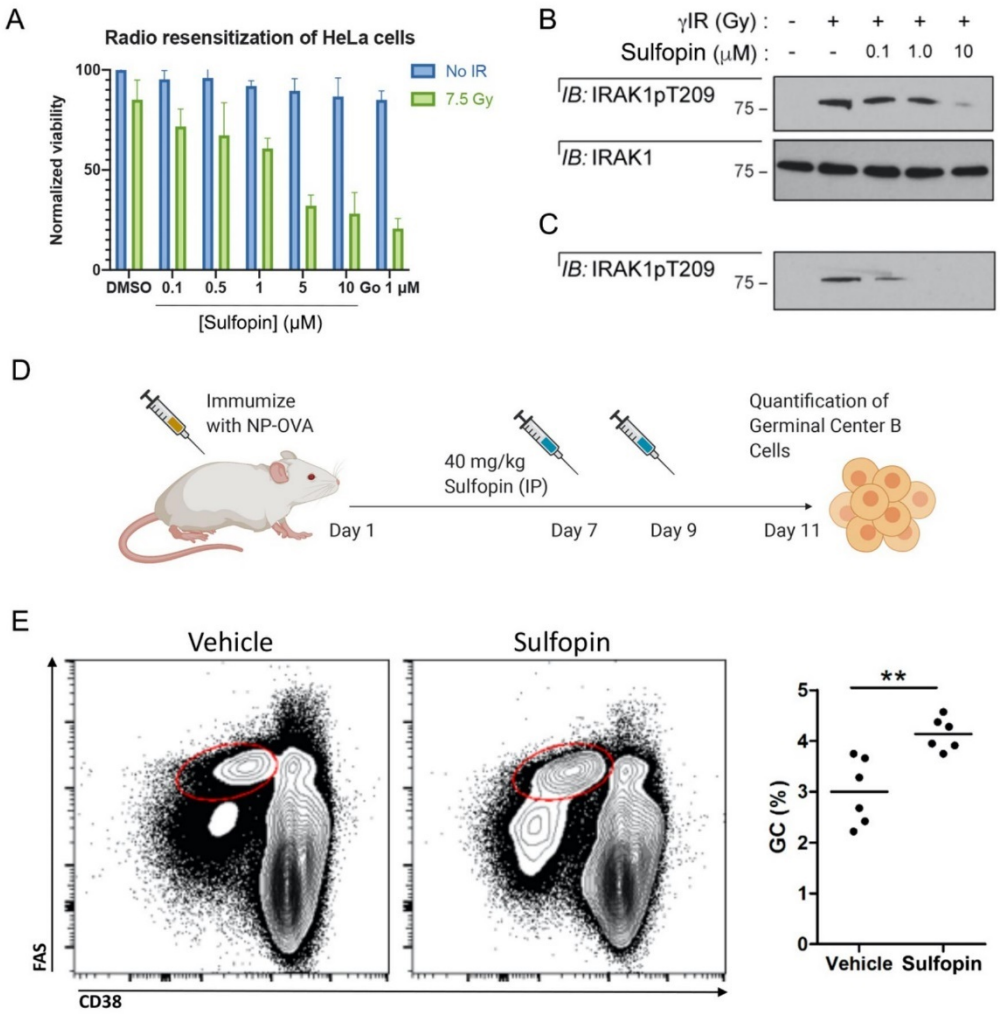


Figure 4. Sulfopin phenocopies Pin1 knockout phenotypes.

A. HeLa cells were treated with either DMSO, Sulfopin, or Go6976 (a Chk1 inhibitor) and exposed to 7.5 Gy IR 1 h after drug treatment. Viability was assessed 3 days post-IR. Sulfopin shows a dose dependent sensitization of the cells to irradiation. **B.** Western blot analysis was performed 24 h post-IR, showing Sulfopin blocked phosphorylation of Thr209 of IRAK1. **C.** A shorter exposure shows that Sulfopin inhibits IRAK1 phosphorylation already at concentrations of 0.1 μM. **D.** A scheme for testing the effect of Sulfopin *in vivo* on germinal center B cells in response to immunization. **E.** Representative flow cytometric plots with Vehicle and Sulfopin (left) and quantification (right) of FAS^{Hi} $CD38^{-}$ germinal center (GC) cells in WT mice 11 days after immunization with NP-OVA. ** $p<0.01$, two tailed Student's t test.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Viability effects of Sulfopin in cancer cell lines

To broadly profile the anti-proliferative activity of Sulfopin, we used the PRISM platform [82] (Broad Institute) to evaluate its potency against 300 human cancer cell lines. The PRISM method enables high-throughput, pooled screening of mixtures of cell lines, which are each labeled with a 24-nucleotide barcode [82]. In all 300 cancer cell lines profiled, Sulfopin demonstrated limited or no anti-proliferative activity after a 5-day treatment, with IC_{50} values > 3 μM. This result aligns with our initial cytotoxicity screening, as well as data from the Cancer Dependency Map, in which Pin1 was not identified as a significant genetic dependency in CRISPR-Cas9 and RNAi screens across hundreds of adherent and suspension cancer cell lines (https://depmap.org/portal/). This suggests that the strong single-agent cytotoxicity of previously published Pin1 inhibitors, such as juglone, is likely attributable to off-target effects [42,83] (Supp Fig. 10).

We next assessed whether Sulfopin treatment might induce antiproliferative activity effects after prolonged treatments (6-8 days). To ensure that target engagement was maintained for the duration of the experiment, we replenished Sulfopin in fresh media every 48 h. We found that Sulfopin treatment impacted the viability of PATU-8988T cells in a Pin1-dependent manner, having no effect on proliferation in the corresponding Pin1 KO cells (Fig. 5A). To evaluate whether this time-dependent growth phenotype was extensible to other cancer types, we performed additional experiments in breast (MDA-MB-468), prostate (PC3), and ovarian (Kuramochi) cancer cell lines. MDA-MB-468 cells showed the most pronounced sensitivity to Sulfopin, which diminished cell viability in a dose- and time-dependent manner, while Sulfopin-AcA did not affect cell proliferation (Fig. 5B). Both PC3 and Kuramochi cells exhibited only very modest sensitivity to Sulfopin treatment.

Given that three-dimensional (3D) culture models can reflect *in vivo* results better than monolayer cell culture[84], we next evaluated the anti-proliferative activity of Sulfopin in PATU-8988T WT or Pin1 KO cells grown in 3D matrigel domes. Following a 9-day treatment, replenishing compound in media every 3 days, Sulfopin demonstrated modest anti-proliferative activity in PATU-8988T WT Pin1 cells, with no effects observed in PATU-8988T Pin1 KO cells, suggesting an on-target phenotype (Fig. 5C).

Collectively, these data suggest that Pin1 inhibition does not cause proliferation defects at short time points, but that it does moderately affect cell proliferation after prolonged treatments, and in 3D cell culture.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

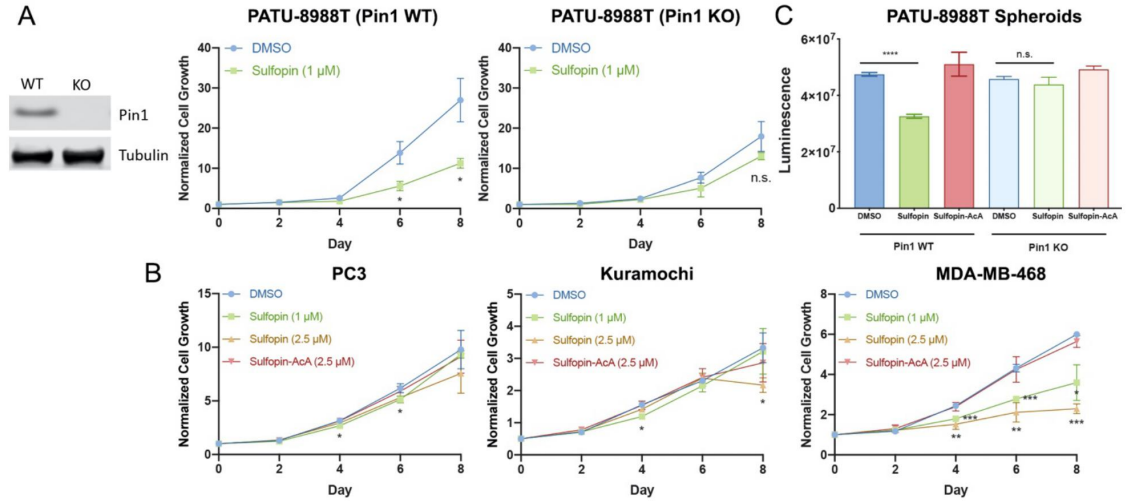


Figure 5. Sulfopin has a Pin1 dependent viability effect on long term exposure.
A. We generated a PATU-8988T Pin1 knockout (KO) cell line, as evidenced by Western blot. Sulfopin (1 μM) has a significant effect on cellular viability after 6 and 8 days in the WT PATU-8988T cells, but shows no significant effect on viability in the Pin1 KO cells (n=3). **B.** Sulfopin shows varying anti-proliferative effects across cancer cell lines, with the most pronounced sensitivity observed in MDA-MB-468 cells (n=3). **C.** Relative viability of PATU-8988T WT and Pin1 KO cells grown in 100% Matrigel domes following treatment with Sulfopin (1 μM; n=9) or Sulfopin-AcA (1 μM; n=9). The non-covalent control Sulfopin-AcA shows no effect in any of the tested systems. Statistical significance calculated using Student's t test with unequal variance (n.s. = $p > 0.05$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$). In all graphs, error bars indicate standard deviation.

Sulfopin downregulates Myc transcriptional activity

We and others have previously shown that Pin1 regulates the c-Myc oncoprotein [85], affecting Myc protein stability [86] as well as its DNA binding and transcriptional activity [22,87]. Pin1 physically interacts with c-Myc [86,88], altering isomerization of P63 following phosphorylation of S62. We have shown that overexpression of Pin1 can lead to an increase in the transcription of c-Myc target genes while knockdown of Pin1 decreases Myc-dependent transcription [22].

To test whether Sulfopin affects Myc transcriptional output, we treated Mino B cells with Sulfopin (1 μM; 6 h; in triplicates) or DMSO and performed a global RNA sequencing analysis to detect differentially expressed genes as the result of this perturbation. 206 genes were found to be significantly down-regulated (Fig. 6A; Supp. Dataset 3). We performed a gene set enrichment analysis of these genes using Enrichr [89], against a dataset of genes identified by ChIP-seq (chromatin immunoprecipitation followed by sequencing) for various transcription factors. Myc target genes in K562 cells and HeLa-S3 cells appeared as the most enriched set and the third

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

most enriched set, respectively (adjusted p-value of $1.99x10^{-16}$ and $2.00x10^{-13}$ respectively; Fig. 6B), validating a significant downregulation of Myc's transcriptional signature by Sulfopin. To further validate the effect of Sulfopin on Myc transcriptional activity, we co-transfected HEK-293 cells with a Myc reporter construct (4x-Ebox-Luciferase) and Pin1. As expected, Pin1 expression increased Myc transcriptional activity, while treatment with 2 μM Sulfopin for 48 h resulted in a significant reduction in relative luciferase activity as compared to DMSO control (Fig. 6C). Together these results suggest that treatment with Sulfopin downregulates Myc target genes and, thus, points to Myc-driven cancers as natural candidates for its therapeutic application. Accordingly, we next evaluated Sulfopin in *in vivo* models of MYCN-driven neuroblastoma (NB).

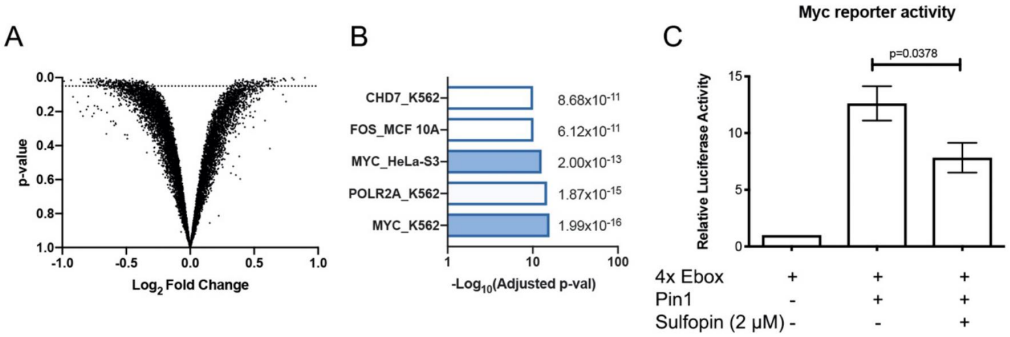


Figure 6. Sulfopin downregulates Myc transcription.
A. Results of an RNA-sequencing experiment, comparing the change in RNA levels between Mino B cells treated with either Sulfopin (1 μM, 6 h, in triplicates) or DMSO. Each dot presents the Log_2 fold change of a transcript (x-axis) vs. the p-value for significance of that change (Student's t test; y-axis). The dotted line indicates p=0.05. 206 genes were significantly down regulated. **B.** Results of a gene set enrichment analysis using Enrichr against the ENCODE TF ChIP-seq set. Two of the most enriched sets are Myc target genes from different cell lines. **C.** HEK293 cells were transfected with a 4x Ebox-luciferase reporter for Myc transcriptional activity levels. Co-transfection with Pin1 increased reporter activity, while 48 h treatment with Sulfopin significantly (Student's t test) reduced this activity compared to DMSO.

Sulfopin blocks tumor initiation and progression in MYCN-driven zebrafish models of neuroblastoma

To evaluate Sulfopin's effects on Myc-driven cancers, we turned to a zebrafish model of neuroblastoma [90–92], a pediatric malignancy derived from the peripheral sympathetic nervous system (PSNS)[93]. During the development of normal zebrafish embryos, neural crest-derived PSNS neuroblasts form the primordial superior cervical ganglia (SCG) and intrarenal gland (IRG) at the age of 3 to 7 days post fertilization (dpf), can be visualized using the dβh:EGFP fluorescent reporter [91] (Fig. 7A). Overexpression of the MYCN oncogene, which is the oncogenic driver in

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

approximately 20% of human high-risk neuroblastomas, causes the fish to develop neuroblast hyperplasia in the PSNS of Tg(dβh:MYCN;dβh:EGFP) transgenic zebrafish (Fig. 7A upper right). These neuroblast hyperplasia rapidly progress into fully transformed tumors that faithfully resemble human high-risk neuroblastoma [90–92]. When Sulfopin was added to the fish water containing Tg(dβh:MYCN;dβh:EGFP) zebrafish (at concentrations of 25-100 μM), the neuroblastoma-initiating hyperplasia was significantly suppressed and fully transformed neuroblastoma did not develop over the treatment period (Fig. 7A,B). This indicates that Sulfopin blocks neuroblastoma initiation *in vivo* in this tumor model. In addition, no evidence of toxicity was observed in embryos treated with Sulfopin at these concentrations, further supporting our findings in mice that Sulfopin is well tolerated by healthy tissues *in vivo*.

We then assessed the anti-tumor activity of Sulfopin on the maintenance of fully transformed neuroblastoma cells *in vivo* in primary tumor derived allograft (PDA) models constructed in transplanted zebrafish embryos. EGFP-labeled neuroblastoma cells were dissected from four-month-old Tg(dβh:MYCN;dβh:EGFP) donor zebrafish, disaggregated, counted and 200-400 GFP-labeled tumor cells were injected intravenously into the Duct of Cuvier (common cardinal vein) of 2 dpf zebrafish embryos [94]. One day after injection, 100 μM Sulfopin or the DMSO control was added to the fish water containing embryos bearing the transplanted EGFP-labeled neuroblastoma cells. After five days, we quantified the area of the EGFP-labeled tumor mass in treated embryos, and discovered that tumor masses in the DMSO-treated embryos grew larger over the five days of treatment, while the tumor masses decreased in size in the Sulfopin treated embryos (Fig. 7C,D). Hence, Sulfopin treatment not only suppressed MYCN-driven neuroblastoma initiation, but also suppressed the growth and survival *in vivo* of transplants of fully transformed primary neuroblastoma tumor cells.

Sulfopin blocks tumor initiation and progression in a MYCN-mouse model of neuroblastoma

Following the encouraging results in the zebrafish models, we also assessed the effects of Sulfopin in a murine model of neuroblastoma - the Th-MYCN GEM (genetically engineered mouse) model of neuroblastoma, in which human MYCN is expressed under the tyrosine hydroxylase promoter [95]. The Th-MYCN faithfully recapitulate the major molecular and histopathologic features of high-risk MYCN-amplified neuroblastoma, and the model has been widely used for therapeutics studies. The study was performed using both male and female hemizygous mice, which spontaneously developed palpable tumors at 50 to 130 days with a 25% penetrance. Once tumors were palpable, mice were randomly assigned to treatment groups, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

treated once per day with either vehicle or 40 mg/kg Sulfopin. In this aggressive model, we assessed the overall survival of the mice. Treated mice showed significant (p=0.0127) increase in overall survival, with an average increase of 10 days for Sulfopin treated mice (Fig. 7E).

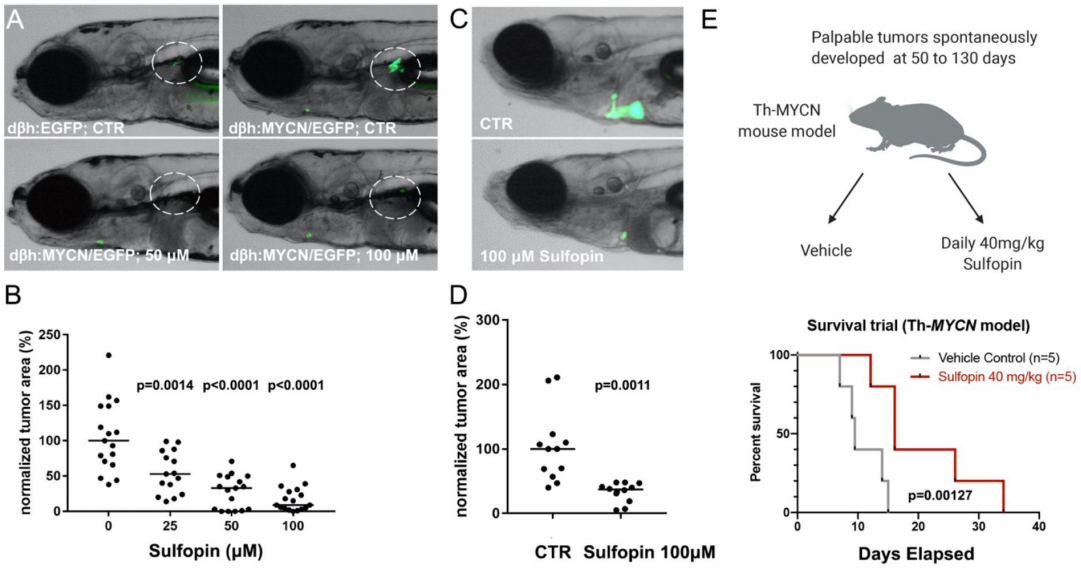


Figure 7. Sulfopin abrogate MYCN driven neuroblastoma growth *in vivo*. A. PSNS cells in the primordial superior cervical ganglia (SCG) and intrarenal gland (IRG; highlighted by dotted circles) in representative embryos of Tg(dβh:EGFP) (upper left) and Tg(dβh:MYCN;dβh:EGFP) (upper right) transgenic zebrafish at 7 dpf. Representative 7 dpf Tg(dβh:MYCN;dβh:EGFP) zebrafish treated with 50 µM Sulfopin (bottom left) and 100 µM Sulfopin (bottom right). The primordial SCG plus IRG areas are highlighted by dotted circles. **B.** Quantification of GFP+ cells in the primordial SCG and IRG of 7 dpf Tg(dβh:MYCN;dβh:EGFP) embryos treated with Sulfopin at multiple doses. A Mann-Whitney test with confidence intervals of 95% was used for the analysis of significance (p-value) and the quantitative data are reported as median. **C.** Representative zebrafish embryos transplanted with neuroblastoma cells isolated from four-month old Tg(dβh:MYCN;dβh:EGFP) donor zebrafish and treated with DMSO control (upper) or 100 µM Sulfopin added to the fish water (lower). **D.** Quantification of EGFP-positive tumor area in zebrafish embryos treated with DMSO control and 100 µM Sulfopin added to the fish water. A Mann-Whitney test with confidence intervals of 95% was used for the analysis of significance (p-value) and the quantitative data are reported as median. **E.** Survival trials in Th-*MYCN* mice. Mice were randomly assigned to treatment groups, and treated daily with either vehicle or 40 mg/kg Sulfopin, as soon as the tumors were palpable. Treated mice showed significant (p=0.0127) increase in overall survival, with an average increase of 10 days survival.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Discussion

Despite extensive, decades-long efforts to discover Pin1 inhibitors, no approach has yielded a compound capable of selectively blocking Pin1 *in vivo*. Here, we describe the development and *in vivo* characterization of Sulfopin, a highly selective and potent, covalent Pin1 inhibitor with low inherent reactivity and negligible toxicity, which blocks tumor initiation and progression in both zebrafish and mouse models of neuroblastoma.

While inhibitors such as juglone, ATRA, ATO and KPT-6566 pioneered the investigation of Pin1 in cancer-related contexts, they all exert their anti-cancer effects in part through Pin1-independent mechanisms [41] that include DNA-damage [39], induction of Pin1 degradation [37] or by directly blocking transcription [83]. Our recently reported covalent peptide inhibitor BJP-06-005-3 [43], while selective in cells, is not suitable for *in vivo* application. In contrast, Sulfopin is highly specific for Pin1, as we established here using multiple orthogonal experimental strategies. Therefore, this highly selective *in vivo* tool allowed us to investigate Pin1 as a *bona fide* drug target in cancer for the first time. Thus far, Pin1 has proven a challenging target for both ligand- [74] and structure-based [76–78] approaches, even in studies employing high-throughput-screenings of up to one million compounds [76]. This is in part due to the shallow nature of the Pin1 active site, which is evolved to bind peptides, further complicated by the phosphate binding site, which favors negatively charged moieties. To overcome this hurdle, we screened a library of electrophilic fragments, which can compensate for sparse protein interfaces by irreversibly binding to the Pin1 Cys113. The screen resulted in a relatively high hit rate of 11% (111 compounds with >50% labeling) compared to previous screens [62]. This might be explained by the high reactivity of Cys113, which was found amongst the most reactive cysteines in several chemoproteomic campaigns [67].

A major finding during our hit optimization campaign was that intrinsic warhead reactivity of Pin1 binders did not correlate with potency (Supp. Fig. 5). Despite the high structural similarity of the optimized compounds, their reactivities spanned over 30-fold (Supp. Table 2), with Sulfopin having the lowest reactivity of the best binders (Fig. 1C), with reactivity nearing that of acrylamides (Supp. Fig. 11). This low reactivity manifested in exquisite cellular selectivity. Two separate chemoproteomic approaches, performed in two different cell lines, both identified Pin1 as the sole target of Sulfopin by a wide margin (Fig. 3). This result is rare for covalent inhibitors [79] and even for FDA approved covalent drugs [96].

This proteomic selectivity for Pin1 allowed us to show that short-term Pin1 inhibition does not induce cytotoxicity, despite previous results reported with less selective inhibitors [39,97–102].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Juglone, for instance, has identical LD_{50} for cell lines lacking Pin1 [42,43] (Supp. Fig. 10). Sulfopin induced little to no cytotoxicity across a panel of 300 cancer cell lines (PRISM screen), though it significantly impacted cancer cell growth after prolonged treatments (6-8 days; Fig. 5A,B) and in 3D-cell culture (Fig. 5C). The irreversible binding of Sulfopin allowed us to demonstrate target engagement in cells and *in vivo* (Fig. 2), showing that its favorable PK/PD profile (Supp. Table 5) translated to complete *in vivo* target engagement and target modulation as demonstrated by the ability of Sulfopin to phenocopy previous genetic results *in vivo* (Fig. 4). This result underscores the utility of Sulfopin as an *in vivo* tool compound, with broad applicability to domains such as immunology, and as we later show, cancer biology.

In agreement with previous studies [22,86], our RNA sequencing experiment suggested Myc target gene downregulation as a major consequence of Pin1 inhibition (Fig. 6A,B). This conclusion is supported by transcriptional suppression in the Myc luciferase reporter assay (Fig. 6C), and the demonstrated efficacy of Pin1 inhibition in various MYCN-driven cancer models (Fig. 7). However, given that Pin1 plays a central role in numerous signaling pathways, it is likely that Myc-downregulation is not the only mechanism at play. Indeed, additional transcription factors showed significant downregulation in the RNA-seq dataset, including RelA (Supp. Dataset 4) which was also previously linked to Pin1 [9,103] and could contribute to oncogenesis. Other pathways that were previously linked to Pin1 have also been implicated in neuroblastoma [104,105]. RNA-seq experiments in additional cell lines (and perhaps additional time points) might shed more light on other processes regulated by Pin1.

Further investigation is needed to establish whether the therapeutic efficacy of Sulfopin can be further enhanced. For instance, we have yet to reach dose limiting toxicity with Sulfopin, and observed dose-dependent responses in zebrafish (Fig. 7B), indicating that treatment with higher doses might enable more pronounced effects in mice. Furthermore, its favorable toxicity profile could enable the use of Sulfopin in combination with other drugs. Since Pin1 regulates numerous cellular pathways, it stands to reason that Pin1 inhibition may be synthetic lethal with the loss of other targets. Ongoing studies will clarify the full therapeutic potential of Sulfopin.

In summary, we present a potent and selective, covalent Pin1 inhibitor with *in vivo* activity and no toxicity that allows investigation of Pin1 biology in physiologically relevant models in health and disease. Using Sulfopin as a pharmacological tool will enable the study of the many diverse processes that are regulated by Pin1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Methods

Electrophile library screening

993 compounds as 20 mM DMSO stocks in 384 well-plate format were transferred into a 384 well plate working copy by combining 0.5 µL of five compounds per well. The catalytic domain of Pin1 (2 µM) in 20 mM Tris, 75 mM NaCl, pH7.5 was incubated with 200 µM for each compound and moderately shaken for 24 h at 4°C. The reaction was stopped by the addition of formic acid to 0.4% final concentration. The LC/MS runs were performed on Waters ACUITY UPLC class H, in positive ion mode using electrospray ionization. UPLC separation using C4 column (300 Å, 1.7 µM, 21 mm×100 mm). the column was held at 40°C, and the autosampler at 10°C. Mobile solution A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile. Run flow was 0.4 mL/min. Gradient used for BSA was 20% B for 2 minutes increasing linearly to 60% B for 4 minutes holding at 60% B for 2 minutes, changing to 0% B in 0.1 minutes and holding at 0% for 1.9 minutes. Gradient for the other proteins was 20% B for 2 minutes increasing linearly to 60% B for 3 minutes holding at 60% B for 1.5 minutes, changing to 0% B in 0.1 minutes and holding at 0% for 1.4 minutes. The mass data was collected at a range of 600-1300 m/z. Desolvation temperature was 500°C with a flow rate of 1000 liter/hour. The voltage used were 0.69 kV for the capillary and 46 V for the cone. Raw data was processed using openLYNX and deconvoluted using MaxEnt. Labeling assignment was performed as previously described[62].

Covalent Docking

Covalent docking was performed using DOCKovalent 3.7[73] against 16 structures of Pin1. PDB codes: 1PIN, 2ITK, 2Q5A, 2XP3, 2ZQV, 2ZR4, 3IK8, 3KAB, 3KCE, 3NTP, 3ODK, 3OOB, 3TC5, 3TCZ, 3TDB, 3WH0. The docked compounds include seven sulpholane hits from the electrophilic library with the following IDs: PCM-0102138, PCM-0102178, PCM-0102105, PCM-0102832, PCM-0102313, PCM-0102760, PCM-0102755. The covalent bond length was set to 1.8Å and the two newly formed bond angles to Cβ-Sγ-C=109.5 ± 5° and Sγ-C-Ligatom=109.5 ± 5°.

Thiol reactivity assay

50 µM DTNB was incubated with 200 µM TCEP in 20 mM sodium phosphate buffer pH 7.4, 150 mM NaCl, for 5 minutes at room temperature, in order to obtain TNB^{2-}. 200 µM compounds were subsequently added to the TNB^{2-}, followed by immediate UV absorbance measurement at 412 nm at 37°C . The UV absorbances were acquired every 15 minutes for 7 h. The assay was performed in a 384 well-plate using a Tecan Spark10M plate reader. Background absorbance of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

compounds was subtracted by measuring the absorbance at 412 nm of each compound in the same conditions without DTNB. Compounds were measured in triplicates. The data was fitted to a second order reaction equation such that the rate constant k is the slope of ln([A][B0]/[B][A0]). Where [A0] and [B0] are the initial concentrations of the compound (200 µM) and TNB^{2-} (100 µM) respectively, and [A] and [B] are the remaining concentrations as a function of time as deduced from the spectrometric measurement. Linear regression using Prism was performed to fit the rate against the first 4 h of measurements.

Pin1 expression and purification

A construct of full-length human Pin1 in a pET28 vector was overexpressed in E. coli BL21 (DE3) in LB medium in the presence of 50 mg/ml of kanamycin. Cells were grown at 37°C to an OD of 0.8, cooled to 17°C, induced with 500 µM isopropyl-1-thio-D-galactopyranoside, incubated overnight at 17°C, collected by centrifugation, and stored at -80°C. Cell pellets were sonicated in buffer A (50 mM hepes 7.5, 500 mM NaCl, 10% glycerol, 20 mM Imidazole, and 7 mM BME) and the resulting lysate was centrifuged at 30,000 xg for 40 min. Ni-NTA beads (Qiagen) were mixed with lysate supernatant for 30 min and washed with buffer A. Beads were transferred to an FPLC-compatible column and the bound protein was washed with 15% buffer B (50 mM hepes 7.5, 500 mM NaCl, 10% glycerol, 250 mM Imidazole, and 3 mM BME) and eluted with 100% buffer B. Thrombin was added to the eluted protein and incubated at 4°C overnight. The sample was concentrated and passed through a Superdex 200 10/300 column (GE helathcare) in a buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol, and 1 mM TCEP. Fractions were pooled, concentrated to approximately 37 mg/ml and frozen at -80°C.

Pin1 crystallization and soaking

Apo protein at a final concentration of 1 mM was crystallized by sitting-drop (200 nL + 200 nL) vapor diffusion at 20°C in the following crystallization buffer: 3 M NH_4SO_4, 100 mM HEPES-pH7.5, 150 mM NaCl, 1% PEG400, and 10 mM DTT. A volume of 200 nL of 1 mM Sulfopin was added directly to crystals for soaking at 20°C for 16 h. Crystals were transferred briefly into crystallization buffer containing 25% glycerol prior to flash-freezing in liquid nitrogen.

Crystallization data collection and structure determination

Diffraction data from complex crystals were collected at beamline 24ID-C of the NE-CAT at the Advanced Photon Source at the Argonne National Laboratory. Data sets were integrated and scaled using XDS[106]. Structures were solved by molecular replacement using the program

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Phaser[107] and the search model PDB entry 1PIN. Iterative manual model building and refinement using Phenix[108] and Coot[109] led to the final models (Supp. Table 4).

Fluorescence polarization (FP) assay

Binding affinity to Pin1 was determined using a fluorescence polarization (FP) assay to assess competition with an N-terminal fluorescein-labeled peptide (peptide core structure: Bth-D-pThr-Pip-Nal), which was synthesized by Proteintech. The indicated concentrations of candidate compounds were pre-incubated for 12 h at 4°C with a solution containing 250 nM glutathione S-transferase (GST)-Pin1, 5 nM of fluorescein-labeled peptide probe, 10 µg/ml bovine serum albumin, 0.01% Tween-20 and 1 mM DTT in a buffer of 10 mM HEPES, 10 mM NaCl and 1% glycerol (pH 7.4). Measurements of FP were made in black 384-well plates (Corning) using an EnVision reader. K_i values obtained from the FP assay results were derived from the Kenakin K_i equation: Kenakin K_i = (Lb)(EC_{50})(K_d)/(Lo)(Ro) + Lb(Ro–Lo + Lb–K_d), where K_d [M]: K_d of the probe, EC_{50} [M]: concentration of unlabeled compound that results in 50% inhibition of binding (obtained from FP assay), total tracer Lo [M]: probe concentration in FP, bound tracer Lb [M]: 85%, fraction of probe bound to Pin1, total receptor Ro [M]: Pin1 concentration in the FP assay, as described[37,110].

Pin1 PPIase activity assay

Inhibition of Pin1 isomerase activity was determined using the chymotrypsin-coupled PPIase assay, using GST-Pin1 and Suc-Ala-pSer-Pro-Phe-pNA peptide substrate, as described previously[75]. GST-Pin1 was pre-incubated with the indicated concentrations of compound for 12 h at 4°C in buffer containing 35 mM HEPES (pH 7.8), 0.2 mM DTT, and 0.1 mg/mL BSA. Immediately before the assay was started, chymotrypsin (final concentration of 6 mg/mL) was added, followed by the addition of the peptide substrate (Suc-Ala-pSer-Pro-Phe-pNA peptide substrate, final concentration 50 mM). The K_i value obtained from the PPIase assay was derived from the Cheng–Prusoff equation, $K_i = IC_{50}/ (1 + S/K_m)$, where K_m is the Michaelis constant for the peptide substrate, S is the initial concentration of the substrate in the assay, and IC_{50} is the half-minimal inhibitory concentration of the inhibitor.

Cell Culture and Reagents.

PATU-8988T (DSMZ), MDA-MB-468 (ATCC), and HeLa (ATCC) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Sigma) and 1% penicillin/streptomycin. PC3 (ATCC), IMR32 (ATCC) and Kuramochi (Panagiotis A.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Konstantinopoulos's laboratory) cells were cultured in RPMI 1640 medium with L-glutamine, supplemented with 10% FBS and 1% penicillin/streptomycin. All cell lines were cultured at 37 °C in a humidified chamber in the presence of 5% CO_2. All cell lines were tested for the absence of Mycoplasma infection on a monthly basis.

Immunoblotting.

Whole cell lysates for immunoblotting were prepared by pelleting cells from each cell line at 4°C (300 g) for 5 minutes. The resulting cell pellets were washed 1x with 5 mL ice-cold PBS and then resuspended in the indicated cell lysis buffer. Lysates were clarified at 14,000 rpm for 15 minutes at 4°C prior to quantification by BCA assay (Pierce, cat#23225). Whole cell lysates were loaded into Bolt 4-12% Bis-Tris Gels (Thermo Fisher, cat#NW04120BOX) and separated by electrophoreses at 95 V for 1.5 h. The gels were transferred to a nitrocellulose membrane using the iBlot Gel Transfer at P3 for 6 minutes (Thermo Fisher, cat#IB23001) and then blocked for 1 h at room temperature in Odyssey blocking buffer (LICOR Biosciences, cat#927-50010). Membranes were probed using antibodies against the relevant proteins at 4°C overnight in 20% Odyssey Blocking Buffer in 1x TBST. Membranes were then washed three times with 1x TBST (at least 5 minutes per wash) followed by incubation with the IRDye goat anti-mouse (LICOR, cat#926-32210) or goat anti-rabbit (LICOR, cat #926-32211) secondary antibody (diluted 1:10,000) in 20% Odyssey Blocking Buffer/1x TBST for 1 h at room temperature. After three washes with 1x TBST (at least 5 minutes per wash), the immunoblots were visualized using the ODYSSEY Infrared Imaging System (LICOR). Antibodies used against various proteins were as follows: Pin1 (1:1,000, Cell Signaling cat#3722), α-Tubulin (1:1,000, Cell Signaling cat#3873), IRAK1 (Cell Signaling Technology #4504), IRAK1 pT209 (Assay Biotech #A1074).

Lysate Pull-Down with Sulfopin-DTB.

The indicated cells were lysed in 50 mM Hepes, pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM TCEP, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor tablet (Roche cat#4693159001). After clarifying (14,000 rpm for 15 min), lysates were incubated with the indicated concentrations of Sulfopin-DTB for 1 h at 4°C, using 500 µg of lysate per sample. Lysates were then incubated with streptavidin agarose resin (30 µL of 1:1 beads:lysis buffer slurry) (Thermo scientific, cat#20349) for 1.5 h at 4°C. Beads were washed four times with 500 µL of buffer (50 mM Hepes, pH 7.5, 10 mM NaCl, 1 mM EDTA, 10 % glycerol), then pelleted by centrifugation and dried. The beads were boiled for 5 minutes at 95°C in 30 µL of 2x LDS + 5% β-mercaptoethanol. Lysates were probed for specified proteins by Western blotting using the Bolt system (Life Technologies).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Cellular Target Engagement – Competition with Sulfopin-DTB.

The indicated cell lines were plated in 10 cm plates with 2.5 million cells per plate in 6 mL of media. The day after plating, cells were treated with DMSO or the indicated concentrations of candidate inhibitor for the indicated time points. The cells were then washed two times with cold PBS (1 mL per 10 cm plate) and collected by scraping with a cell scraper. Cells were lysed in 50 mM Hepes, pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM TCEP, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor tablet (Roche) – using 210 µL of cell lysis buffer per 10 cm plate of cells. After clarifying (14,000 rpm for 15 min), 9 µL of each lysate sample was combined with 4x LDS + 10% β-mercaptoethanol (in a ratio of 3:1), boiled for 5 minutes, and set aside for the input loading control (later to be loaded directly on the gel). Then, 200 µL of each lysate sample was incubated with 1 µM of Sulfopin-DTB for 1 h at 4°C and processed as in "lysate pull-down with Sulfopin-DTB" (above).

Radiosensitization studies

AlamarBlue-based cell viability assays were performed as previously described[31]. Briefly, HeLa cells were seeded at a density of 200 cells/well in a 96-well plate. After 16 h, cells were treated with Sulfopin and exposed to 7.5 Gy IR using an X-RAD 320 PRECISION X-RAY irradiator 1 hour after drug treatment. At 3 days post-IR, cells were incubated with AlamarBlue (ThermoFisher) at a final concentration of 10%. At 4 days post-IR, absorbance was measured at a wavelength of 570 nm with a 600 nm reference wavelength. Relative fluorescence was calculated using cell-free wells as a control reference, and percentage survival was calculated by comparing with DMSO-treated, non-irradiated controls. Sulfopin efficacy was assessed at 24 h post-IR by Western blot using anti-IRAK1 (Cell Signaling Technology #4504) and anti-IRAK1pT209 (Assay Biotech #A1074) antibodies.

In vivo germinal centers evaluation

WT mice (C57BL/6) were provided by Harlan, Israel. All experiments with mice were approved by the Weizmann Institute IACUC committee. Mice were immunized by injection of OVA coupled to the hapten 4-hydroxy-3-nitrophenylacetyl (NP-OVA) precipitated in alum (Imject® Alum, Thermo Scientific) into the hind footpads (25ul). Single cell suspensions were obtained by forcing popliteal lymph nodes through a 70 µm mesh into ice cold FACS buffer (EDTA 1 mM and 2% serum in PBS). Cells were incubated with 2 µg/ml anti-16/32 (clone 93) for blockage of FC receptors for 5-10 min. Cell suspensions were washed and incubated with fluorescently-labeled antibodies (B220 V500, FAS FITC, CD38 Alexa fluor 700; Biolegend) for 20-40 min. GC cells were gated as

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

live/single, $B220^{+}$ $CD38^{Lo}$ FAS^{Hi}. Cell suspensions were analyzed by Cytoflex (Beckman) flow cytometer.

Cell Viability Assays: Growth Over Time in 2D-Adherent Monolayer Cell Culture.

The indicated cell lines were plated at a density of 500 cells per well (except PATU-8988T cells, which were plated at 100 cells per well to avoid over-confluence by day 8) in 100 µL of media in a 96-well white clear bottom plate (Corning cat#3903), with one plate per time point (Day 0, 2, 4, 6, 8). Cells were treated the day after plating with 1 µL of DMSO, Sulfopin, or Sulfopin-AcA to give the indicated concentrations, and were then incubated at 37°C 5% CO_2. Every 48 h, the media was aspirated and replaced with fresh media containing fresh compound or DMSO. When the indicated time points had been reached, cell viability was evaluated using the CellTiter-Glo Luminescent Cell Viability Assay (Promega cat#G7570) according to the manufacturer's standards, measuring luminescence using an Envision plate-reader. The Day 0 time point plates were read the day after plating, prior to compound treatment. N=3 biological replicates were used for each treatment condition.

Cell Viability Assay: 5 day treatment

PATU-8988T cells were plated in flat bottom 96-well plates (Corning cat #3903) at a density of 1,000 cells per well in 100 µL media and were treated the next day with 1 µL of the indicated compounds in a three-fold dilution series. The cells were incubated at 37°C 5% CO_2 for 5 days. Anti-proliferative effects were assessed by CellTiter-Glo Luminescent Cell Viability Assay (Promega cat #G7570) according to the manufacturer's standards, measuring luminescence using an Envision plate-reader. N=3 biological replicates were used for each treatment condition.

Assessing Antiproliferative Activity in PATU-8988T 3D Cell Culture.

Prepare PATU-8988T (WT or $Pin1^{-/-}$) Matrigel suspensions by resuspending cells in 100% cold Matrigel (kept on ice). Plate 1 dome per well in 24-well plate (Greiner CELLSTAR), with 50 µL of cells/Matrigel suspension per dome, and 1,000 cells per dome. After plating the domes, place 24-well plate on top of T175 (previously filled with autoclaved water) in the incubator, let solidify for 15 min. Keeping the 24-well plate on the T175 filled with water, move to the TC hood and carefully add 500 µL of cold DMEM (+10% FBS/1% P/S) per well, then place 24-well plate in incubator. The next day, treat with DMSO, Sulfopin (1 µM), or Sulfopin-AcA (1 µM). Every 3 days, carefully aspirate off the media, add 500 µL of fresh media and retreat. After 9 days, aspirate off the media, add 300 µL of 3D CellTiter-Glo (Promega cat#G9681) per well, shake plate for 1 h. Measure

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

luminescence using an Envision plate-reader. N=3 biological replicates were used for each treatment condition.

RNA sequencing

Mino cells (acquired from the ATCC) were grown at 37°C in a 5% CO_2 humidified incubator and cultured in RPMI-1640 (Biological industries), supplemented with 15% Fetal bovine serum (biological industries) and 1% pen-strep solution (biological industries). 11×10^6 cells were incubated with 1 µM Sulfopin (0.02% DMSO) or with 0.02% DMSO in triplicates for 6 h. Total RNA was isolated with RNeasy kit (Qiagen). RNA libraries were prepared from 2 µg total RNA using SENSE mRNA-Seq library prep kit V2 (lexogen). Total RNA and library quality was analyzed using Qubit fluorometric and TapeStation analysis (Agilent). Samples were sequenced using NextSeq 500/550 High Output Kit v2.5 (illumina) on NextSeq550.

RNA-seq reads were aligned to the human genome (hg19 assembly) using STAR[111] and gene expression was determined using RSEM[112] and RefSeq annotations. Differential expression was computed using DESeq2[113] with default parameters. Genes with baseMean >50 that were downregulated with $P<0.05$ were further analysed using Enrichr[89].

Profiling of Sulfopin-DTB reactive cysteines by CITe-Id

PATU-8988T cells were cultured in DMEM supplemented with 10% Fetal Calf Serum. DMSO (Control) or Sulfopin were diluted in fresh media (final DMSO concentration < 0.1%, final Sulfopin concentration: 100 nM, 500 nM and 1 µM) and added to sub-confluent cultures. After 5 h of incubation at 37°C, cells were harvested using a cell scraper and centrifuged at 300 × g for 3 minutes at 4°C. Cell pellets were washed with ice-cold PBS and centrifuged again. A total of 3 washes were performed before freezing cell pellets at -80°C. This procedure was performed twice on cells independently cultured a week apart. Frozen pellets were then processed essentially as described[79], except that pre-cleared lysates were treated with 2 µM Sulfopin-DTB overnight at 4 °C. Protein desalting, digestion, enrichment of desthiobiotin modified peptides, iTRAQ stable isotope labeling, peptide clean-up, and multidimensional LC/MS-MS analysis were then performed exactly as described[79]. Data processing and database search was performed as described[79], except that spectral processing accounted for Sulfopin-DTB specific fragment ions, and Sulfopin-DTB labeling of cysteine was considered as a variable modification by Mascot (version 2.6.2). Inhibitor concentrations and ratios were used to generate a trendline for each labeled site with the slope corresponding to the competitive dose response for each modified cysteine site.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Profiling of Sulfopin reactive cysteines by rdTOP-ABPP

MDA-MB-231 cells were cultured at 37°C under 5% CO_2 atmosphere in DMEM culture medium supplemented with 10% FBS and 1% PS. Cells were grown to 70% confluence and incubated with DMSO or 5 μM Sulfopin for 2 h with serum-free medium. Cells were harvested, lysed by sonication in ice-cold PBS containing 0.1% TritonX-100 and centrifuged at 100,000g for 30 min to remove cell debris. Then protein concentrations were determined by BCA protein assay. Proteomes were normalized to 2 mg/mL in 1 mL for each sample.

Each of the DMSO and Sulfopin incubated proteomes was treated with 100 μM iodoacetamide-alkyne (IAyne) for 1 h at room temperature. The proteomes were then reacted with 1 mM CuSO4, 100 μM TBTA ligand, 100 μM biotin-acid-N3 tag and 1 mM TCEP for 1 h. After click reaction, the proteomes were centrifuged at 8000g for 5 min, then the precipitated proteins were washed for two times using cold methanol. The proteomes were resuspended in 1.2% SDS/PBS and diluted to 0.2% SDS/PBS. Finally, the samples were prepared, analyzed on LC-MS/MS and quantified according to the published rdTOP-ABPP protocol[80]. Briefly, the beads from trypsin digestion were washed and resuspended in 100 μL of TEAB buffer. 8 μL of 4% $D^{13}CDO$ or HCHO was added to the Sulfopin or DMSO sample respectively. At the same time, 8 μL of 0.6 M $NaBH_3CN$ was added and the reaction was lasted for 2 h at room temperature. Then the beads were washed again and the modified peptides were cleaved by 2% formic acid. LC-MS/MS data was analyzed by ProLuCID[114] with static modification of cysteine (+57.0215 Da) and variable oxidation of methionine (+15.9949 Da). The isotopic modifications (+28.0313 and +34.0631 Da for light and heavy labeling respectively) are set as static modifications on the N-terminal of a peptide and lysines. Variable modification on cysteines is set at +322.23688 Da. The ratios were quantified by the CIMAGE software[115].

Myc luciferase reporter assay

HEK-293 cells were maintained in DMEM supplemented with 10% standard fetal bovine serum (FBS), 2.5 mM L-glutamine, NEEA and 1x penicillin/streptomycin. Cells were passaged to 80% confluence in 6-well plates and transfected with 4x-Ebox-Luc and Pin1-Flag plasmids as indicated and β-galactosidase as an internal control[22] using Lipofectamine 3000 (Invitrogen, Carlsbad, CA) following manufacturer's protocol. Sulfopin treatments were performed at the time of transfection and cells were harvested 48 h later.

Cell were washed with PBS, and then lysed in 1X cell lysis buffer (Promega, Madison, WI). Lysates were sonicated for 10 pulses at an output = 1 and 10% duty (Branson), and incubated on ice for 20 min. Lysates were then cleared by centrifugation at 14K rpm for 10 min at 4ºC.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Luciferase activity was measured using the Promega Luciferase Assay Kit and Berthold luminometer (Bundoora, Australia) and normalized to β-galactosidase activity [116].

Zebrafish Neuroblastoma Models

All zebrafish studies and maintenance of the animals were performed in accordance with Dana-Farber Cancer Institute IACUC-approved protocol #02-107. For *in vivo* drug treatment, 3-day-old zebrafish embryos were placed in 48-well plates with 5 embryos per well, and treated with DMSO control or Sulfopin in standard egg water.

For neuroblastoma transplantation, zebrafish neuroblastoma cells were harvested by dicing 4-month-old Tg(dβh:MYCN;dβh:EGFP) transgenic zebrafish in PBS. The cell suspension was filtered with Falcon 40-μm cell strainer (Corning, Corning, NY, USA) and loaded into thin-wall borosilicate glass capillary needles (1.0 mm OD, 0.75 mm ID; World Precision Instruments, Sarasota, FL, USA). The recipients, 2-day-old zebrafish Casper embryos, were manually dechorionized and anaesthetized with 0.003% tricaine (Sigma) before being positioned on a 10 cm Petri dish coated with 1% agarose. Intravenous tumor transplantation was performed as described[94], with 200~400 cells injected into the Duct of Cuvier of each recipient. One day later, the 3-day-old recipients were randomly divided into 48-well plates and treated with DMSO control or Sulfopin in standard egg water for five days, with drug refreshment on the second day of the treatment.

A Nikon SMZ1500 microscope equipped with a Nikon digital sight DS-U1 camera was used for capturing both the bright field and fluorescent images from live zebrafish and embryos. For PSNS and neuroblastoma quantification, all animals in the same experiments were imaged under the same conditions and the acquired fluorescent images were quantified using ImageJ software by measuring the area of the EGFP fluorescent tumor mass.

Mice studies (PK/PD/Tox)

PK data was obtained as fee-for-service from Scripps Florida. For the toxicity study 18 mice were used for the following arms: 3 control mice injected with vehicle, 3 mice 10 mg/kg Sulfopin injected once daily, 3 mice 20 mg/kg Sulfopin injected once daily, 3 mice 20 mg/kg Sulfopin injected once every other day, 3 mice 40 mg/kg Sulfopin injected once daily, 3 mice 40 mg/kg Sulfopin injected once every other day. Formulation was (5% NMP, 5% solutol, 20% DMSO). Mice were treated for 14 days.

PD Study was performed at the Dana-Farber Cancer Institute, and the procedure was approved by IACUC under protocol 16-015. Mice were treated for 3 total doses spanning 2

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

consecutive days with vehicle, or Sulfopin (10 or 20 mpk) by oral gavage, after which the organs were harvested 4 h after the last dose. The spleen of each mouse was ground and lysed in 300 µL of 50 mM Hepes, pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM TCEP, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor tablet (Roche). After clarifying (14,000 rpm for 15 min), the lysates were normalized by BCA and diluted to a final concentration of 2.5 µg/µL. 200 µL of 2.5 µg/µL of each spleen sample was then incubated with 1 µM of Sulfopin-DTB for 1 h at 4°C and processed as in "lysate pull-down with Sulfopin-DTB" (above). To prepare the input samples, each original lysate sample (prior to pull-down) was combined with 4x LDS + 10% β-mercaptoethanol (in a ratio of 3:1), boiled for 5 minutes, and 25 µg of each sample was then loaded on the gel.

Murine Neuroblastoma models

All experiments were approved by The Institute of Cancer Research Animal Welfare and Ethical Review Body and performed in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986, the United Kingdom National Cancer Research Institute guidelines for the welfare of animals in cancer research[117] and the ARRIVE (animal research: reporting *in vivo* experiments) guidelines[118].

Transgenic Th-*MYCN* mice were genotyped to detect the presence of the human *MYCN* transgene[119]. The study was performed using both male and female hemizygous mice, which developed palpable tumors at 50 to 130 days with a 25% penetrance. Tumor development was monitored weekly by palpation by an experienced animal technician. Mice with palpable tumors at >/= 4-5 mm were then enrolled into two groups. Group 1: Animals will all received Sulfopin at 40mg/kg, once per day by oral gavage. Group 2: Animals will all received Vehicle (5% NMP, 5% kolliphor, 20% DMSO) once per day by oral gavage. Studies were terminated when the mean diameter of the tumor reached 15 mm. Tumor volumes were measured by Vernier caliper across two perpendicular diameters, and volumes were calculated according to the formula: $V = 4/3\pi [(d1 + d2) / 4]^3$. Mice were housed in specific pathogen-free rooms in autoclaved, aseptic microisolator cages (maximum of four mice per cage).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Acknowledgments

N.L. is the incumbent of the Alan and Laraine Fischer Career Development Chair; N.L. would like to acknowledge funding from the Israel Science Foundation (grant No. 2462/19), The Rising Tide Foundation, The Israel Cancer Research Fund, the Israeli Ministry of Science. Technology (grant No. 3-14763), and the Moross integrated cancer center. N.L. is also supported by the Helen and Martin Kimmel Center for Molecular Design, Joel and Mady Dukler Fund for Cancer Research, the Estate of Emile Mimran and Virgin JustGiving and the George Schwartzman Fund. C.D. was supported by the Minerva Fellowship program of the Max Planck Society, funded by the German Federal Ministry for Education and Research. The work was supported in part by NIH grant R01CA205153 to K.P.L., N. S. G. and X. Z. Z. N.S.G. was also supported by the Hale Center for Pancreatic Research. Y.C. & C.W. thank the Computing Platform of the Center for Life Science at Peking University for supporting the proteomic data analysis. S.D.P acknowledges funding from the Linde Family Foundation. Part of this research was conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines (NIGMS P41 GM103403) and SBGrid compiled software (eLife 2013;2:e01456). J.A.M. acknowledges support from the National Institutes of Health (CA233800) and the Mark Foundation for Cancer Research. B.J.P. was supported by the Ruth L. Kirschstein NRSA Individual Predoctoral Fellowship (F31 CA225066), the Training Grant in Chemical Biology (NIH 5 T32 GM007306-41), the Training Grant in Pharmacological Sciences (NIH 5 T32 GM095450-04, B.J.P.), and the Chleck Foundation (B.J.P./Z.M.D.). B.N. was supported by an American Cancer Society Postdoctoral Fellowship (PF-17-010-01-CDD) and the Katherine L. and Steven C. Pinard Research Fund (N.S.G. and B.N.). L. C. is supported by the Cancer Research UK Program Grant (C34648/A18339 and C34648/A14610). Y. J. is supported by a Children with Cancer UK Research Fellowship (2014/176). R.C.S. acknowledges funding support from NCI R01s CA196228 and CA186241 and foundation support from The Brenden-Colson Center for Pancreatic Care. We thank Igor Ulitski for help with RNA-seq analysis and Milka Kostic for critical reading of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Conflict of Interests

N.S.G. is a Scientific Founder and member of the Scientific Advisory Board (SAB) of C4, 506 Therapeutics, Syros, Soltego, Gatekeeper and Petra Pharmaceuticals and has received research funding from Novartis, Astellas, Taiho and Deerfield. N.L. is a member of the SAB of Trilogy Sciences, and has received research support from Teva. J.A.M. has received support through sponsored research agreements with AstraZeneca and Vertex. J.A.M. serves on the SAB of 908 Devices. C.M.B. is an employee of AstraZeneca. C.D., B.J.P., D.Z., S.H., X.L., K.P.L., X.Z.Z., A.T.L., N.S.G. and N.L. are inventors on a patent application related to the inhibitors described in this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Pawson, T. & Scott, J. D. Protein phosphorylation in signaling--50 years and counting. *Trends Biochem. Sci.* **30**, 286–290 (2005).
3. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
4. Lu, K. P. & Zhou, X. Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat. Rev. Mol. Cell Biol.* **8**, 904–916 (2007).
5. Lam, P. B. *et al.* Prolyl isomerase Pin1 is highly expressed in Her2-positive breast cancer and regulates erbB2 protein stability. *Mol. Cancer* **7**, 91 (2008).
6. Liao, Y. *et al.* Peptidyl-prolyl cis/trans isomerase Pin1 is critical for the regulation of PKB/Akt stability and activation phosphorylation. *Oncogene* **28**, 2436–2445 (2009).
7. Lee, T. H. *et al.* Essential role of Pin1 in the regulation of TRF1 stability and telomere maintenance. *Nat. Cell Biol.* **11**, 97–105 (2009).
8. Ryo, A., Nakamura, M., Wulf, G., Liou, Y. C. & Lu, K. P. Pin1 regulates turnover and subcellular localization of beta-catenin by inhibiting its interaction with APC. *Nat. Cell Biol.* **3**, 793–801 (2001).
9. Ryo, A. *et al.* Regulation of NF-kappaB signaling by Pin1-dependent prolyl isomerization and ubiquitin-mediated proteolysis of p65/RelA. *Mol. Cell* **12**, 1413–1426 (2003).
10. Shen, M., Stukenberg, P. T., Kirschner, M. W. & Lu, K. P. The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes Dev.* **12**, 706–720 (1998).
11. Stukenberg, P. T. & Kirschner, M. W. Pin1 acts catalytically to promote a conformational change in Cdc25. *Mol. Cell* **7**, 1071–1083 (2001).
12. Xiang, K. *et al.* Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

13. Zhou, X. Z. *et al.* Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol. Cell* **6**, 873–883 (2000).
14. Brown, N. R., Noble, M. E., Endicott, J. A. & Johnson, L. N. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat. Cell Biol.* **1**, 438–443 (1999).
15. Chen, Y. *et al.* Prolyl isomerase Pin1: a promoter of cancer and a target for therapy. *Cell Death Dis.* **9**, 883 (2018).
16. Zhou, X. Z. & Lu, K. P. The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nat. Rev. Cancer* **16**, 463–478 (2016).
17. Bao, L. *et al.* Prevalent overexpression of prolyl isomerase Pin1 in human cancers. *Am. J. Pathol.* **164**, 1727–1737 (2004).
18. Lu, K. P. Prolyl isomerase Pin1 as a molecular target for cancer diagnostics and therapeutics. *Cancer Cell* **4**, 175–180 (2003).
19. Tan, X. *et al.* Pin1 expression contributes to lung cancer: Prognosis and carcinogenesis. *Cancer Biol. Ther.* **9**, 111–119 (2010).
20. Li, Q. *et al.* The rs2233678 polymorphism in PIN1 promoter region reduced cancer risk: a meta-analysis. *PLoS One* **8**, e68148 (2013).
21. Rustighi, A. *et al.* The prolyl-isomerase Pin1 is a Notch1 target that enhances Notch1 activation in cancer. *Nature Cell Biology* vol. 11 133–142 (2009).
22. Farrell, A. S. *et al.* Pin1 regulates the dynamics of c-Myc DNA binding to facilitate target gene regulation and oncogenesis. *Mol. Cell. Biol.* **33**, 2930–2949 (2013).
23. Brenkman, A. B. *et al.* The peptidyl-isomerase Pin1 regulates p27kip1 expression through inhibition of Forkhead box O tumor suppressors. *Cancer Res.* **68**, 7597–7605 (2008).
24. Basu, A. *et al.* Proteasomal degradation of human peptidyl prolyl isomerase pin1-pointing phospho Bcl2 toward dephosphorylation. *Neoplasia* **4**, 218–227 (2002).
25. Boldetti, A., Guarnaccia, V., Rambaldi, A. & Parrella, E. Inhibition of the peptidyl-prolyl-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

isomerase Pin1 enhances the responses of acute myeloid leukemia cells to retinoic acid via stabilization of RARα and PML-RARα. *Cancer Res.* (2009).

26. Girardini, J. E. *et al.* A Pin1/Mutant p53 Axis Promotes Aggressiveness in Breast Cancer. *Cancer Cell* vol. 20 79–91 (2011).
27. Wulf, G., Garg, P., Liou, Y.-C., Iglehart, D. & Lu, K. P. Modeling breast cancer in vivo and ex vivo reveals an essential role of Pin1 in tumorigenesis. *EMBO J.* **23**, 3397–3407 (2004).
28. D'Artista, L. *et al.* Pin1 is required for sustained B cell proliferation upon oncogenic activation of Myc. *Oncotarget* **7**, 21786–21798 (2016).
29. Zheng, M. *et al.* Inhibition of the prolyl isomerase Pin1 enhances the ability of sorafenib to induce cell death and inhibit tumor growth in hepatocellular carcinoma. *Oncotarget* **8**, 29771–29784 (2017).
30. Ding, Q. *et al.* Down-regulation of myeloid cell leukemia-1 through inhibiting Erk/Pin 1 pathway by sorafenib facilitates chemosensitization in breast cancer. *Cancer Res.* **68**, 6109–6117 (2008).
31. Liu, P. H. *et al.* An IRAK1-PIN1 signalling axis drives intrinsic tumour resistance to radiation therapy. *Nat. Cell Biol.* **21**, 203–213 (2019).
32. Luo, M.-L. *et al.* The Rab2A GTPase promotes breast cancer stem cells and tumorigenesis via Erk signaling activation. *Cell Rep.* **11**, 111–124 (2015).
33. Rustighi, A. *et al.* Prolyl-isomerase Pin1 controls normal and cancer stem cells of the breast. *EMBO Mol. Med.* **6**, 99–119 (2014).
34. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).
35. Liou, Y.-C. *et al.* Loss of Pin1 function in the mouse causes phenotypes resembling cyclin D1-null phenotypes. *Proceedings of the National Academy of Sciences* vol. 99 1335–1340 (2002).
36. Hennig, L. *et al.* Selective inactivation of parvulin-like peptidyl-prolyl cis/trans isomerases by juglone. *Biochemistry* **37**, 5953–5960 (1998).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

37. Wei, S. *et al.* Active Pin1 is a key target of all-trans retinoic acid in acute promyelocytic leukemia and breast cancer. *Nat. Med.* **21**, 457–466 (2015).
38. Kozono, S. *et al.* Arsenic targets Pin1 and cooperates with retinoic acid to inhibit cancer-driving pathways and tumor-initiating cells. *Nat. Commun.* **9**, 3069 (2018).
39. Campaner, E. *et al.* A covalent PIN1 inhibitor selectively targets cancer cells by a dual mechanism of action. *Nat. Commun.* **8**, 15772 (2017).
40. Lu, Z. & Hunter, T. Prolyl isomerase Pin1 in cancer. *Cell Res.* **24**, 1033–1049 (2014).
41. Moore, J. D. & Potter, A. Pin1 inhibitors: Pitfalls, progress and cellular pharmacology. *Bioorg. Med. Chem. Lett.* **23**, 4283–4291 (2013).
42. Fila, C., Metz, C. & van der Sluijs, P. Juglone inactivates cysteine-rich proteins required for progression through mitosis. *J. Biol. Chem.* **283**, 21714–21724 (2008).
43. Pinch, B. J. *et al.* Discovery and Characterization of a Potent and Selective Pin1 Inhibitor Targeted to an Active Site Cysteine. *Nat. Chem. Biol.* (2020) doi:In press.
44. Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chem. Int. Ed Engl.* **55**, 13408–13421 (2016).
45. Noe, M. C. & Gilbert, A. M. Targeted covalent enzyme inhibitors. in *Annual Reports in Medicinal Chemistry* vol. 47 413–439 (Elsevier, 2012).
46. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med. Chem.* **2**, 949–964 (2010).
47. Ghosh, A. K., Samanta, I., Mondal, A. & Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **14**, 889–906 (2019).
48. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
49. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* **4**, 1046–1061 (2014).
50. Ward, R. A. *et al.* Structure- and reactivity-based development of covalent inhibitors of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

activating and gatekeeper mutant forms of the epidermal growth factor receptor (EGFR). *J. Med. Chem.* **56**, 7025–7048 (2013).

51. Bauer, R. A. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **20**, 1061–1073 (2015).
52. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
53. Patricelli, M. P. *et al.* Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discov.* **6**, 316–329 (2016).
54. Janes, M. R. *et al.* Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell* **172**, 578–589.e17 (2018).
55. Shin, Y. *et al.* Discovery of N-(1-Acryloylazetidin-3-yl)-2-(1H-indol-1-yl)acetamides as Covalent Inhibitors of KRASG12C. *ACS Med. Chem. Lett.* **10**, 1302–1308 (2019).
56. Leach, A. R., Hann, M. M., Burrows, J. N. & Griffen, E. J. Fragment screening: an introduction. *Molecular BioSystems* vol. 2 429 (2006).
57. Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* **12**, 5–7 (2013).
58. Erlanson, D. A., McDowell, R. S. & O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **47**, 3463–3482 (2004).
59. Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**, 4990–5003 (2012).
60. Hall, R. J., Mortenson, P. N. & Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **116**, 82–91 (2014).
61. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856–864 (2001).
62. Resnick, E. *et al.* Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Am. Chem. Soc. **141**, 8951–8968 (2019).

63. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proceedings of the National Academy of Sciences* vol. 97 9367–9372 (2000).
64. Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: Fragment-Based Drug Discovery. *Annual Review of Biophysics and Biomolecular Structure* vol. 33 199–223 (2004).
65. Johansson, H. *et al.* Fragment-Based Covalent Ligand Screening Enables Rapid Discovery of Inhibitors for the RBR E3 Ubiquitin Ligase HOIP. *J. Am. Chem. Soc.* **141**, 2703–2712 (2019).
66. Kathman, S. G., Xu, Z. & Statsyuk, A. V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J. Med. Chem.* **57**, 4969–4974 (2014).
67. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).
68. Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527–541.e29 (2017).
69. Roberts, A. M. *et al.* Chemoproteomic Screening of Covalent Ligands Reveals UBA5 As a Novel Pancreatic Cancer Target. *ACS Chem. Biol.* **12**, 899–904 (2017).
70. Grossman, E. A. *et al.* Covalent Ligand Discovery against Druggable Hotspots Targeted by Anti-cancer Natural Products. *Cell Chem Biol* **24**, 1368–1376.e4 (2017).
71. Counihan, J. L., Wiggenhorn, A. L., Anderson, K. E. & Nomura, D. K. Chemoproteomics-Enabled Covalent Ligand Screening Reveals ALDH3A1 as a Lung Cancer Therapy Target. *ACS Chem. Biol.* **13**, 1970–1977 (2018).
72. Anderson, K. E., To, M., Olzmann, J. A. & Nomura, D. K. Chemoproteomics-Enabled Covalent Ligand Screening Reveals a Thioredoxin-Caspase 3 Interaction Disruptor That Impairs Breast Cancer Pathogenicity. *ACS Chemical Biology* vol. 12 2522–2528 (2017).
73. London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10**, 1066–1072 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

74. Zhang, Y. *et al.* Structural basis for high-affinity peptide inhibition of human Pin1. *ACS Chem. Biol.* **2**, 320–328 (2007).
75. Yaffe, M. B. *et al.* Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **278**, 1957–1960 (1997).
76. Guo, C. *et al.* Structure-based design of novel human Pin1 inhibitors (I). *Bioorg. Med. Chem. Lett.* **19**, 5613–5616 (2009).
77. Dong, L. *et al.* Structure-based design of novel human Pin1 inhibitors (II). *Bioorg. Med. Chem. Lett.* **20**, 2210–2214 (2010).
78. Guo, C. *et al.* Structure-based design of novel human Pin1 inhibitors (III): optimizing affinity beyond the phosphate recognition pocket. *Bioorg. Med. Chem. Lett.* **24**, 4187–4191 (2014).
79. Browne, C. M. *et al.* A Chemoproteomic Strategy for Direct and Proteome-Wide Covalent Inhibitor Target-Site Identification. *J. Am. Chem. Soc.* **141**, 191–203 (2019).
80. Yang, F., Gao, J., Che, J., Jia, G. & Wang, C. A Dimethyl-Labeling-Based Strategy for Site-Specifically Quantitative Chemical Proteomics. *Anal. Chem.* **90**, 9576–9582 (2018).
81. Phan, R. T., Saito, M., Kitagawa, Y., Means, A. R. & Dalla-Favera, R. Genotoxic stress regulates expression of the proto-oncogene Bcl6 in germinal center B cells. *Nat. Immunol.* **8**, 1132–1139 (2007).
82. Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016).
83. Chao, S. H., Greenleaf, A. L. & Price, D. H. Juglone, an inhibitor of the peptidyl-prolyl isomerase Pin1, also directly blocks transcription. *Nucleic Acids Res.* **29**, 767–773 (2001).
84. Baker, L. A., Tiriac, H., Clevers, H. & Tuveson, D. A. Modeling pancreatic cancer with organoids. *Trends Cancer Res.* **2**, 176–190 (2016).
85. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* **2**, 764–776 (2002).
86. Yeh, E. *et al.* A signalling pathway controlling c-Myc degradation that impacts oncogenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

transformation of human cells. *Nat. Cell Biol.* **6**, 308–318 (2004).

87. Su, Y. *et al.* Post-translational modification localizes MYC to the nuclear pore basket to regulate a subset of target genes involved in cellular responses to environmental signals. *Genes Dev.* **32**, 1398–1419 (2018).
88. Helander, S. *et al.* Pre-Anchoring of Pin1 to Unphosphorylated c-Myc in a Fuzzy Complex Regulates c-Myc Activity. *Structure* **23**, 2267–2279 (2015).
89. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).
90. Zhu, S. *et al.* Activated ALK collaborates with MYCN in neuroblastoma pathogenesis. *Cancer Cell* **21**, 362–373 (2012).
91. He, S. *et al.* Synergy between loss of NF1 and overexpression of MYCN in neuroblastoma is mediated by the GAP-related domain. *Elife* **5**, (2016).
92. Zimmerman, M. W. *et al.* MYC Drives a Subset of High-Risk Pediatric Neuroblastomas and Is Activated through Mechanisms Including Enhancer Hijacking and Focal Enhancer Amplification. *Cancer Discov.* **8**, 320–335 (2018).
93. Campbell, K. *et al.* Association of MYCN copy number with clinical features, tumor biology, and outcomes in neuroblastoma: A report from the Children's Oncology Group. *Cancer* **123**, 4224–4235 (2017).
94. He, S. *et al.* Neutrophil-mediated experimental metastasis is enhanced by VEGFR inhibition in a zebrafish xenograft model. *J. Pathol.* **227**, 431–445 (2012).
95. Chesler, L. & Weiss, W. A. Genetically engineered murine models – Contribution to our understanding of the genetics, molecular pathology and therapeutic targeting of neuroblastoma. *Seminars in Cancer Biology* vol. 21 245–255 (2011).
96. Lanning, B. R. *et al.* A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **10**, 760–767 (2014).
97. Uchida, T. *et al.* Pin1 and Par14 peptidyl prolyl isomerase inhibitors block cell proliferation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Chem. Biol. **10**, 15–24 (2003).

98. Tatara, Y., Lin, Y.-C., Bamba, Y., Mori, T. & Uchida, T. Dipentamethylene thiuram monosulfide is a novel inhibitor of Pin1. *Biochem. Biophys. Res. Commun.* **384**, 394–398 (2009).
99. Yoon, H.-E. *et al.* Inhibition of Plk1 and Pin1 by 5'-nitro-indirubinoxime suppresses human lung cancer cells. *Cancer Lett.* **316**, 97–104 (2012).
100. Subedi, A. *et al.* Discovery of novel selenium derivatives as Pin1 inhibitors by high-throughput screening. *Biochem. Biophys. Res. Commun.* **474**, 528–533 (2016).
101. Lee, N. Y. *et al.* The prolyl isomerase Pin1 interacts with a ribosomal protein S6 kinase to enhance insulin-induced AP-1 activity and cellular transformation. *Carcinogenesis* **30**, 671–681 (2009).
102. Kanaoka, R. *et al.* Pin1 Inhibitor Juglone Exerts Anti-Oncogenic Effects on LNCaP and DU145 Cells despite the Patterns of Gene Regulation by Pin1 Differing between These Cell Lines. *PLoS One* **10**, e0127467 (2015).
103. Fan, G. *et al.* Peptidyl-prolyl isomerase Pin1 markedly enhances the oncogenic activity of the rel proteins in the nuclear factor-kappaB family. *Cancer Res.* **69**, 4589–4597 (2009).
104. Cole, K. A. *et al.* RNAi screen of the protein kinome identifies checkpoint kinase 1 (CHK1) as a therapeutic target in neuroblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3336–3341 (2011).
105. Hogarty, M. D. & Maris, J. M. PI3King on MYCN to improve neuroblastoma therapeutics. *Cancer cell* vol. 21 145–147 (2012).
106. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
107. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
108. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.20.998443. this version posted March 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

109. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

110. Auld, D. S. *et al.* Receptor Binding Assays for HTS and Drug Discovery. in *Assay Guidance Manual* (eds. Sittampalam, G. S. et al.) (Eli Lilly & Company and the National Center for AdvancingTranslational Sciences, 2012).

111. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

112. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

113. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

114. Xu, T. *et al.* ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J. Proteomics* **129**, 16–24 (2015).

115. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).

116. Sears, R., Ohtani, K. & Nevins, J. R. Identification of positively and negatively acting elements regulating expression of the E2F2 gene in response to cell growth signals. *Mol. Cell. Biol.* **17**, 5227–5235 (1997).

117. Workman, P. *et al.* Guidelines for the welfare and use of animals in cancer research. *British Journal of Cancer* vol. 102 1555–1577 (2010).

118. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *J. Pharmacol. Pharmacother.* **1**, 94–99 (2010).

119. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J.* **16**, 2985–2995 (1997).

Sulfopin, a selective covalent inhibitor of Pin1, blocks Myc-driven tumor initiation and growth *in vivo*

Christian Dubiella[1,*], Benika J. Pinch[2,3,4,*], Daniel Zaidman[1], Theresa D. Manz[2,3,5], Evon Poon[6], Shuning He[7], Efrat Resnick[1], Ellen M. Langer[8,9], Colin J. Daniel[8,9], Hyuk-Soo Seo[2], Ying Chen[10], Scott B. Ficarro[2,11,12,13], Yann Jamin[14], Xiaolan Lian[15,16,17], Shin Kibe[15,16,17], Shingo Kozono[15,16,17], Kazuhiro Koikawa[15,16,17], Zainab M. Doctor[2,3], Behnam Nabet[2,3], Christopher M. Browne[2,3,18], Annan Yang[2,19], Liat Stoler-Barak[20], Richa B. Shah[21,22], Nick E. Vangos[2], Ezekiel A. Geffken[2], Roni Oren[23], Samuel Sidi[21,22], Ziv Shulman[20], Chu Wang[10], Jarrod A. Marto[2,11,12,13], Sirano Dhe-Paganon[2], Thomas Look[7,24], Xiao Zhen Zhou[15,16,17], Kun Ping Lu[15,16,17], Rosalie C. Sears[8,9,25], Louis Chesler[6], Nathanael S. Gray[2,3,#], Nir London[1,#]

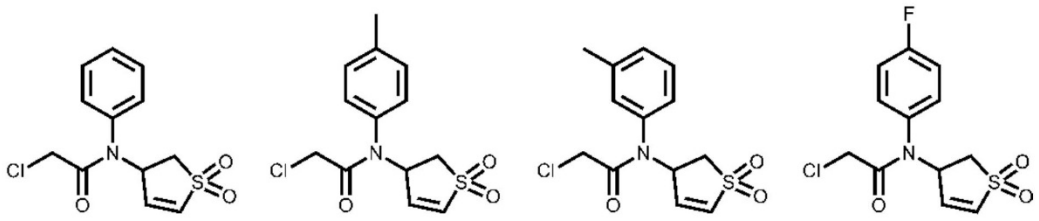
[1] Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[2] Department of Cancer Biology, Dana–Farber Cancer Institute, Boston, MA 02115, USA.
[3] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA.
[4] Department of Chemistry and Chemical Biology; Department of Chemical Biology, Harvard University, Cambridge, MA, 02138, USA.
[5] Department of Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbruecken, Germany.
[6] Division of Clinical Studies, The Institute of Cancer Research, London, UK.
[7] Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[8] Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon.
[9] Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon.
[10] College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.
[11] Department of Oncologic Pathology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[12] Blais Proteomics Center, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, 02115, USA.
[13] Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA.
[14] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK.
[15] Department of Medicine, Division of Translational Therapeutics, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA.
[16] Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA.
[17] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
[18] Currently at: Discovery Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Boston, MA USA
[19] Department of Medical Oncology, Dana–Farber Cancer Institute, Boston, MA 02115, USA.
[20] Department of Immunology, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[21] Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[22] Department of Cell, Developmental and Regenerative Biology, The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[23] Department of Veterinary Resources, The Weizmann Institute of Science, Rehovot, 7610001, Israel.
[24] Division of Pediatric Hematology/Oncology Boston Children's Hospital, Boston, MA, 02115, USA.
[25] Brenden-Colson Center for Pancreatic Care, Oregon Health & Science University, Portland, Oregon.

[*] equal contribution.
[#] e-mail: nir.london@weizmann.ac.il; nathanael_gray@dfci.harvard.edu

Keywords: covalent fragments; covalent probes; Pin1; neuroblastoma; Myc

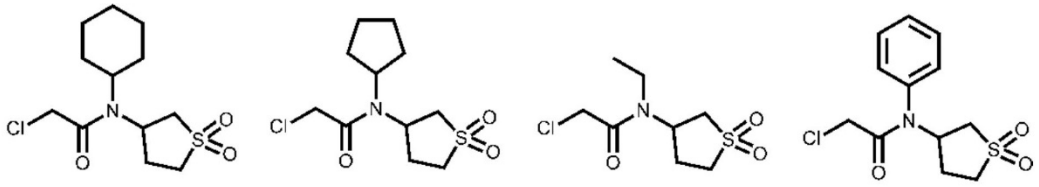
Supplementary Figures

2-Sulfolenes:

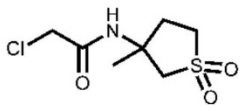


PCM-0102372 PCM-0102539 PCM-0102579 PCM-0102868

Sulfolanes:

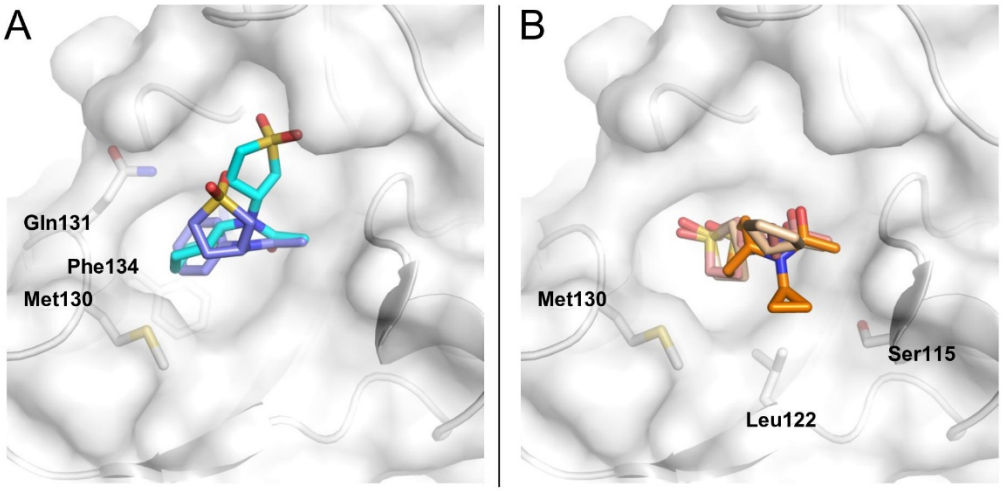


PCM-0102760 PCM-0102313 PCM-0102105 PCM-0102755

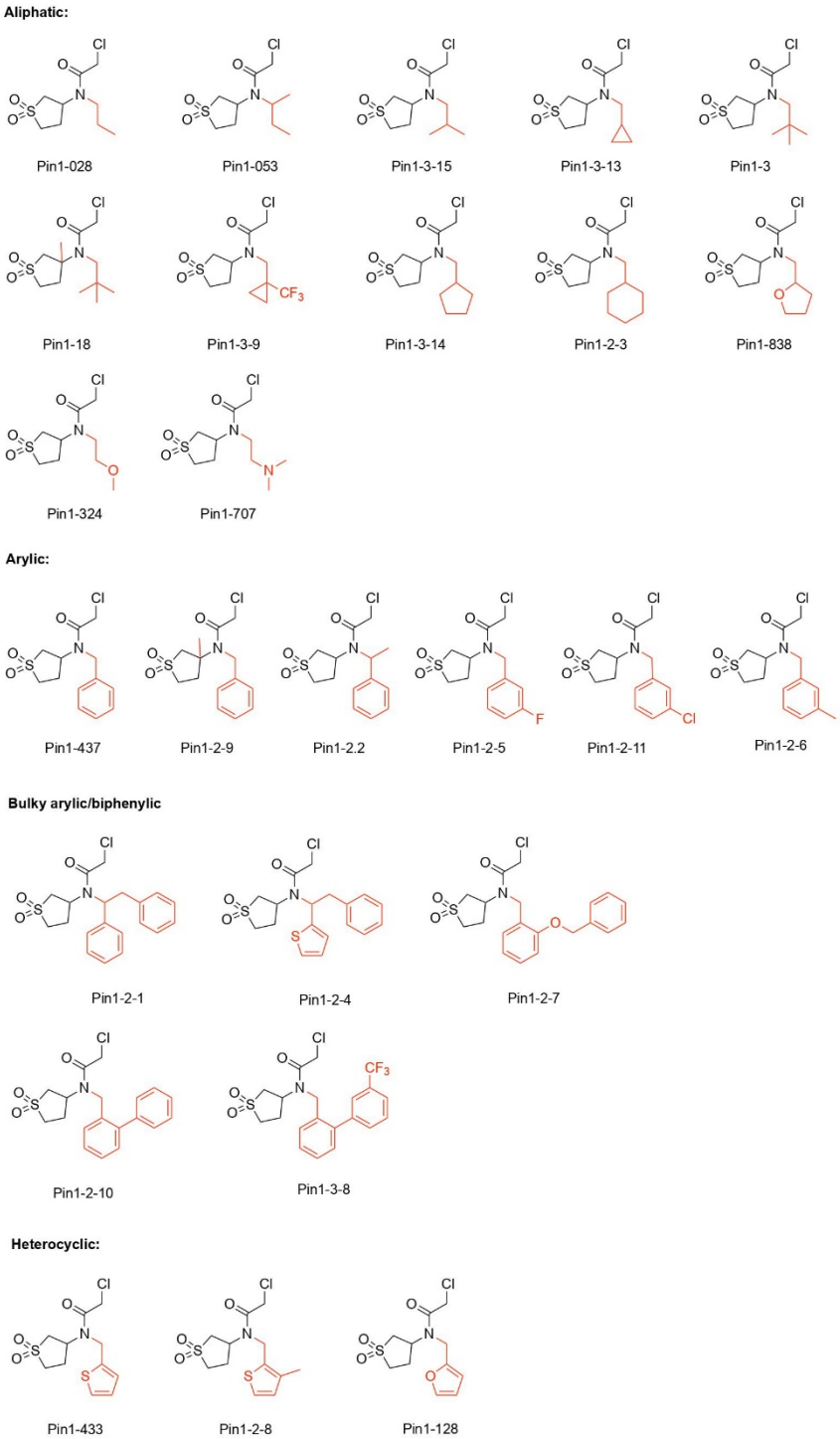


PCM-0102230

Supplementary Figure 1. Structures of 2-sulfolenes and sulfolane hits that labeled >75% in the primary screen.

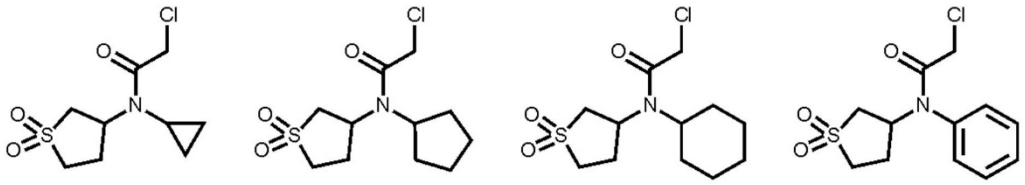


Supplementary Figure 2. Examples of docking poses with selected screening compounds bound to Pin1 (pdb code: 2ZQV, 2.5 Å).
A. The phenyl- and cyclohexyl residues of PCM-010275 (purple) and PCM-010276 (cyan), respectively, protrude into a hydrophobic cavity build up by Met130, Gln131 & Phe134. **B**. The cyclopropyl of PCM-010283 (orange) covers a shallow hydrophobic patch formed by Ser115, Leu122 and Met130. The ethyl- and the cyclopropyl moiety of PCM-010210 (brown) and PCM-0120231 (light brown), respectively, protrude into the solvent.



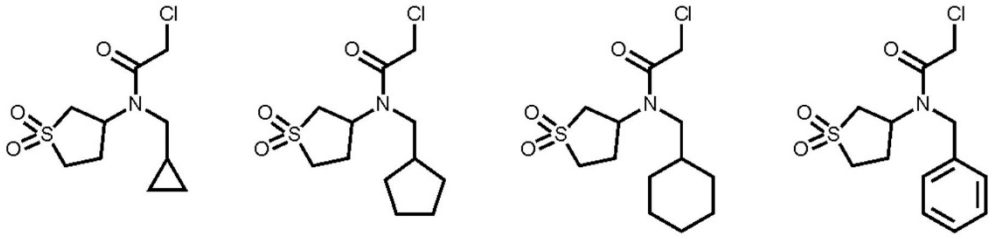
Supplementary Figure 3. Sulfolane analogs tested for Pin1 irreversible labeling. The varying lipophilic residues are denoted in red.

Sulfolane fragments with no Pin1 labeling at 2 µM (protein:compound) 1h.



PCM-0102832 PCM-0102313 PCM-0102760 PCM-0102755

Matched molecular pairs with improved labeling under the same conditions



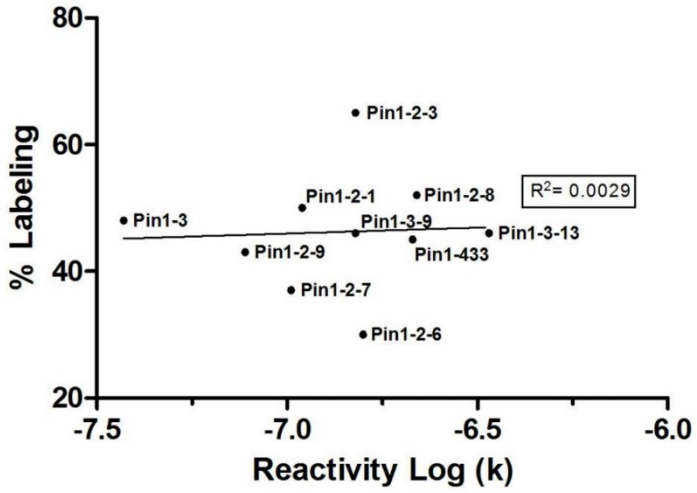
Pin1-3-13
46% labeling

Pin1-3-14
27% labeling

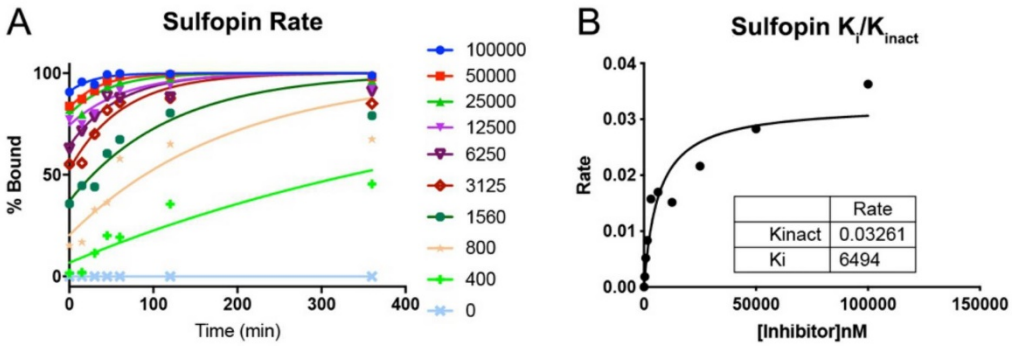
Pin1-2-3
65% labeling

Pin1-437
27% labeling

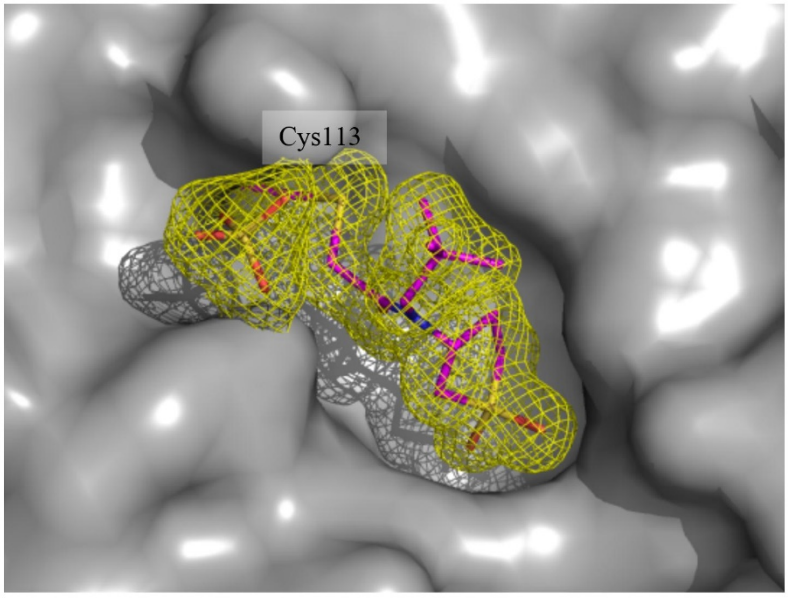
Supplementary Figure 4. Matched molecular pair analysis highlights the importance of a methylene linker between the amide and lipophilic moiety.



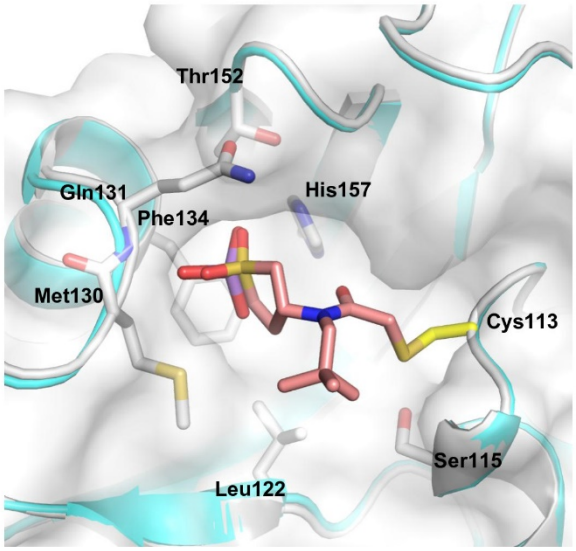
Supplementary Figure 5. Correlation between labeling percentage and reactivity
% Labeling and reactivity (log (k)) of the top ten binders shows no correlation ($R^2 = 0.003$).



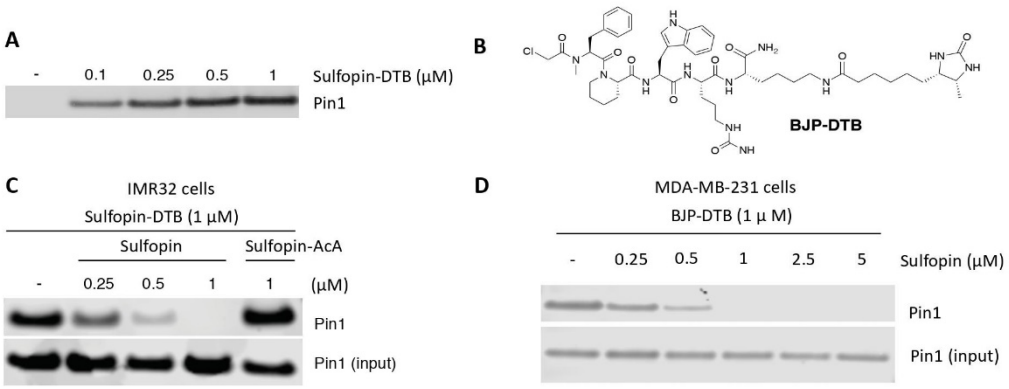
Supplementary Figure 6. Determination of Sulfopin K_{inact}/K_i. A. Determining labeling rates for various concentrations of Sulfopin. B. Plotting rates as a function of Sulfopin concentration to extrapolate K_{inact} and K_i.



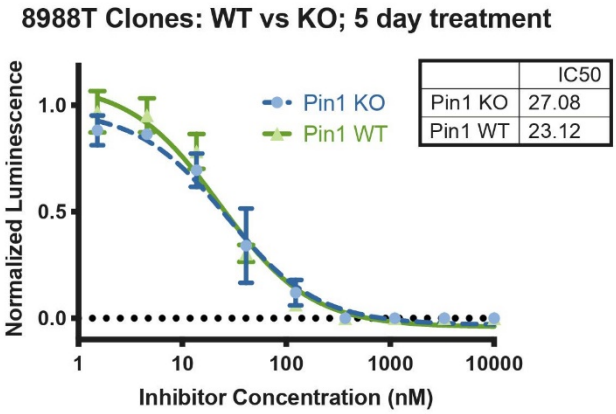
Supplementary Figure 7. Continuous electron density between Cys113 and Sulfopin confirms the formation of the covalent bond.



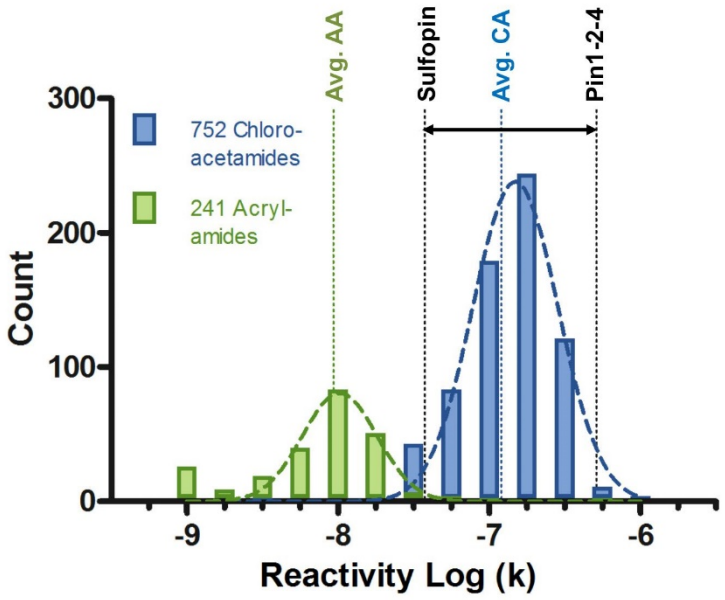
Supplementary Figure 8. Superposition of Pin1 (white) in complex with Sulfopin (pink; PDB: 6VAJ, 1.4Å) and Pin1 (cyan) in complex with arsenic trioxide (ATO, purple) (PDB: 6DUN, 1.6 Å).
The sulfolane moiety of Sulfopin and ATO occupy the hydrophobic Pro-binding pocket formed by M130, Q131, F134, Thr152 and H157. The sulfonyl oxygens (red) of Sulfopin and ATO analogously mediate hydrogen bonds with the backbone amide of Q131 and the imidazole NH of H157.



Supplementary Figure 9. Lysate and cellular engagement of Sulfopin-DTB and Sulfopin
A. Sulfopin-DTB engages Pin1 in PATU-8988T cell lysates.
B. BJP-DTB is a desthiobiotin probe based on our recently reported covalent peptide Pin1 inhibitior.
C. Sulfopin engages Pin1 in IMR32 cells.
D. Sulfopin engages Pin1 in MDA-MB-231 cells.



Supplementary Figure 10. Juglone's viability effects are not Pin1-mediated. PATU 8988T cells either WT for Pin1 or KO (see Fig. 4A) were treated with Juglone for 5 days. The similar IC_{50} indicates that Juglone's viability effects are Pin1 independent.



Supplementary Figure 11. Reactivity distribution of optimized Pin1 covalent inhibitors compared to the electrophile fragment library. The most reactive Pin1-2-4 is ~14x more reactive than the least reactive of the top analogs - Sulfopin, which shows a reactivity close to an acrylamide.

Supplementary Tables

Compound	Labeling [%]
PCM-0102372	100
PCM-0102760	100
PCM-0102539	100
PCM-0102579	100
PCM-0102868	100
PCM-0102230	87
PCM-0102105	85
PCM-0102755	83
PCM-0102313	83
PCM-0102178	72
PCM-0102832	72
PCM-0103082	69
PCM-0102138	56
PCM-0102896	42

Supplementary Table 1. Labeling of Pin1 by fragments of the electrophilic library that contain 2-sulfolene or sulfolane scaffolds.
2 μM of Pin1 were incubated with compounds at 200 μM concentration for 24 h at RT. Labeling in % was assigned via intact protein LC/MS. See Supp. Dataset 1 for structures of compounds.

Compound	Labeling [%][a]	Reactivity k [M^-1*s^-1][b]	Reactivity Log k[c]
Pin1-2-3	65	1.53E-07	-6.82
Pin1-2-8	52	2.19E-07	-6.66
Pin1-2-1	50	1.09E-07	-6.96
Pin1-3	48	3.73E-08	-7.43
Pin1-3-13	46	1.50E-07	-6.82
Pin1-3-9	46	3.42E-07	-6.47
Pin1-433	45	2.13E-07	-6.67
Pin1-2-9	43	7.68E-08	-7.11
Pin1-2-7	37	1.02E-07	-6.99
Pin1-3-7	36	1.12E-07	-6.95
Pin1-2-6	30	1.58E-07	-6.80
Pin1-053	28	1.24E-07	-6.91
Pin1-2-2	27	8.06E-08	-7.09
Pin1-3-14	27	7.03E-08	-7.15
Pin1-437	27	1.51E-07	-6.82
Pin1-128	25	1.47E-07	-6.83
Pin1-2-10	25	1.30E-07	-6.89
Pin1-2-5	24	1.31E-07	-6.88
Pin1-3-8	23	8.22E-08	-7.09
Pin1-3-15	21	7.77E-08	-7.11
Pin1-2-11	19	1.15E-07	-6.94
Pin1-838	16	1.41E-07	-6.85
Pin1-028	16	1.59E-07	-6.80
Pin1-324	12	1.59E-07	-6.80
Pin1-707	0	1.17E-09	-8.93
PCM-0102138	0	1.20E-07	-6.92
PCM-0102178	0	1.30E-07	-6.89
PCM-0102105	0	1.10E-07	-6.96
PCM-0102832	0	6.02E-08	-7.22
PCM-0102313	0	1.07E-07	-6.97
PCM-0102760	0	1.00E-07	-7.00
PCM-0102755	0	1.54E-07	-6.81
PCM-0102230	0	8.87E-08	-7.05

Supplementary Table 2. Characterization of sulfolane analogs and primary screening hits.
[a] 2 μM Pin1 were incubated with compounds at 2 μM concentration for 1 h at RT. Labeling in % was assigned via intact protein LC/MS. See Supp. Fig. 1 & 3 for structures of compounds.
[b] The second-order rate constant for thiol reactivity as determined by a DTNB assay.
[c] Log_{10} value of the reactivity rate constant.

Compound	Labeling [%]	Ki [nM][a]	Reactivity k [M^-1*s^-1]	Log k	EC_{50} IMR90 [µM][b]
Pin1-2-3	65	17/46	1.53E-07	-6.82	7.5
Pin1-2-8	52	133	2.19E-07	-6.66	5.1
Pin1-2-1	50	7\58	1.09E-07	-6.96	2.8
Pin1-3	48	17/20/56/110	3.73E-08	-7.43	>25
Pin1-3-13	46	121	1.50E-07	-6.82	N.D.
Pin1-3-9	46	411	3.42E-07	-6.47	N.D.
Pin1-433	45	40/194	2.13E-07	-6.67	8.9
Pin1-2-9	43	83	7.68E-08	-7.11	11.3
Pin1-2-7	37	39	1.02E-07	-6.99	6.1
Pin1-2-6	30	194	1.58E-07	-6.80	5.6

Supplementary Table 3. Activity of the top ten labeling compounds.
[a] Ki values were determined in FP assays after 14 h incubation with compound. Multiple values represent results from multiple independent repetitions of the assay.
[b] EC_{50} values are from a CellTiterGlo viability experiment (48h) with IMR90 cells.
See Supp. Fig. 3 for structures of compounds.

Supplementary Table 4. Crystallization conditions and data collection and refinement statistics for crystal structures

RCSB accession code	6VAJ
Data collection [a]	
Space group	P 4_3 2_1 2
Cell dimensions	
a, *b*, *c* (Å)	48.96 48.96 137.04
a, b, g (°)	90.00 90.00 90.00
Resolution (Å)	39.84 - 1.42 (1.471 - 1.42) [b]
R_{pim}	0.01849 (0.5658)
Redundancy	6.2 (6.3)
Completeness (%)	99.38 (99.72)
$I / \sigma I$	17.67 (1.54)
Structure solution	
PDB entries used for molecular replacement	1PIN
Refinement	
No. reflections	32262 (3163)
R_{work}	0.1923 (0.3278)
R_{free}	0.2144 (0.3227)
No. atoms	1384
Macromolecules	1229
Ligand/ion	23
Water	132
B-factors	31.41
Macromolecules	30.11
Ligand/ion	50.67
Water	40.23
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.19
Ramachandran	
Preferred	100.0%
Allowed	0.0%
Not Allowed	0.0%

[a] A single crystal was used to collect data for each of the structures reported here.

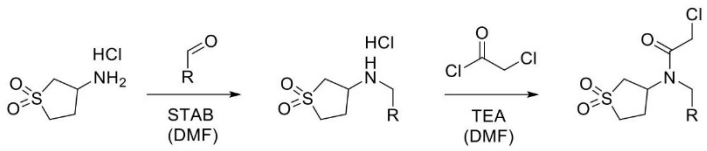
[b] Values in parentheses are for highest-resolution shell.

Supplementary methods

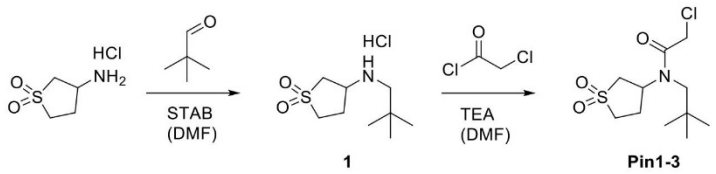
Synthetic chemistry methods

All solvents and reagents used for organic synthesis were purchased from Sigma-Aldrich, Merck, Baker, Acros and used without further purification. Building blocks for synthesis were purchased from Enamine and MolPort chemical suppliers. Purification of precursors was performed using an automated Flash chromatography system (CombiFlash® Systems, Teledyne Isco, USA) with RediSep Rf Normal-phase Flash Columns. Final compounds were purified by semi-preparative HPLC on a Waters Prep 2545 Preparative Chromatography System, with UV/Vis detector 2489, using XBridge® Prep C18 10µm 10x250 mm Column (PN: 186003891, SN:161I3608512502). Reaction progress was monitored using a Waters UPLC-MS system: Acquity UPLC® H class with PDA detector, Acquity UPLC® BEH C18 1.7 µm 2.1x50 mm Column (PN:186002350, SN 02703533825836), Waters SQ detector 2. Spectral analysis by ^{1}H- and ^{13}C-NMR was obtained on a Bruker Avance -300 MHz and 400 MHz spectrometer, equipped with a QNP probe. Chemical shifts (δ_H & δ_C) are quoted in ppm to the nearest 0.1 ppm, and referenced to trimethylsilane (TMS). Coupling constants (J) are reported in Hertz (Hz) to the nearest 0.1 Hz. High resolution electron-spray mass spectrometry (HRMS-ESI) was performed on a Xevo G2-XS QTOF Mass Spectrometer (Waters Corporation, USA) and reported values are within the error limit of ± 2.2 ppm mass units.

General procedure for chloroacetamide preparation



3-aminosulfolane hydrochloride (1 eq.) was added to a solution of triethylamine (TEA) (0.9 eq.) in dry dimethylformamide (DMF) and stirred for 1 h at room temperature (RT). Afterwards, aldehyde (1.1 eq.) and acetic acid (0.2 eq.) were added to the reaction mixture and stirred at RT for 1 h. Then sodium triacetoxyborohydride (STAB) (2.1 eq.) was added at once to the mixture and stirred overnight at RT. After evaporation of the solvent, the residue was dissolved with sat. aq. $NaHCO_3$ and the aq. solution was extracted with ethyl acetate (EA) (2x). The organic layers were combined, dried over Na_2SO_4 and filtered. Evaporation of the solvent yielded the secondary amine as hydrochloride, which was used without purification in the next step. Secondary amine hydrochloride (1 eq.) was dissolved in dry DMF and cooled to 0 °C. Subsequently, 2-chloroacetyl chloride (1.2 eq.) and TEA (1.2 eq.) were added dropwise at 0 °C and stirred for 30 min. Afterwards the reaction mixture was allowed to reach RT and stirred for 1 h. The reaction was quenched at 0 °C by the addition of water and filtered. Purification by RP-HPLC (linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded the corresponding chloroacetamide.



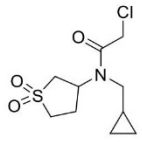
1 Pin1-3

2-chloro-*N*-(sulfolan-3-yl)-*N*-neopentylacetamide (Sulfopin / Pin1-3)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (100 mg, 0.583 mmol, 1 eq.) was added to a solution of triethylamine (TEA) (73 µL, 0.524 mmol, 0.9 eq.) in dry dimethylformamide (DMF) (1.4 mL) and stirred for 1 h at room temperature (RT). Afterwards, pivaldehyde (70 µL, 0.641 mmol, 1.1 eq.) and one drop of acetic acid (0.2 eq.) were added to the reaction mixture and stirred at RT for 1 h. Then sodium triacetoxyborohydride (STAB) (259 mg, 1.223 mmol, 2.1 eq.) was added at once to the mixture and stirred overnight at RT. After evaporation of the solvent, the residue was dissolved with sat. aq. $NaHCO_3$ (0.5 mL) and the aq. solution was extracted with ethyl acetate (EA) (2x 1 mL). The organic layers were combined, dried over Na_2SO_4 and filtered. Evaporation of the solvent yielded the secondary amine **1** as white solid (86.2 mg, 0.42 mmol, 72% (crude product)), which was used without purification in the next step.

Secondary amine **1** hydrochloride (100 mg, 0.487 mmol, 1 eq.) was dissolved in dry DMF (1 mL) and cooled to 0 °C. Subsequently, 2-chloroacetyl chloride (47 µL, 0.584 mmol, 1.2 eq.) and TEA (81 µL, 0.584 mmol, 1.2 eq.) were added dropwise at 0 °C and stirred for 30 min. Afterwards the reaction mixture was allowed to reach RT and stirred for 2 h. The reaction was quenched at 0 °C by the addition of water (2 mL). Purification by RP-HPLC (t_R = 16 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-3 (59.8 mg, 0.212 mmol, 44%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.11 (d, *J*=5.5 Hz, 2 H), 3.91 - 4.00 (m, 1 H), 3.65 - 3.79 (m, 2 H), 3.25 - 3.33 (m, 1 H), 3.11 - 3.20 (m, 2 H), 3.00 - 3.09 (m, 1 H), 2.46 - 2.60 (m, 2 H), 1.03 (s, 9 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.0, 62.4, 57.6, 50.3, 49.0, 42.1, 33.6, 28.0, 26.6 ppm.
HRMS (ESI): *m/z* calcd. for $C_{11}H_{20}ClNO_3SNa^+$ $[M+Na]^+$: 304.0750; found 304.0746.



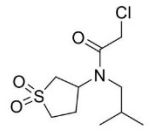
Pin1-3-13

2-chloro-*N*-(cyclopropylmethyl)-*N*-(sulfolan-3-yl)acetamide (Pin1-3-13)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (100 mg, 0.583 mmol, 1 eq.), TEA (73 µL, 0.524 mmol, 0.9 eq.) in dry DMF (1.4 mL) were stirred for 1 h at RT. Cyclopropanecarboxaldehyde (48 µL, 0.641 mmol, 1.1 eq.), one drop of acetic acid and STAB (259 mg, 1.223 mmol, 2.1 eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine as hydrochloride (63.1 mg, 0.280 mmol, 48% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (27 µL, 0.335 mmol, 1.2 eq.) and TEA (47 µL, 0.335 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (63.1 mg, 0.280 mmol, 1 eq.) in dry DMF (0.6 mL) and stirred for 30 min. Quenching with water (2 mL), filtering, purification by RP-HPLC (t_R = 13.5

min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-3-13 (24.9 mg, 0.094 mmol, 34%) as white powder.

^{1}H NMR (500MHz, $CDCl_3$): δ = 4.27 - 4.41 (m, 1 H), 4.13 (s, 2 H), 3.52 - 3.73 (m, 2 H), 3.17 - 3.38 (m, 3 H), 3.02 - 3.14 (m, 1 H), 2.45 - 2.61 (m, 2 H), 0.91 - 1.04 (m, 1 H), 0.73 (d, *J*=6.6 Hz, 2 H), 0.36 (d, *J*=4.4 Hz, 2 H) ppm.
^{13}C NMR (126MHz, $CDCl_3$): δ = 166.7, 54.4, 54.1, 50.7, 50.2, 41.9, 26.4, 11.1, 4.3 ppm.
HRMS (ESI): *m/z* calcd. for $C_{10}H_{16}ClNO_3SNa^+$ $[M+Na]^+$: 288.0437; found 288.0435.

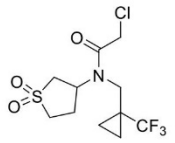


Pin1-3-15

2-chloro-*N*-(sulfolan-3-yl)-*N*-isobutylacetamide (Pin1-3-15)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (90 mg, 0.524 mmol, 1 eq.), TEA (66 µL, 0.474 mmol, 0.9 eq.) in dry DMF (1.3 mL) were stirred for 1 h at RT. Isobutyraldehyde (57 µL, 0.629 mmol, 1.2 eq.), one drop of acetic acid (0.2 eq.) and STAB (233 mg, 1.101 mmol, 2.1 eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine as hydrochloride (78.18 mg, 0.343 mmol, 66% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (33 µL, 0.412 mmol, 1.2 eq.) and TEA (57 µL, 0.412 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (78.18 mg, 0.487 mmol, 1 eq.) in dry DMF (0.7 mL) and stirred for 30 min. Quenching with water (2 mL), filtering, purification by RP-HPLC (t_R = 14 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-3-15 (29.22 mg, 0.412 mmol, 32%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.10 (s, 2 H), 3.99 - 4.08 (m, 1 H), 3.56 - 3.77 (m, 2 H), 3.11 - 3.22 (m, 3 H), 3.01 - 3.10 (m, 1 H), 2.44 - 2.60 (m, 2 H), 1.86 - 1.98 (m, 1 H), 0.99 (t, *J*=6.6 Hz, 6 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 167.0, 58.0, 55.2, 50.5, 49.7, 42.0, 28.4, 26.2, 19.9, 19.7 ppm
HRMS (ESI): *m/z* calcd. for $C_{10}H_{18}ClNO_3SNa^+$ $[M+Na]^+$: 209.0594; found 290.0590.



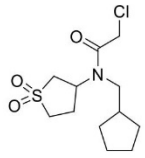
Pin1-3-9

2-chloro-*N*-(sulfolan-3-yl)-*N*-((1-(trifluoromethyl)cyclopropyl)methyl)acetamide (Pin1-3-9)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (55 mg, 0.320 mmol, 1 eq.), TEA (40 µL, 0.288 mmol, 0.9 eq.) in dry DMF (0.8 mL) were stirred for 1 h at RT. 1-(Trifluoromethyl)cyclopropanecarbaldehyde (48.7 mg, 0.352 mmol, 1.1 eq.), one drop of acetic acid and STAB (143 mg, 0.673 mmol, 2.1 eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine as hydrochloride (72 mg, 0.245 mmol, 76% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (22 µL, 0.270 mmol, 1.1 eq.) and TEA (41 µL, 0.294 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (72 mg, 0.245 mmol, 1 eq.) in

dry DMF (0.6 mL) and stirred for 30 min. Quenching with water (2 mL), filtration, purification by RP-HPLC (t_R = 16.5 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-3-9 (18.8 mg, 0.056 mmol, 23%) as white powder.

^{1}H NMR (500MHz, $CDCl_3$): δ = 4.11 - 4.18 (m, 2 H), 3.98 - 4.10 (m, 1 H), 3.57 - 3.73 (m, 4 H), 3.10 - 3.22 (m, 1 H), 3.02 - 3.09 (m, 1 H), 2.47 - 2.56 (m, 2 H), 1.18 - 1.28 (m, 2 H), 0.83 - 0.91 (m, 2 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 167.2, 127.4, 125.2, 55.3, 51.9, 50.4, 49.3, 41.5, 26.0, 25.4, 8.3 ppm.
HRMS (ESI): *m/z* calcd. for $C_{11}H_{15}ClNO_3SNa^+$ $[M+Na]^+$: 356.0311; found 356.0306.

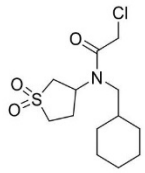


Pin1-3-14

2-chloro-*N*-(cyclopentylmethyl)-*N*-(sulfolan-3-yl)acetamide (Pin1-3-14)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (100 mg, 0.583 mmol, 1 eq.), TEA (73 µL, 0.524 mmol, 0.9 eq.) in dry DMF (1.4 mL) were stirred for 1 h at RT. Cyclopentanecarboxaldehyde (68 µL, 0.641 mmol, 1.1 eq.), one drop of acetic acid and STAB (259 mg, 1.223 mmol, 2.1 eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine (95.68 mg, 0.377 mmol, 65% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (36 µL, 0.452 mmol, 1.2 eq.) and TEA (63 µL, 0.452 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (95.68 mg, 0.377 mmol, 1 eq.) in dry DMF (1 mL) and stirred for 30 min. Quenching with water (2 mL), filtration, purification by RP-HPLC (t_R = 17.5 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded Pin1-3-14 (23.4 mg, 0.08 mmol, 21%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.11 (s, 3 H), 3.57 - 3.76 (m, 2 H), 3.26 - 3.39 (m, 2 H), 3.02 - 3.20 (m, 2 H), 2.47 - 2.58 (m, 2 H), 2.11 - 2.21 (m, 1 H), 1.77 - 1.88 (m, 2 H), 1.60 - 1.76 (m, 4 H), 1.17 - 1.29 (m, 2 H) ppm.
^{13}C NMR (125 MHz, $CDCl_3$): δ = 166.8, 55.3, 55.1, 50.5, 49.7, 42.0, 40.2, 30.4, 30.4, 26.3, 24.9, 24.9 ppm.
HRMS (ESI): *m/z* calcd. for $C_{12}H_{20}ClNO_3SNa^+$ $[M+Na]^+$: 316.0750; found 316.0743.



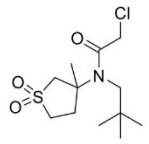
Pin1-2-3

2-chloro-*N*-(cyclohexylmethyl)-*N*-(sulfolan-3-yl)acetamide (Pin1-2-3)

Following the general procedure of chloroacetamide preparation, 3-aminosulfolane hydrochloride (75 mg, 0.437 mmol, 1 eq.) in dry DMF (1.1 mL) were stirred for 1 h at RT. Cyclohexanecarboxaldehyde (58 µL, 0.481 mmol, 1.1 eq.) and STAB (139 mg, 0.655 mmol, 1.5

eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine as hydrochloride (72.11 mg, 0.269 mmol, 62% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (25 µL, 0.323 mmol, 1.2 eq.) and TEA (45 µL, 0.323 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (72 mg, 0.269 mmol, 1 eq.) in dry DMF (0.6 mL) and stirred for 30 min. Quenching with water (2 mL), filtration, purification by RP-HPLC (t_R = 18.5 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded Pin1-2-3 (9.1 mg, 0.030 mmol, 11%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.09 (s, 2 H), 4.00 - 4.07 (m, 1 H), 3.56 - 3.77 (m, 2 H), 3.01 - 3.27 (m, 4 H), 2.45 - 2.59 (m, 2 H), 1.66 - 1.88 (m, 5 H), 1.51 - 1.63 (m, 1 H), 1.13 - 1.33 (m, 3 H), 0.89 - 1.03 (m, 2 H) ppm.
^{13}C NMR (125 MHz, $CDCl_3$): δ = 167.0, 57.1, 55.3, 50.5, 49.7, 42.0, 38.0, 30.9, 30.8, 26.2, 26.1, 25.8 ppm.
HRMS (ESI): *m/z* calcd. for $C_{13}H_{22}ClNO_3SNa^+$ $[M+Na]^+$: 330.0907; found 330.0901.

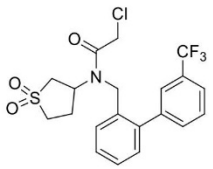


Pin1-18

2-chloro-*N*-(3-methyl-sulfolan-3-yl)-*N*-neopentylacetamide (Pin1-18)

Following the general procedure of chloroacetamide preparation, 3-amino-3-methylsulfolane hydrochloride (200 mg, 1.077 mmol, 1 eq.), TEA (135 µL, 0.969 mmol, 0.9 eq.) in dry DMF (2.7 mL) were stirred for 1 h at RT. Pivaldehyde (129 µL, 1.185 mmol, 1.1 eq.), two drops of acetic acid and STAB (479 mg, 2.262 mmol, 2.1 eq.) were added and stirred overnight at RT. After workup and evaporation, the secondary amine (267.6 mg, 1.046 mmol, 97% (crude product)) was used without purification in the next step. 2-chloroacetyl chloride (17.98 µL, 0.235 mmol, 1.2 eq.) and TEA (32.7 µL, 0.235 mmol, 1.2 eq.) were added dropwise to cooled (0 °C) secondary amine hydrochloride (50 mg, 0.195 mmol, 1 eq.) in dry DMF (0.4 mL) and stirred for 30 min. Quenching with water (2 mL) and purification by RP-HPLC (t_R = 17 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) yielded Pin1-18 (11.94 mg, 0.04 mmol, 21%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 3.96 - 4.44 (m, 2 H), 3.68 (d, *J*=13.2 Hz, 1 H), 3.51 - 3.65 (m, 1 H), 3.41 (s, 2 H), 3.12 - 3.37 (m, 2 H), 2.13 - 2.54 (m, 2 H), 1.77 (s, 3 H), 1.04 (s, 9 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.0, 60.9, 55.0, 51.0, 43.5, 33.7, 29.1, 27.1 ppm.
HRMS (ESI): *m/z* calcd. for $C_{12}H_{22}ClNO_3SNa^+$ $[M+H^+]$: 318.0907; found 318.0914.



Pin1-3-8

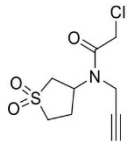
2-chloro-*N*-(sulfolan-3-yl)-*N*-((3'-(trifluoromethyl)-[1,1'-biphenyl]-2-yl)methyl)acetamide (Pin1-3-8)

3-aminosulfolane hydrochloride (100 mg, 0.583 mmol, 1 eq.) was added to a solution of TEA (73.1 µL, 0.524 mmol, 0.9 eq.) in MeOH (2 mL) and stirred for 1 h at room temperature (RT). Afterwards, 3'-(trifluoromethyl)-1,1'-biphenyl-2-carboxaldehyde (182 µL, 0.728 mmol, 1.25 eq.) and acetic acid (6.67 µL, 0.117 mmol, 0.2 eq.) were added to the reaction mixture and stirred at RT for 1 h. Then sodium cyanoborohydride (92 mg, 1.456 mmol, 2.5 eq.) was added at once to the mixture and stirred overnight at RT. After evaporation of the solvent, the residue was dissolved with sat. aq. $NaHCO_3$ (0.5 mL) and the aq. solution was extracted with ethyl acetate (EA) (2x 1 mL). The organic layers were combined, dried over Na_2SO_4 and filtered. Evaporation of the solvent yielded the secondary amine **1** as white solid (231.8 mg, 0.571 mmol, 98% (crude product)), which was used without purification in the next step. The secondary amine hydrochloride (100 mg, 0.271 mmol, 1 eq.) was dissolved in dry DMF (0.5 mL) and cooled to 0 °C. Subsequently, 2-chloroacetyl chloride (26 µL, 0.325 mmol, 1.2 eq.) and TEA (56.6 µL, 0.406 mmol, 1.5 eq.) were added dropwise at 0 °C and stirred for 30 min. Afterwards the reaction mixture was allowed to reach RT and stirred for 2 h. The reaction was quenched at 0 °C by the addition of water (2 mL). Purification by RP-HPLC (t_R = 22 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-3-8 (40.69 mg, 0.091 mmol, 34%) as white powder.

^{1}H NMR (500 MHz, CDCl3): δ = 7.73 (d, J=7.7 Hz, 1 H), 7.65 (t, J=7.7 Hz, 1 H), 7.56 - 7.61 (m, 1 H), 7.43 - 7.56 (m, 3 H), 7.21 - 7.39 (m, 2 H), 5.06 (br. s., 1 H), 4.69 (br. s., 1 H), 4.50 - 4.61 (m, 2 H), 4.42 (br. s., 1 H), 4.20 (s, 1 H), 3.23 - 3.45 (m, 2 H), 3.06 - 3.18 (m, 1 H), 2.95 - 3.05 (m, 1 H), 2.25 - 2.43 (m, 2 H) ppm.

^{13}C NMR (126 MHz, $CDCl_3$): δ = 167.2, 140.3, 132.2, 131.0, 129.5, 129.2, 129.0, 128.6, 125.7, 124.9, 62.7, 53.0, 50.6, 40.4, 26.3 ppm.

HRMS (ESI): *m/z* calcd. for $C_{20}H_{19}ClF_3NO_3SNa^+$ $[M+Na]^+$: 468.0624; found 468.0616.

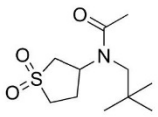


Pin1-4

2-chloro-*N*-(sulfolan-3-yl)-*N*-(prop-2-yn-1-yl)acetamide (Pin1-4)

3-(prop-2-yn-1-ylamino)sulfolane hydrochloride (100 mg, 0.477 mmol, 1 eq.) was dissolved in dry DMF (1 mL) and cooled to 0 °C. Subsequently, 2-chloroacetyl chloride (44 µL, 0.572 mmol, 1.2 eq.) and TEA (120 µL, 0.858 mmol, 1.8 eq.) were added dropwise at 0 °C and stirred for 30 min. Afterwards the reaction mixture was allowed to reach RT and stirred for 30 min. The reaction was quenched at 0 °C by the addition of water (2 mL). Subsequent filtration, purification by RP-HPLC (t_R = 11 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Pin1-4 (52 mg, 0.209 mmol, 44%) as white powder.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.90 - 5.03 (m, 1 H), 4.16 - 4.25 (m, 4 H), 3.31 - 3.51 (m, 3 H), 3.02 - 3.20 (m, 1 H), 2.39 - 2.89 (m, 3 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 167.0, 74.6, 52.1, 50.9, 41.5, 35.7, 26.3 ppm.
HRMS (ESI): *m/z* calcd. for $C_9H_{12}ClNO_3SNa^+$ $[M+Na]^+$: 272.0124; found 272.0119.



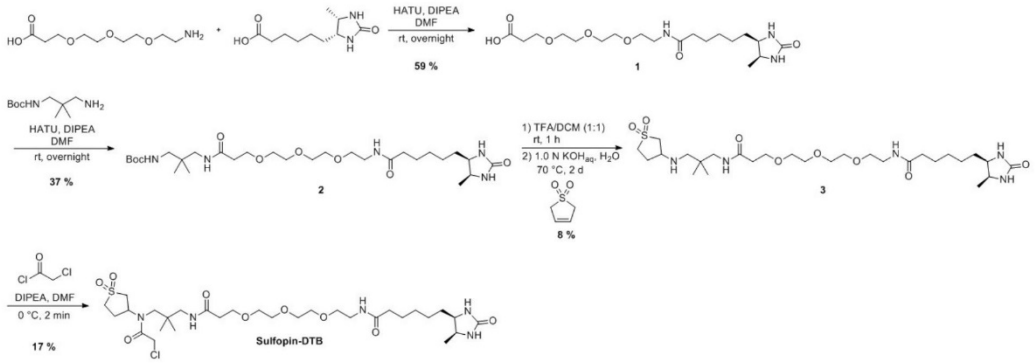
Sulfopin-AcA

***N*-(sulfolan-3-yl)-*N*-neopentylacetamide (Sulfopin-AcA)**

3-(neopentylamino)sulfolane hydrochloride (75 mg, 0.310 mmol, 1 eq.) was dissolved in dry DMF (1 mL) and acetic anhydride (64 µL, 0.682 mmol, 2.2 eq.) and TEA (30 µL, 0.217 mmol, 0.7 eq.) were added dropwise. The reaction was stirred at RT overnight. The reaction was quenched by the addition of water (2 mL). Filtration, purification by RP-HPLC (t_R = 8 min, linear gradient 5 → 95% ACN/H_2O + 0.1% TFA in 30 min) and lyophilization yielded chloroacetamide Sulfopin-AcA (8.75 mg, 0.035 mmol, 34%) as white powder.

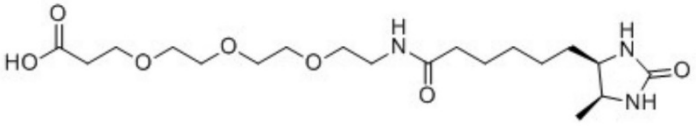
^{1}H NMR (500 MHz, $CDCl_3$): δ = 3.87 - 3.97 (m, 1 H), 3.75 - 3.84 (m, 1 H), 3.63 - 3.74 (m, 1 H), 3.28 (d, *J*=15.4 Hz, 1 H), 3.07 - 3.15 (m, 2 H), 2.99 - 3.06 (m, 1 H), 2.47 - 2.59 (m, 2 H), 2.12 - 2.18 (m, 3 H), 0.98 - 1.08 (m, 9 H) ppm.
^{13}C NMR (126 MHz, $CDCl_3$): δ = 172.9, 63.1, 57.0, 50.4, 49.4, 33.8, 28.1, 26.9, 23.2 ppm.
HRMS (ESI): *m/z* calcd. for $C_{11}H_{21}NO_3SNa^+$ $[M+Na]^+$: 270.1140; found 270.1140.

General scheme of the DTB-Probe synthesis



Experimental Details for Individual Compound Synthesis

19-((4R,5S)-5-methyl-2-oxoimidazolidin-4-yl)-14-oxo-4,7,10-trioxa-13-azanonadecanoic acid



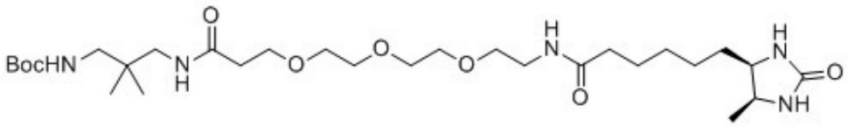
1

6-((4R,5S)-5-methyl-2-oxoimidazolidin-4-yl)hexanoic acid (96.8 mg, 0.452 mmol), HATU (171.9 mg, 0.452 mmol) and DIPEA (240 µL, 1.36 mmol) were dissolved in DMF (1.0 mL) and stirred

at r.t. for 15 min. The resulting reaction mixture was added dropwise to a solution of 3-(2-(2-(2-aminoethoxy)ethoxy)ethoxy)propanoic acid (100.0 mg, 0.452 mmol) in DMF (1.0 mL) until all the starting material was consumed. The crude was purified by HPLC to yield the product as a yellow oil (142 mg, 0.340 mmol, 59%).

MS (ESI): *m/z* calcd. for $C_{19}H_{36}N_3O_7^+$ $[M+H]^+$: 418.25; found 418.4.

tert-butyl(2,2-dimethyl-23-((4R,5S)-5-methyl-2-oxoimidazolidin-4-yl)-5,18-dioxo-8,11,14-trioxa-4,17-diazatricosyl)carbamate

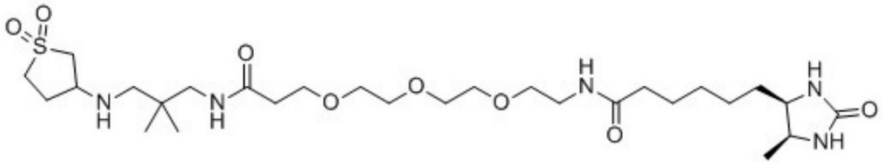


2

Intermediate 1 (130.0 mg, 0.311 mmol), tert-butyl (3-amino-2,2-dimethylpropyl)carbamate (75.5 mg, 0.373 mmol), HOBt (75.7 mg, 0.560 mmol) and DIPEA (163 µL, 0.933 mmol) were dissolved in DMF (1.0 mL). EDC·HCl (107.4 mg, 0.560 mmol) was added to the reaction mixture, which was stirred at r.t. for 16 h. The crude was purified by HPLC to yield the product as a dark yellow oil (82 mg, 0.136 mmol, 37%).

MS (ESI) *m/z* calcd. for $C_{29}H_{56}N_5O_8^+$ $[M+H]^+$: 602.41; found 602.6.

N-(16-((1,1-dioxidotetrahydrothiophen-3-yl)amino)-15,15-dimethyl-12-oxo-3,6,9-trioxa-13-azahexadecyl)-6-((4R,5S)-5-methyl-2-oxoimidazolidin-4-yl)hexanamide

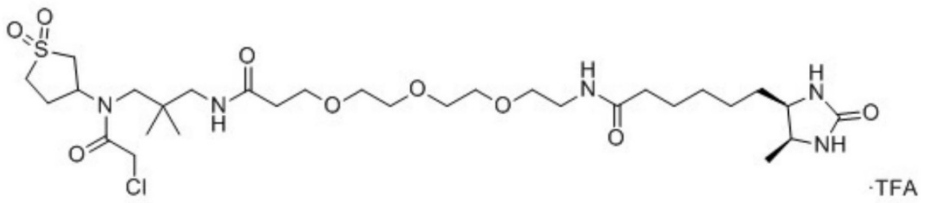


3

Intermediate 2 (52.5 mg, 0.0872 mmol) and TFA (1.0 mL) were stirred at r.t. in DCM (1.0 mL) for 1 h. The solvent was removed *in vacuo* and the residue was dissolved in H_2O (0.57 mL). 2,5-dihydrothiophene 1,1-dioxide (17.5 mg, 0.148 mmol) and 1.0 n aq. KOH (500 µL) were added to the solution and stirred at 70 °C for 2 d. The reaction mixture was concentrated *in vacuo* and the residue was purified by HPLC to yield the product as a yellow solid (6 mg, 0.010 mmol, 8%).

MS (ESI) *m/z* calcd. for $C_{28}H_{54}N_5O_8S^+$ $[M+H]^+$: 620.37; found 620.5.

N-(19-chloro-17-(1,1-dioxidotetrahydrothiophen-3-yl)-15,15-dimethyl-12,18-dioxo-3,6,9-trioxa-13,17-diazanonadecyl)-6-((4R,5S)-5-methyl-2-oxoimidazolidin-4-yl)hexanamide



Sulfopin-DTB

Intermediate 3 (6 mg, 0.010 mmol) and DIPEA (16 µL, 0.092 mmol) were dissolved in DMF (1.0 mL). The solution was cooled to 0 °C. 2-chloroacetyl chloride (18 µL) in DMF (300 µL) was added dropwise to the reaction mixture until the starting material was consumed. The crude was purified by HPLC to yield the title compound Pin1-3-DTB as a light-yellow solid (1.3 mg, 0.002 mmol, 17%).

^{1}H NMR (500 MHz, DMSO-d_6) δ = 7.81 (s, 1H), 6.28 (s, 1H), 6.11 (s, 1H), 4.33 (s, 1H), 4.02 – 3.92 (m, 1H), 3.61 (t, *J* = 6.3 Hz, 5H), 3.49 (s, 11H), 3.33 (s, 3H), 3.23 (s, 2H), 3.18 (d, *J* = 5.8 Hz, 3H), 3.10 (dt, *J* = 13.2, 6.8 Hz, 3H), 2.99 (s, 2H), 2.38 (s, 3H), 2.09 – 2.02 (m, 3H), 1.46 (d, *J* = 7.3 Hz, 2H), 1.28 (d, *J* = 47.9 Hz, 3H), 0.95 (d, *J* = 6.4 Hz, 3H), 0.88 (s, 6H) ppm.

MS (ESI) *m/z* calcd. for $C_{30}H_{55}ClN_5O_9S^+$ $[M+H]^+$: 696.34; found 696.6.

10 Diazirine-tagged Muraymycin Analogs as Photoaffinity Labeling Probes to Study a Novel Class of Antibiotics

10.1 Introduction

10.1.1 Activity-Based Protein Profiling and Photoaffinity Labeling as Tools in Drug Development

Activity-Based Protein Profiling (ABPP): Methods and Applications

Understanding the function of a protein in the complex environment of a cell or tissue, including its activation state, active site, localization within the cell, up- and downstream signaling pathways, PPIs, expression levels, and native substrates, is crucial for drug discovery campaigns. These features are important parts of understanding target biology, providing key links between a protein of interest and its potential to function as a therapeutic target (see Chapter 1). Likewise, it is important to thoroughly characterize novel drug candidates, especially in terms of their selectivity, in order to understand their modes of action, including on- and off-target effects. One method identifying protein expression levels and therefore protein functions in a global context is mass spectrometry-based proteomics analysis[1,2]. Over time, advanced chemical proteomic approaches have been developed and applied to small molecule discovery and characterization methods, such as drug screens, target identification and protein localization studies in the context of drug binding, as well as evaluation methods of downstream inhibitor effects on the proteome and phosphoproteome[3,4].

One important technique in chemical proteomics is ABPP, which arose from advances made in different areas including genomics[5] and mass spectrometry analysis[6]. Utilizing chemical probes that recognize and bind to the active site of their target protein in a complex system such as cells or tissues, this technique enables the functional investigation of target proteins[7], in contrast to mere quantitative methods that commonly determine mRNA or protein levels[8]. The general strategy behind this method is to covalently bind a protein by an ABP or A*f*BP *via* direct crosslinking or after activation of a photoreactive moiety, respectively[3]. The ligand-target complexes are then enriched or detected *via* affinity tags such as biotin[3]. ABPs/A*f*BPs are therefore commonly equipped with a reactive warhead at the target-binding structure of the molecule, that can thus form a covalent bond with the targeted protein, a linker as a spacer, and either a bioorthogonal group that can serve as an attachment point for the introduction of a tag or a tag directly attached to the linker[5,9] (Figure 10.1.1).

Covalent small molecule binders of proteins generally carry electrophilic warhead moieties such as Michael-acceptors, epoxides, or β-lactones/-lactams[5]. These groups irreversibly label specific amino acids, commonly at the active site, of a protein upon binding of the probe (Figure 10.1.2) and can therefore be found in ABP scaffolds[3]. However, most reported protein binders reversibly engage their target enzyme. For these ligands, PAL groups can be introduced into their structure, creating A*f*BPs that form highly reactive intermediates upon UV irradiation and resulting in covalent bond formation between ligand and target protein[10]. Widely used PAL groups include aryl azides, benzophenones, and diazirines[10]. Recently, α-pyrones, pyrimidones[11], and tetrazoles[12]

(Figure 10.1.2) have been exploited as photoactive groups for ABPP studies that potentially show reduced non-specific labeling due to their unique modes of action[13]. Appropriate linkers, connecting the covalent affinity probe with a respective tag or bioorthogonal group for further complex enrichment or identification, also serve as a spacer to ensure optimal function of the target binding moiety on one side as well as the tag/bioorthogonal group function on the other. While these linkers usually consist of alkyl chains, polyethylene glycol (PEG) chains, or peptides[5] (Figure 10.1.2), some minimalist linkers include the photoactive moiety adjacent to the bioorthogonal group, having the advantage of leaving the original scaffold of the bioactive molecule mostly unchanged[14] (Figure 10.1.1). In some contexts, cleavable linkers have been used as well, as they offer advantages in target enrichment and isolation of proteins (not shown here)[15].



Figure 10.1.1 | General Structure of ABPs/A*f*BPs. Overview of common probe designs. ABPs/A*f*BPs consist of a small molecule-based protein binder (blue diamond) that is decorated with a reactive warhead (yellow star). A linker (black line) connects the target binding moiety with a tag (green oval) for further analysis/detection either *via* direct linkage or a bioorthogonal handle (dark/light brown structures). If a reactive warhead is not part of the target-binding structure, a minimalist linker can be applied, which would have the advantage of incorporating the warhead moiety as well as the bioorthogonal handle within the linker, therefore leaving the target binding molecule mostly unchanged.

As seen with minimalist linkers, a probe should ideally exhibit the same properties and target interactions as the respective bioactive molecule. Therefore, using bioorthogonal reactive groups instead of directly attaching the tag to a probe can be of advantage. Here, smaller moieties such as azides or alkynes replace bulky tags, which can not only lead to an enhanced cellular permeability but can also increase the versatility of a probe as many different and complex tags can be introduced after a probe bound its target(s)[16]. In ABPP, Staudinger ligation, click chemistry, or inverse electron-demand (IED) Diels-Alder reactions are often applied to attach tags to bioorthogonal handles[4], with the alkyne group being the most commonly used moiety for ABPs/A*f*BPs[3].

Generally, the choice of the reporter group is dependent on the desired function of a probe. Biotin is a widely applied tag since it can be used for detection and purification or enrichment of a target protein. In contrast, fluorescent or radio-labeled tags are limited to the detection of a protein; however, these groups offer several advantages over biotin as the read-out of the samples is often faster and more sensitive[5] (Figure 10.1.2).

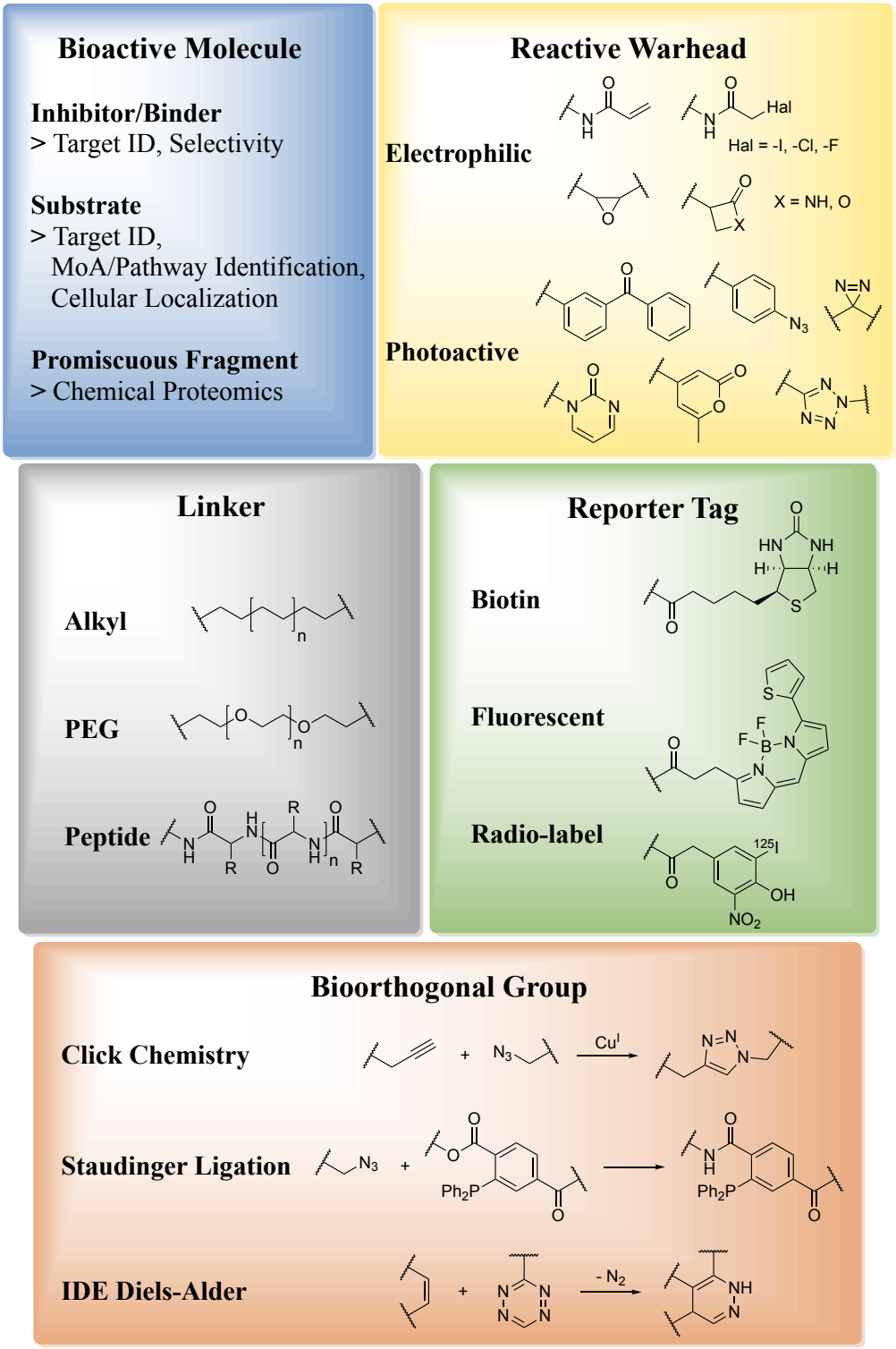


Figure 10.1.2 | Overview of Commonly Used ABP/A*f*BP Moieties. Examples of some ABPP probe moieties, mentioned in this chapter, are given (same color code as used in Figure 10.1.1). MoA; Mode(s) of Action.

Initially, ABPP probe-target enzyme complexes were analyzed using gel-based approaches, where the samples would be separated by 1D- or 2D-gel electrophoresis, followed by trypsin digest and mass-spectrometry analysis of the resulting peptides in order to identify the labeled proteins[1] (Figure 10.1.3). Later, using streptavidin- or antibody-coated beads for biotin or fluorescent tag pulldown, respectively, samples incubated with an ABP/A*f*BP could be directly digested by trypsin, affinity-purified using the beads, and then analyzed by mass-spectrometry after elution[17]. This procedure simplified the gel-based approach to a straightforward, LC-MS-focused approach for the identification of labeled active sites[17] (Figure 10.1.4).

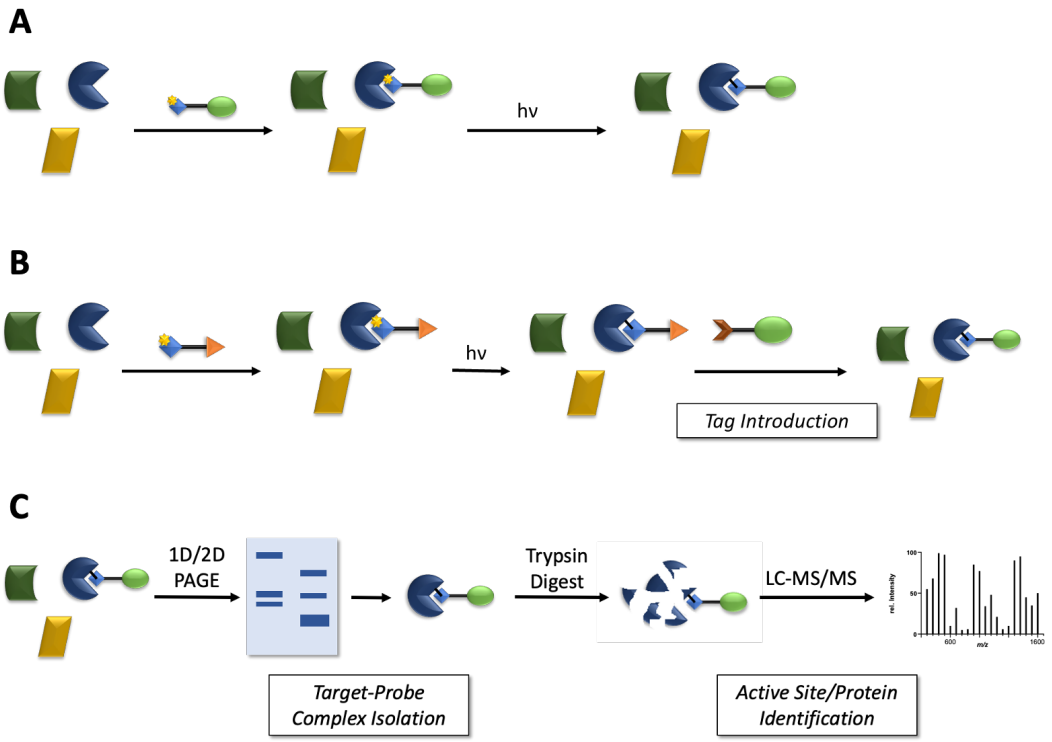


Figure 10.1.3 | ABPP Analysis of Target Binding Sites with Photoaffinity Probes. Complex protein mixtures, e.g. whole cells, are incubated with A*f*BPs (same color code as described in Figure 10.1.1). Subsequent UV irradiation causes formation of covalent target-probe complexes with either an already tagged probe (**A**) or with a probe that is equipped with a bioorthogonal handle for tag introduction in an additional follow-up step (**B**). The resulting labeled complexes can then be further analyzed. In earliest ABPP methods, this was achieved merely by gel electrophoresis and is now commonly coupled with complex isolation, trypsin digest, and LC-MS/MS analysis for active site/protein identification (**C**)[1].
PAGE, polyacrylamide gel electrophoresis.

A variation of this initial bead-based approach is the ABPP-multidimensional protein identification technology (MudPIT), in which tagged ligand-target protein complexes are first purified and then trypsin digested so that all fragments of the resulting target protein peptides can be analyzed by multi-layered LC-MS/MS, providing excellent resolution and quantitative read-outs[17] (Figure 10.1.4 B). Combining the initial bead-based approach with the MudPIT strategy led to the development of tandem orthogonal proteolysis (TOP) ABPP, in which purified target-ligand complexes are trypsin digested and the labeled

peptide fragment isolated in a second bead-based purification step, comparable to the MudPIT method[17] (Figure 10.1.4 C). TOP-ABPP was further improved by the use of isotope-labeled probes, resulting in better and more reliable read-outs – the so-called "isoTOP-ABPP" method[18]. Besides LC-MS(/MS), capillary electrophoresis, microarray analysis[17], and microfluidic chip-based[18] platforms were also applied to these ABPP methods.

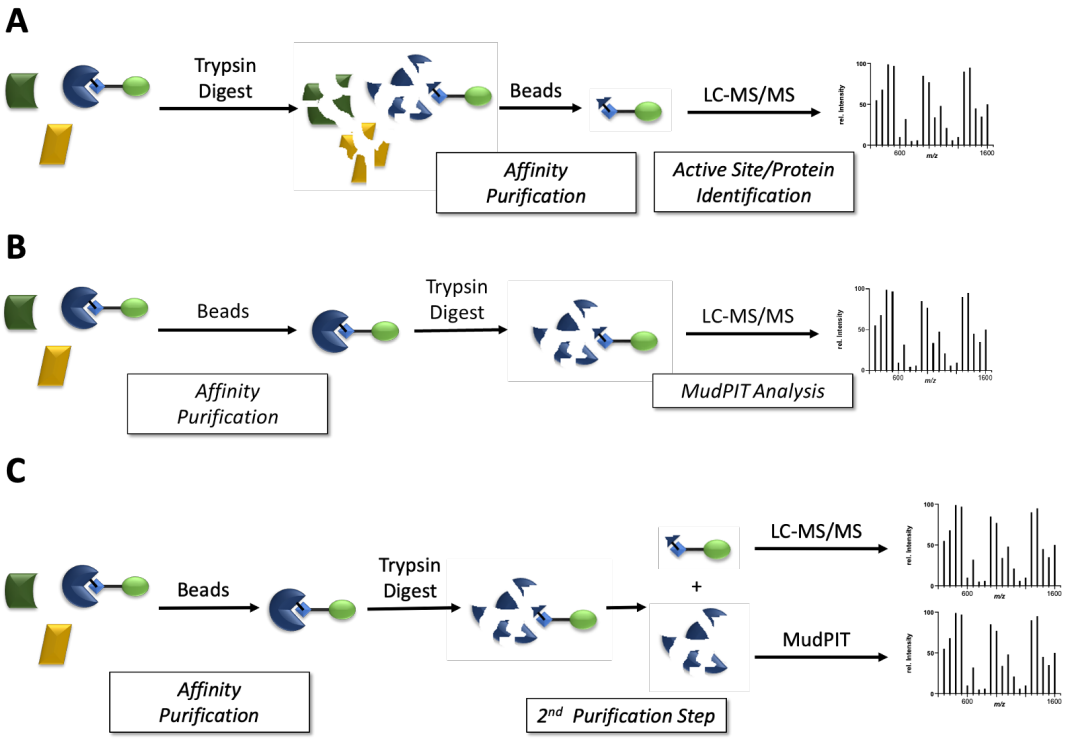


Figure 10.1.4 | Bead-based ABPP Strategies. A. Trypsin digest of the whole sample followed by affinity purification and LC-MS/MS analysis for active site/protein identification. **B.** Affinity purification of intact target-probe complexes, followed by trypsin digest and comparative analysis of all peptide fragments, serves as a target identification method, also providing quantitative read-outs (ABPP-MudPIT). **C**. Combination of A and B led to the development of TOP-ABPP. Following a first affinity purification of the intact target-probe complex and trypsin digest, a second affinity purification step separates labeled binding-site peptides from unlabeled peptides of target proteins. LC-MS/MS analysis of both fractions gives information about identity and binding site of the target protein as well as quantitative read-outs.

Comparative strategies using gel-based approaches can help identify (unknown) target proteins. For example, comparing labeling patterns of promiscuous A*f*BPs from healthy and diseased cells can provide valuable information about potential disease-associated targets[7] (Figure 10.1.5 A). Competitive ABPP approaches are used to identify small molecules that bind specific enzymes by outcompeting an A*f*BP[7] (Figure 10.1.5 B). Applying this competition platform with a fluorophore-equipped probe further improves the read-out of this approach to enable high-throughput screens (fluorescence polarization-ABPP; FluoPol-ABPP)[19], while improved fluorescent probes in quenched near-infrared fluorescent (qNIRF) ABPP are highly useful for diagnostics and monitoring purposes, since their detection is only possible after covalent binding of a target structure[18].

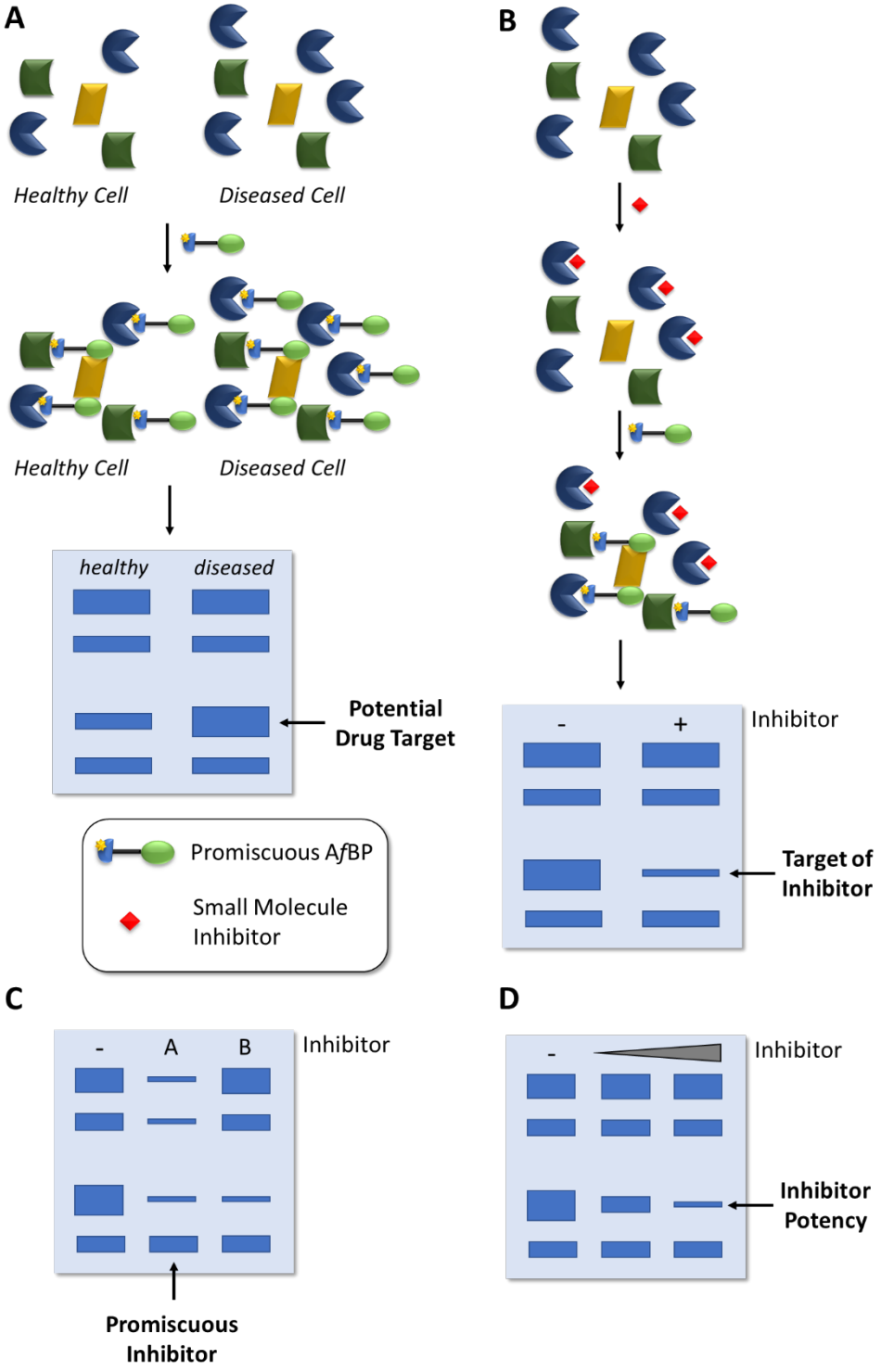


Figure 10.1.5 | Examples for ABPP Experiment Set-ups. A. Whole cells or cell lysates can be analyzed using promiscuous A*f*BPs. Comparison of gel electrophoretic analysis of healthy and abnormal cells can lead to the identification of potential drug targets in a comparative experiment set-up. **B.** Pre-incubation with a small molecule before labeling with a promiscuous A*f*BP can help in the identification of inhibitor targets in competitive experiment set-ups. By combination of competitive and comparative set-ups, selectivity profiles of different inhibitors can be compared (**C**), or, if different concentrations of the same inhibitor are employed (**D**), this method can be used to determine inhibitor potencies in complex biological media.

Combining comparative and competitive strategies by, for example, comparing competitive labeling patterns of different drugs or different drug concentrations, can indicate selectivity advantages or drug potency, respectively[5] (Figure 10.1.5 C and D). Further refinement of ABPP techniques, such as quantitative methods[20], including SILAC[21] or label-free methods[22], as well as organelle-specific proteomics[23,24], have added a broad variety of biochemical techniques to the toolbox for medicinal chemistry campaigns and drug discovery.

As already mentioned, ABPP strategies are widely used for drug target identification experiments, including small molecule selectivity profiling, due to their applicability in complex systems such as cells or tissues. A study analyzing FDA-approved drugs with covalent target-binding modes of action, such as tetrahydrolipstatin, monoamine oxidase (MAO) inhibitors, acetylsalicylic acid, ibrutinib, β-lactams, omeprazole, and others, highlights the importance and power of these technologies[25]. In these screens, many off-targets were identified, which were either useful in the context of other diseases, such as fatty acid synthetase (FAS) off-target activity of tetrahydrolipstatin, or demonstrated high promiscuity of the drug, as in the case of acetylsalicylic acid[25].

Screening for off-targets is especially important for enzyme families with highly conserved binding sites, such as kinases. For example, enzyme family specific "kinomebeads", which are covered with a mixture of different promiscuous kinase inhibitors, are able to identify and bind a great variety of kinases and are therefore used to study selectivity profiles of small molecule kinase inhibitors *via* competition experiments[26]. Other successful probe-based strategies to characterize target proteins of small molecules include the use of non-specific, lipid-based[27], or cysteine-directed probes[28]. Over time, ABPP strategies have been applied to identify potential new drug targets or inhibitor scaffolds for a variety of enzymes and enzyme classes, such as serine hydrolases, proteases, protein kinases (including PI3K), phosphodiesterases (PDEs), cytochrome P450s (CYPs), histone deacetylases, as well as proteins involved in infectious diseases, e.g. malaria[9]. Besides target identification and selectivity profiling of small molecules, the discussed techniques have also been applied for the characterization of drug metabolites *in vivo*[29], including CYP activity of small molecules, and can provide important information in the context of potential drug interactions[30].

Another important application of ABPs/A*f*BPs is the investigation of enzymatic pathways and functional studies of proteins. For example, sialic acid 9-*O*-acetylesterase (SAE) was identified as a serine hydrolase despite its lack in sequence homology to common serine hydrolases, using an A*f*BP[31]. Similarly, probes derived from protein substrates were designed for a variety of other enzyme classes, including agmatine[32] and arginine deiminases, acetyl-coenzyme A (CoA)-binding enzymes, protein tyrosine phosphatases, nitrilases, protein kinases, proteases, neuramidases, sulfatases[33], cysteine-, aspartyl-, and metallo-hydrolases, glycosidases, histone deacetylases, as well as oxidoreductases[34], thus enabling the study of their function. In cancer, ABPP techniques have helped to elucidate autophagy mechanisms[35] and putative driver genes[36] in human cells. Beyond human cell culture, ABPP has also been used to investigate functions of plant-based enzymes[37] and microbial processes, such as bioenergy conversion[38], microbial communities[39], virulence factors[40], and bacteria-host interactions[41], clearly highlighting the richness and diversity in ABPP applications. Some other implementations of ABPP methods cover mapping of

protein binding sites[42] and bioimaging[3], including diagnostics and monitoring of diseases[18].

Photo-Affinity Labeling (PAL) Warheads

PAL has been a successful strategy of turning reversible ligands of proteins into covalent chemical probes to study target profiles of bioactive compounds as well as to interrogate functions of a protein of interest. Since this chapter focuses on PAL probes, different PAL strategies and modes of action will be discussed with the emphasis on diazirines as photoactive warheads.

A*f*BPs and their structures have already been described (Figures 10.1.1 and 10.1.2). Generally, photoactive groups are introduced to the scaffold of reversible protein binders or the linker of a probe and are meant to covalently label the bound target enzyme upon UV irradiation. Using photoactive warheads for irreversible target labeling has already been reported in the early 1960s, describing a carbene-generating diazo warhead[43]. Within the following decade, aryl azides, benzophenones, and diazirines were established as PAL groups and are still the most commonly used warheads for PAL today[43]. However, in the past decade, warheads consisting of novel scaffolds such as pyrones, pyrimidones[11], and tetrazoles[12] were added to the PAL toolbox (Figure 10.1.2). Incorporating PAL into ABPP strategies has thus resulted in the development of A*f*BP probes for the investigation of binding sites of small molecules[44], and were used to investigate proteins, such as vasoactive intestinal peptide receptor 1 (VPAC1)[45] and nicotinic receptors[46], and also PPIs[47]. Furthermore, functional studies of plasma proteins[48] or even whole biomembranes[49] were performed using PAL-based probes, and applications of A*f*BPs in target identification investigations for proteins such as carbohydrate-binding proteins, kinases, and phosphatases[50] can be found in the literature. Notably, A*f*BPs are also useful bioimaging agents[13], highlighting the versatility of this technique.

Despite its application to activate A*f*BPs, UV irradiation can generally cause damage to biomolecules. Since proteins absorb UV light at 200 and 280 nm due to their peptide bonds and aromatic amino acids, respectively[51], longer UV irradiation wavelengths would be needed for the activation of a photoactive warhead to avoid damage to proteins. For example, benzophenones form reactive diradical species after irradiation at 350-360 nm[51]. However, the long irradiation times necessary for sufficient conversion lead to an increase in unspecific labeling events[51]. Moreover, their bulky structure makes it difficult to incorporate benzophenone moieties into small molecule structures without potentially altering the probes' target-binding properties relative to the parent compounds[51] (Figure 10.1.6). Nevertheless, benzophenones have been helpful tools in drug development; for instance, for studying the target engagement of PDE inhibitors[52]. Here, by incorporating the bulky photoactive group within the bioorthogonal linker of the probe, interference in the probe's binding properties was avoided[52].

More commonly used PAL warheads are aryl azides, as they are easily accessible both commercially and synthetically[51]. Unfortunately, aryl azides come with several disadvantages (Figure 10.1.6). For example, wavelengths as short as 250 nm are required to yield the reactive nitrene species, causing potential damage to proteins and other biological molecules in the samples, while their labeling efficiency is generally lower than that of carbenes[51]. However, it was found that longer irradiation wavelengths of ~340 nm can be used if the aryl component carries a nitro group[53]. Furthermore, to avoid

undesirable rearrangement reactions after irradiation, fluoro-substituted analogs have shown to prevent phenylnitrene formation[54]. Despite these issues, including their lower chemical stability as compared to other photoactive moieties[53], aryl azide-based photoaffinity warheads have been successfully incorporated into peptide substrates for the active conformation of the Abl tyrosine kinase, due to their ability to function as phenylalanine side chain substitutes[55]. Incorporation into linkers of DNA probes for the identification of DNA-binding proteins[56] or diubiquitin probes to study DUBs[57] are some other examples of the application of aryl azide-based warheads, highlighting their value as tools for the interrogation of biological systems.

Because of the described disadvantages of benzophenones and aryl azides, diazirines are the most commonly applied PAL warheads to date[51]. While benzophenones and aryl azides require an aromatic ring system as part of their photoactive warhead, diazirines can be aryl- or alkyl-based, with the latter being not only very small in size but also providing additional desirable features for A*f*BPs. For example, long wavelengths of 350-355 nm are sufficient to induce the formation of carbenes from diazirines, which also exhibit relatively low unspecific labeling due to their short half-life[51]. While a short half-life thus results in overall better selectivity of a probe, it also results in low overall labeling yields[51]. Furthermore, diazirines show favorable stability properties for optimal handling and synthetic use[51]. In summary, considering the highly reactive nature[58] as well as the excellent labeling selectivity[59] of the carbene intermediates, the labeling properties of diazirines have been shown to be superior to other PAL groups[58] (Figure 10.1.6).

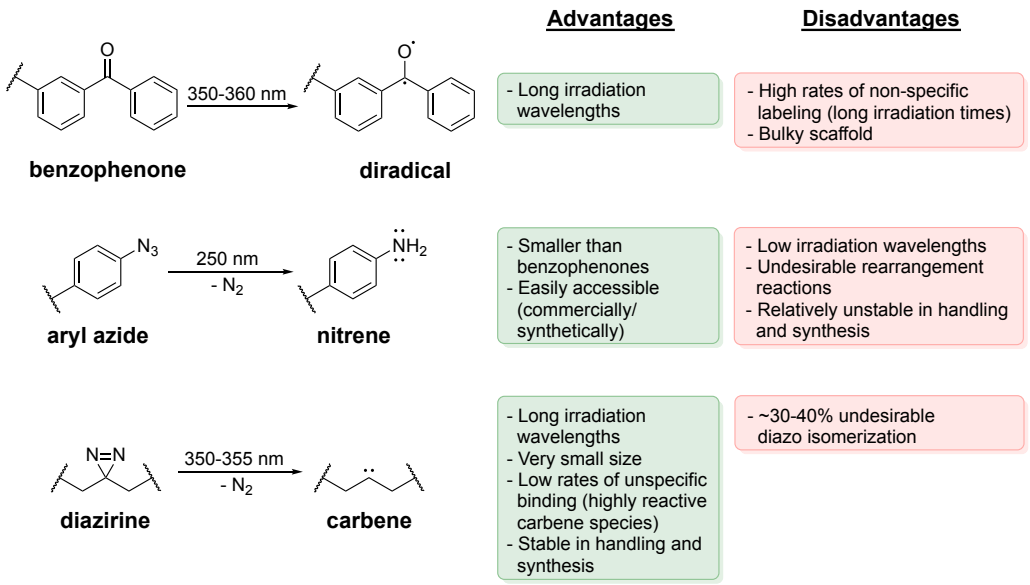


Figure 10.1.6 | Advantages and Disadvantages of Common PAL Warheads. A summary of important advantages (green boxes) and disadvantages (red boxes) for the common PAL warheads benzophenones, aryl azides, and diazirines is given.

As a result, diazirine-labeled probes have been used in a vast variety of studies, including the mapping of binding sites of proteins such as mammalian *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) hydrolase[60], Na^+/K^+-ATPase[61], sweet taste receptor[62], lactate dehydrogenase[63], ion channel binding sites of general anesthetics[64], and topoisomerase II[65]. Other applications include the target identification of small molecules

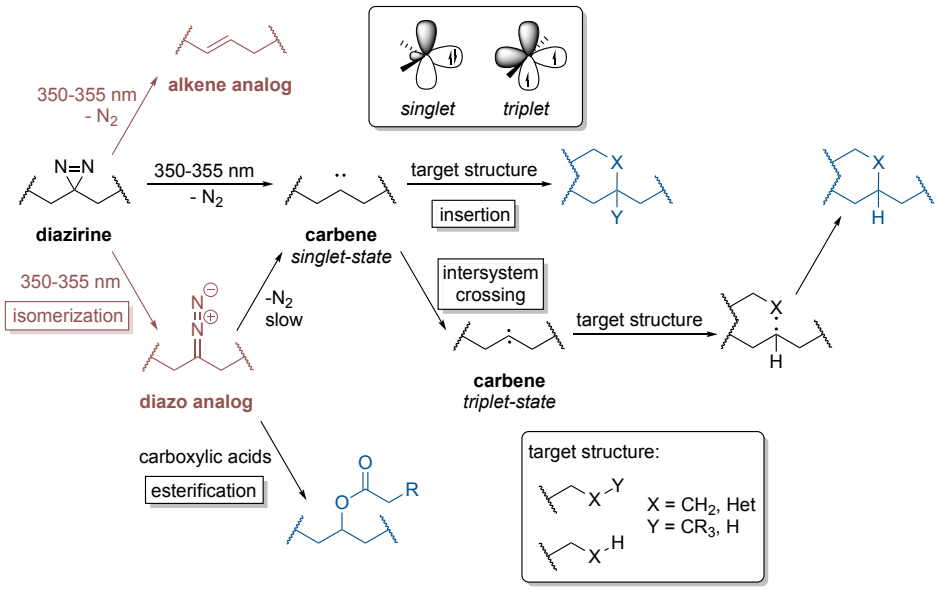
or also larger biomolecules (for example RNA) for selectivity, metabolism, and mode of action studies. These include for example kinase inhibitors[14], quorum sensing autoinducers[66], ethanol[67], siRNA[68], microRNA (miRNA)[69,70], DNA[71], oligonucleotides[72], $\Delta^{8/9}$-tetrahydrocannabinol (THC)[73], histone-based peptides[74], ligands of membrane proteins[75], such as isoprenoid[76] and GLUT1 ligands[77], sialic acid[78], nicotinic acid adenine dinucleotide phosphate (NAADP)[79], PI-4,5-P_2[80], phosphotyrosine as a protein recognition feature[81], sphingosine[82], and neurosteroids[83]. Furthermore, diazirine-labeled probes have been applied to protein isolation and purification techniques[84] as well as to determine the location of small molecules within cells, for example, in the investigation of cholesterol trafficking[85]. In certain contexts, A*f*BPs can be developed as potential new drugs themselves, as shown by the oncologic therapeutic RH1 (2,5-diaziridinyl-3-(hydroxymethyl)-6-methyl-1,4-benzoquinone), which is transformed into a DNA-alkylating agent upon metabolic activation by different reductases, which are commonly overexpressed in certain solid tumors like lung, colon, pancreatic, and breast cancers[86].

Upon UV irradiation, aryl and alkyl diazirines form highly reactive singlet carbene species with a half-life of ~1 ns while releasing molecular nitrogen (N_2)[87]. Because singlet carbenes have an empty p-orbital, they readily form bonds with nearby C-C/H or heteroatom-H bonds *via* insertion reactions[88]. However, about 30% of diazirines are commonly isomerized into diazo groups, which slowly release nitrogen to form singlet state carbenes, resulting in a majority of non-specific binding events[87]. Notably, recent studies have shown that the diazo intermediates can form covalent bonds with carboxylic acids, as found in glutamic and aspartic acid side chains, to form esters[89]. To reduce non-specific labeling, trifluoromethyl phenyl diazirines can be applied as they exhibit stabilization of their respective diazo intermediate compared to simple phenyl diazirines[87]. In some cases, such as for trifluoromethyl phenyl diazirines[88] or singlet state carbenes, triplet carbenes are formed, which react with heteroatom/C-H bonds *via* a radical intermediate[87]. While current literature suggests that alkyl-derived diazirines undergo diazo isomerization for ~40% of the probe – in contrast to ~30% for aryl-based diazirine analogs – other intramolecular rearrangement reactions are also reported, including rearrangements in the excited state (RIES) yielding C=C-double bonds[90], indicating that the diazirine disintegration upon UV irradiation is potentially even more complex (Scheme 10.1.1).

The main strategy to access diazirines synthetically is through the conversion of a carbonyl moiety into a diaziridine, followed by its oxidation to the desired diazirine (Scheme 10.1.2). Several synthetic routes to yield diaziridines from ketones or aldehydes, including co-treatment of a carbonyl with ammonia and hydroxylamine-*O*-sulfonic acid (HAOSA) as well as oxidation steps using basic iodine solutions or silver nitrate, are reported[88]. The aromatic trifluoromethyl phenyl diazirine is most commonly synthesized from the respective carbonyl, using hydroxylamine and tosyl chloride to yield a tosylated oxime, which can then be treated with liquid ammonia (NH_3) resulting in a diaziridine intermediate[91]. For aliphatic diazirines, commonly employed synthetic routes involve treating ketones with liquid ammonia, resulting in an imine, which is subsequently converted into a diaziridine intermediate *via* reaction with HAOSA and oxidized with iodine in the presence of triethylamine (TEA)[92]. A recent paper further describes simplified one-pot versions of those conditions, using potassium *tert*-butoxide (*t*-BuOK) or potassium hydroxide (KOH) and air to convert the diaziridine, formed by treating a ketone

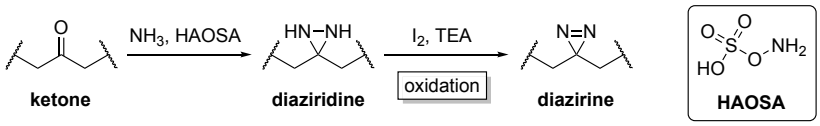
with HAOSA and liquid NH_3, to the respective diazirine[93]. Diazirine-labeled precursors are then incorporated into the ligand of interest. These synthetic procedures almost exclusively consist of amide formations, alkylations, and esterifications, while further improvements were made to broaden the applicability of diazirine chemistry to other common reaction types in drug discovery, including Suzuki couplings[94] and Staudinger ligations[95].

Scheme 10.1.1 | UV Irradiation of Diazirines and their Products[a]



[a] Undesirable isomerization reactions or transformations are depicted in brown. Final ligand-target complexes are highlighted in dark blue. Het, heteroatom.

Scheme 10.1.2 | General Synthetic Route to Access Diazirines from Ketones



10.1.2 Antibiotics – An Overview

Documents from 3,500 years ago provide evidence that humans had found ways to treat infectious diseases, using molds from old bread and certain soils to prevent infections of open wounds[96]. However, it was not until the discovery and study of bacteria in the 17th century as well as the understanding of the connection between pathogens and diseases introduced by Robert Koch and Louis Pasteur in the late 19th century that investigations and developments of antibacterial agents could be undertaken[97] (Figure 10.1.7). For example, starting with Alexander Fleming's fortunate observation of penicillin's antibacterial effects from an unintendedly grown mold on a petri dish in 1928[98], the discovery and development of penicillin is probably the most famous example of how antibiotics have entered and elevated medicine in the 20th century.

Notably, in the early 1900s, Emmerich and Löw had already isolated an antibacterial mixture from *Pseudomonas aeruginosa* to treat patients with infectious diseases, while Paul Ehrlich was the first to perform a drug screen for antibacterial agents[96]. Even though he initially screened for dyes that would stain bacteria, Ehrlich uncovered the antibacterial activity of what later became the antibiotic class of sulfonamides, which inspired him to screen other compound libraries and led to the discovery of the arsenic prodrug antibiotic salvarsan for the treatment of syphilis in 1910[96]. Based off Ehrlich's work, Gerhard Domagk refined the antibacterial dyes, resulting in the development of sulfonamides in the early 1930s[96] and marking the beginning of the era of efficient "anti-infective medicine"[98]. Another important development, marking this period, was the commercial production of a penicillin-containing liquid as an antibacterial drug in the 1940s – following its initial isolation by Florey and Chain at Oxford University –, which became a famous life-saving agent in World War II[98].

From the 1940s to the 1960s, almost 30 new classes of antibiotics were discovered, marking this era as the "Golden Age" of antibiotics[96] (Figure 10.1.7). This incredible progress in the field of antibiotics is mainly owed to Selman Waksman, who intensely studied *Actinomycetes* found in soil samples[96]. Among the identification of various antibacterial secondary metabolites from this class of bacteria, 55% of newly discovered antibiotics in the period of 1945 to 1978 were isolated from a single genus of *Actinomycetes* - the *Streptomyces*[96]. One important product of these discoveries was streptomycin, a potent agent in the treatment of tuberculosis (TB) and even multi-drug-resistant TB, which is still in use today[99].

Interestingly, from the 27 newly identified antibiotic classes in the 20 years of the Golden Age, ~70% were natural products or derived from natural products[96]. These numbers highlight the richness and value of natural sources for the identification of bioactive compounds. In parallel to the discovery of naturally occurring antibiotics, medicinal chemistry campaigns arose, which initially aimed to improve identified antibacterial scaffolds to make them useful drugs[100]. For example, the semi-synthetically produced dihydrostreptomycin exhibits similar antibacterial activity to streptomycin, while being more stable[100]. This so-called "Medicinal Chemistry Era", starting out with semi-synthetic approaches in the 1940s, further initiated fully synthetic strategies to uncover novel antibacterial scaffolds, leading to the development of chloramphenicol as a clinically useful antibiotic in 1949[100] (Figure 10.1.7). Due to the ever-rising issue of drug-resistant bacteria, coupled with the extreme drop in the development of new antibacterial scaffolds (especially

antibiotics with novel modes of action) after the Golden Age[96], other strategies to tackle these challenges are needed and will be discussed in the course of this chapter.

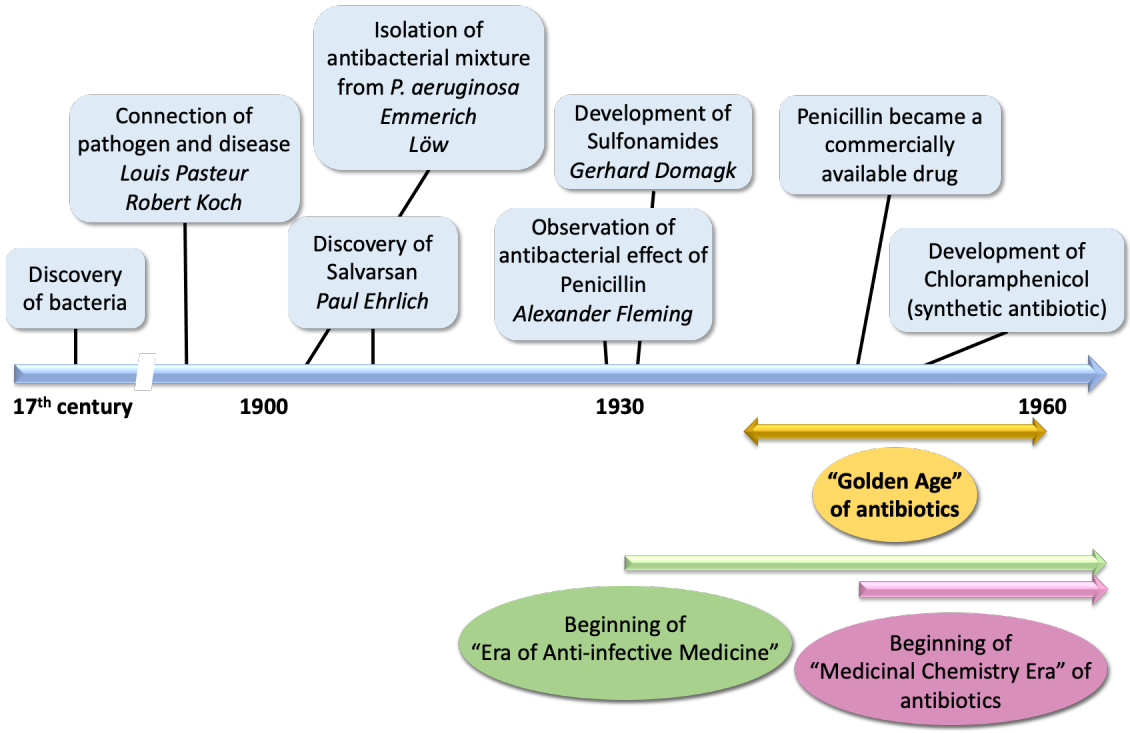


Figure 10.1.7 | Timeline of Early Antibiotic Discovery and Development. A timeline showing important developments and discoveries in antibacterial research (blue boxes; important scientists are listed in *italic*), mentioned in the text. Key eras are marked underneath the main timeline.

Common Antibiotic Targets

As a result of Waksman's efforts and the Golden Era, combined with medicinal chemistry efforts and modern approaches, antibiotics targeting key processes of bacterial cell growth, metabolism, and survival mechanisms have already been developed. To avoid undesired off-target effects in human cells, features unique for prokaryotes are commonly modulated by antibiotics in clinical use and development (Figure 10.1.8). For example, in order for DNA to replicate, the DNA double helix has to be temporarily unwound to enable enzymes to bind to ssDNA. The proteins responsible for this mechanism are topoisomerases, which can be divided into class I and class II enzymes, which either perform single-strand or double-strand breaking of the DNA, respectively[101]. Gyrase and topoisomerase IV are the two class II topoisomerases that can only be found in bacteria and were identified to be selective targets of quinolone-type antibiotics[101]. Recent efforts have further identified other inhibitors of this mechanism, such as gyrase-specific aminocoumarines[102], simocyclinones[103], and non-quinolone scaffolds[101], while topoisomerase I inhibitors are also being investigated[104]. Other DNA-targeting antibiotics include nitrofurans and azoles, which act through DNA damage mechanisms, along with guanine-binding antibiotics of the phenazine-type[96].

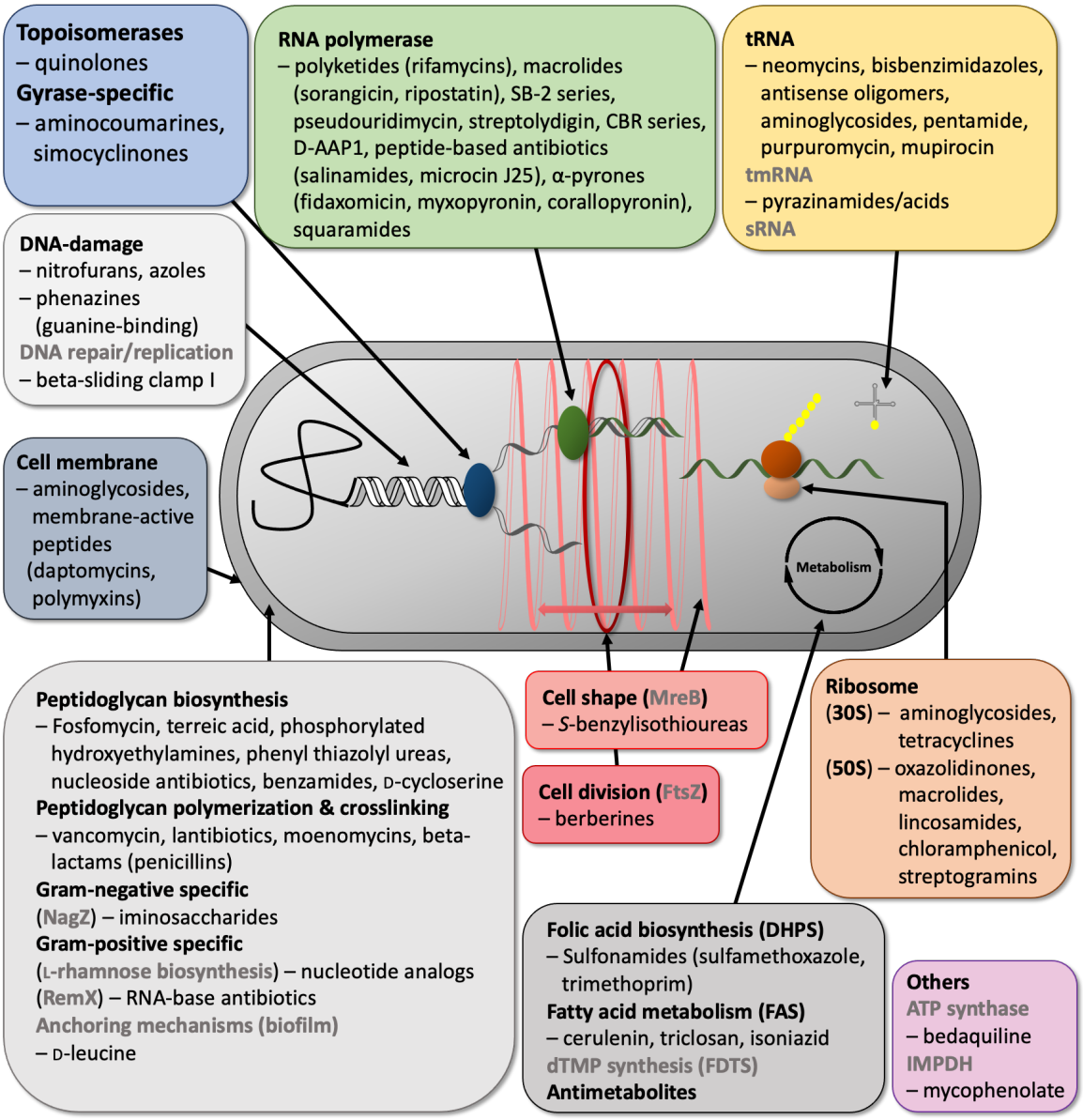


Figure 10.1.8 | Targets of Antibacterial Agents. Established targets of antibiotics (bold, black), as well as currently investigated targets (bold, dark gray) are summarized with examples of respective antibiotics mentioned in this chapter. Common bacterial targets include DNA (DNA-damaging agents, topoisomerase inhibitors), RNA polymerase, tRNA, ribosomes (mainly 30S and 50S unit), the cellular membrane, metabolic pathways (folic acid and fatty acid biosynthesis, as well as antimetabolites), and the bacterial cell wall (peptidoglycan biosynthesis, polymerization, and crosslinking). Other targets, such as tmRNA and sRNA, dTMP biosynthesis, cell division and shape processes (FtsZ, MreB), Gram-specific cell wall features (NagZ, L-rhamnose biosynthesis, RemX), and cell wall anchoring mechanisms, as well as ATP synthase and IMPDH are discussed in the literature.

The first step in protein synthesis is the transcription of DNA to mRNA. Hence, the bacterial RNA polymerase poses an attractive target for antibiotics, with multiple subunits providing various potential binding sites[105]. For instance, rifamycins (polyketides), are well-known inhibitors of this RNA polymerase binding to the DNA/RNA-channel[105]. Furthermore, other antibacterial agents have been reported to inhibit this step of protein

biosynthesis including sorangicin (macrolide), antibiotics from the SB-2 series (transcription initiation inhibitors), pseudouridimycin (nucleoside analog), salinamides, streptolydigin, CBR-series-based antibiotics, D-AAP1, microcin J25 (peptide), as well as fidaxomicin, myxopyronin, corallopyronin, ripostatin, and squaramides, which prevent necessary conformational changes of the RNA polymerase[105].

Also, ribosomes are targets of certain antibiotics e.g., due to the pronounced differences of prokaryotic to eukaryotic ribosomal RNA (rRNA). The most well-known examples of ribosome-targeting antibiotics, mainly targeting rRNA substructures, are aminoglycosides and tetracyclines for the 30S subunit (16S rRNA), and oxazolidinones, macrolides, lincosamides, chloramphenicol, and streptogramins for the 50S subunit (23S rRNA)[106-108]. Transfer RNAs (tRNAs), being actively involved in protein synthesis, are targeted by antibiotics *via* various mechanisms, including prevention of tRNA maturation by binding premature tRNA (neomycins and bis-benzimidazoles) or *via* ribonuclease P (RNase P) inhibition (antisense oligomer), direct binding of tRNAs (aminoglycosides, pentamidine, purpuromycin), aminoacylation prevention by inhibition of aminoacyl tRNA synthetases (aaRSs) (mupirocin), and other mechanisms that are currently being investigated[109,110].

So far, almost every step in protein translation, such as initiation, elongation, and termination, has been found to be targeted by various antibiotics[111], highlighting the importance of this mechanism for bacterial survival and the opportunity for achieving selectivity over eukaryotic translation mechanisms *via* several bacteria-specific target structures. Finally, other RNA molecules, such as tmRNA – hybrid t/mRNAs –, which are involved in the recycling of ribosomes and the removal of incorrectly translated proteins, have been targets of pyrazinamide or respective acid analogs, while small regulatory RNAs (sRNAs) are also being considered potential new targets for antibiotics[107].

Additional successful strategies of targeting bacteria have aimed to inhibit metabolic pathways, including folic acid and fatty acid metabolism or nucleoside biosynthesis. For example, folic acid is an important cofactor for various enzymatic processes in humans and bacteria[112]. Since substrate-related folic acid analogs are commonly used as targets to treat eukaryotic diseases, such as cancer, one strategy to selectively target bacteria is to inhibit their unique dihydropteroate synthase (DHPS) using sulfamethoxazole and trimethoprim, a common combination treatment known as "co-trimoxazole"[112].

Enzymes involved in fatty acid metabolism of prokaryotes are also different from eukaryotes as they use type II instead of type I FAS, making this essential metabolic pathway an attractive target for antibiotics[113]. For instance, the biosynthetic assembly of fatty acid chains is performed by enzymes such as FabH/F/B and can be targeted by antibiotics like cerulenin, while other enzymes involved in reduction reactions, including FabI enoyl-acyl carrier protein (ACP) reductase and homologs, are targeted by triclosan and isoniazid[113]. Moreover, other pathways, which are part of fatty acid metabolism, such as the bacterial acetyl-CoA carboxylase and acetyltransferases, are currently being investigated as potential antibiotic targets[113].

The flavin-dependent thymidylate synthase (FDTS) is a prokaryotic-specific enzyme responsible for the *de novo* synthesis of 2'-deoxythymidine-5'-monophosphate (dTMP), which presents another promising target for antibiotics with first inhibitors being developed[114]. Furthermore, antimetabolite strategies were successfully applied to inhibit effects of nitrogen metabolism by phosphinothricin (glufos; glutamic acid antimetabolite) or tabtoxinine-β-lactam (TβL), L-threonine biosynthesis by 2-amino-5-phosphono-3-

pentenoic acid (APPA; homoserine phosphate antimetabolite), the isoprenoid biosynthetic pathway by FR-33289 (deoxyxylulose phosphate antimetabolite), and CoA-dependent mechanisms by CJ-15,801 (pantothenic acid antimetabolite), but also include broadly acting antimetabolites such as alanine and arginine antimetabolite, argolaphos A[115].

One of the most obvious differences between prokaryotic and eukaryotic cells – besides the absence or presence of a nucleus – is the bacterial cell wall that is missing in eukaryotic cells. Since the antibacterial target discussed in this thesis is part of the bacterial cell wall's biosynthetic pathways, these and other bacterial cell wall targets will be discussed in detail in following chapters.

Non-specific disruption of the integrity of the cellular membrane is the mode of action of a variety of different antibiotics. For example, aminoglycosides or peptides, such as daptomycin and its analogs (lipopeptides), as well as polymyxins, potentially increase cellular permeability by polarization mechanisms of the bacterial membrane[115]. Generally, these antibiotics are proposed to assemble within the membranes to form pores, thus increasing diffusion of i.a. antibiotic agents[115]. This mechanism has also been confirmed for antimicrobial agents naturally occurring on the human skin, such as beta defensin 2[115].

Besides the described, most explored target strategies of antibacterial therapeutics, recent progress has been made in the identification of other promising targets for novel antibiotics, such as ATP synthase that is inhibited by bedaquiline[116], prokaryotic guanosine 5'-monophosphate (GMP) synthesis *via* inosine 5'-monophosphate dehydrogenase (IMPDH) inhibition by mycophenolate[117], bacterial actin-homolog MreB (from murein region "e" B gene) involved in bacterial cell shape and division processes that is inhibited by *S*-benzylisothiourea compounds[118], as well as the essential cell division mediator filamenting temperature-sensitive protein Z (FtsZ) that is inhibited by berberine[119]. Targeting FtsZ-ZipA (Z interaction protein A) PPIs and other PPI-inhibitor strategies (i.e. β-sliding clamp inhibitors), as well as disrupting DNA replication and repairing mechanisms[120], represent new approaches for the development of antibiotics. Lastly, some studies have implicated that most antibiotic treatments lead to oxidative stress in the bacteria cell ultimately causing cell death, which suggests that the potency of antibiotics could be increased by co-treatment with other agents targeting cellular stress response pathways[121].

Antibiotic Resistance – An Old but Ever-growing Problem

Figure 10.1.9 shows a timeline comparing development and introduction of an antibiotic class vs. first appearance of bacteria exhibiting resistance to that class[108]. Notably, most resistances are developed right after introduction of the antibiotic class, as in the case of aminoglycosides, glycopeptides, or quinolones, while some antibiotic resistances already occurred during the developmental stages before their introduction, as seen with tetracyclines and lipopeptides. This data highlights how fast and inevitably bacteria are able to evolve to overcome antibiotics' modes of action; a phenomenon that naturally also applies to specific antibiotics within a specific class[122].

Common mechanisms of resistance (Figure 10.1.10) include the decrease of antibiotic penetration through changes in pore-forming outer membrane proteins (Omps), which are crucial for the uptake of antibiotics such as β-lactams[123]. In a similar matter, bacteria can develop an increased ability to actively export antibiotics, which is seen for a majority of antibiotic classes including tetracyclines, β-lactams, macrolides, aminoglycosides,

sulfamides, lincosamides, chloramphenicol, and others[124]. For this, efflux pumps (Figure 10.1.11), such as small multidrug resistant (SMR) transporters, major facilitator superfamily (MFS) transporters, ATP-binding cassette (ABC) transporters, or multidrug and toxic compound extrusion (MATE) transporters, can be found among Gram-negative and Gram-positive bacteria, whereas resistance nodulation division (RND) transporters are only seen in Gram-negative prokaryotes[124].

Intracellular processes of resistance can include overproduction of the target enzyme (sulfonamides)[124], mutations of the targeted structures (quinolones)[125], non-genetic modifications of the target protein (chloramphenicol, quinolones, and macrolides)[126], and bypassing of the targeted pathways (sulfonamides and trimethoprim)[124]. Antibiotics can also be direct targets for antibiotic resistance mechanisms. One well-known example of this strategy is the deactivation of β-lactams by a vast variety of β-lactamases found in bacteria[127]. Ultimately, prokaryotes have developed community-based approaches to protect themselves from antibiotic agents, such as biofilms[128]. These films are found in surface-based infections and are particularly hard to treat, since antibiotics struggle with not only the physical barrier created by the film matrix but also the increase in mutually shared resistance mechanisms of bacteria within a biofilm, which are communicated by elaborate signaling pathways, including quorum sensing networks and available free DNA[128].

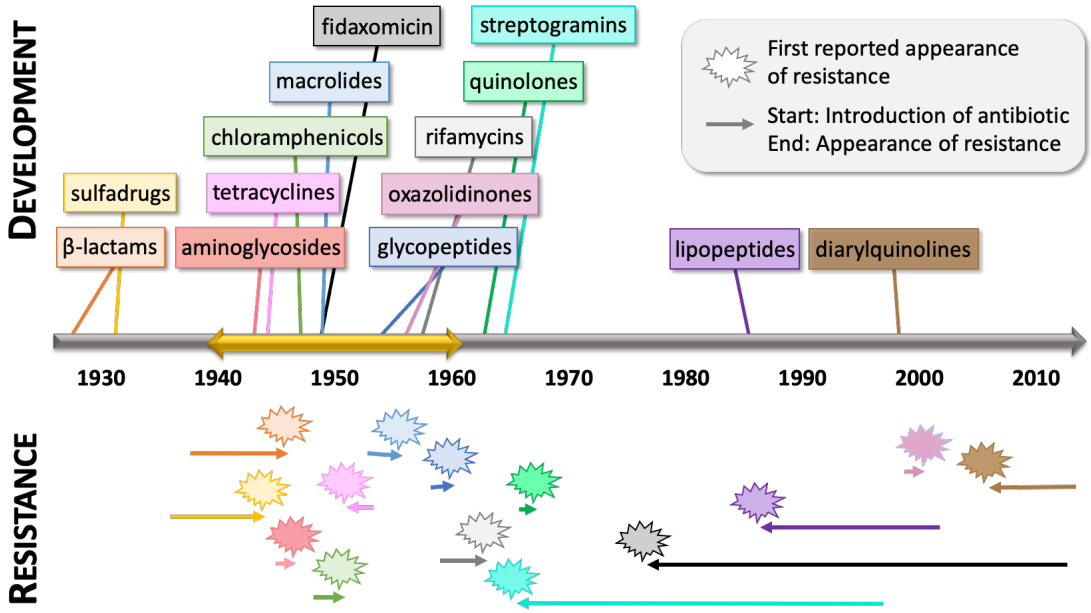


Figure 10.1.9 | Development of Antibiotics and Appearance of Resistance. A timeline showing the development (above timeline) of different classes of antibiotics (color-coded). Below the timeline, first appearances of resistant bacteria and the relation to the introduction of the respective antibiotic class are depicted using the same color code. The arrows reflect the relationship between introduction of the antibiotic class and the first appearance of resistant bacteria, in which the start of the arrow marks the introduction and the end of the arrow the appearance of resistance (highlighted by explosion bubble). Therefore, arrows going from left to right represent antibiotic classes, where antibiotic resistance appeared after their introduction, while arrows going from right to left show antibiotic classes, for which development of antibiotic resistance already occurred before their introduction. The Golden Age is marked as a yellow two-sided arrow. The graphic is based on information taken from Lewis, *Nat. Rev. Drug Discov.* **2013**, *12* (5), 371-387[108].

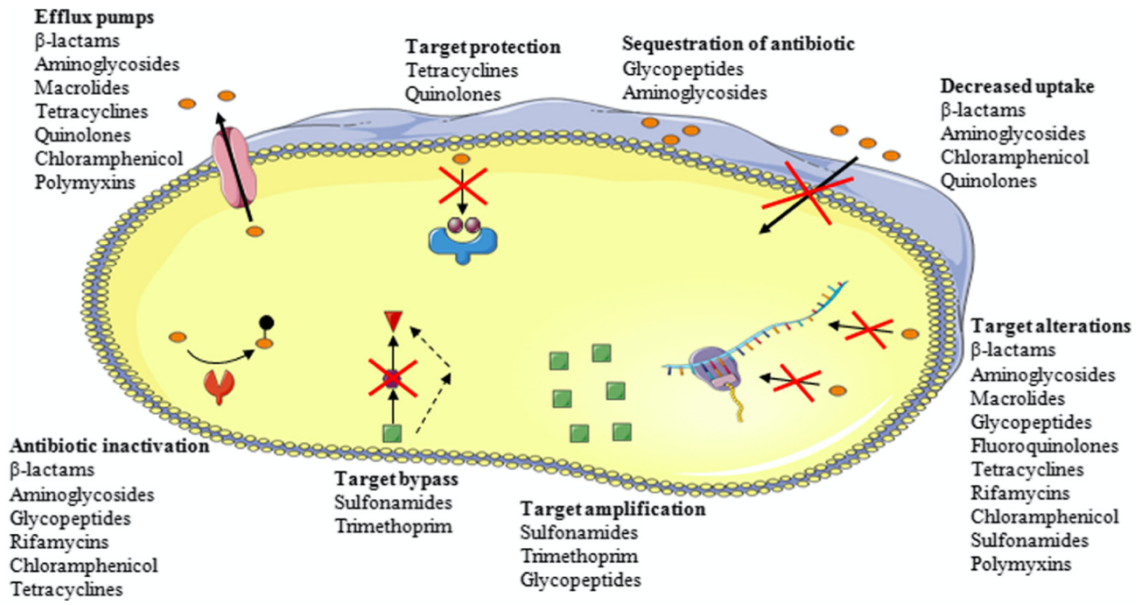


Figure 10.1.10 | Bacterial Resistance Mechanisms. Summary of the most common resistance mechanisms (affected antibiotics listed below) developed by bacteria to fight antibacterial treatments. Taken from Yilmaz *et al.*, *Biochem. Pharmacol.* **2017**, *133*, 43-62[124].

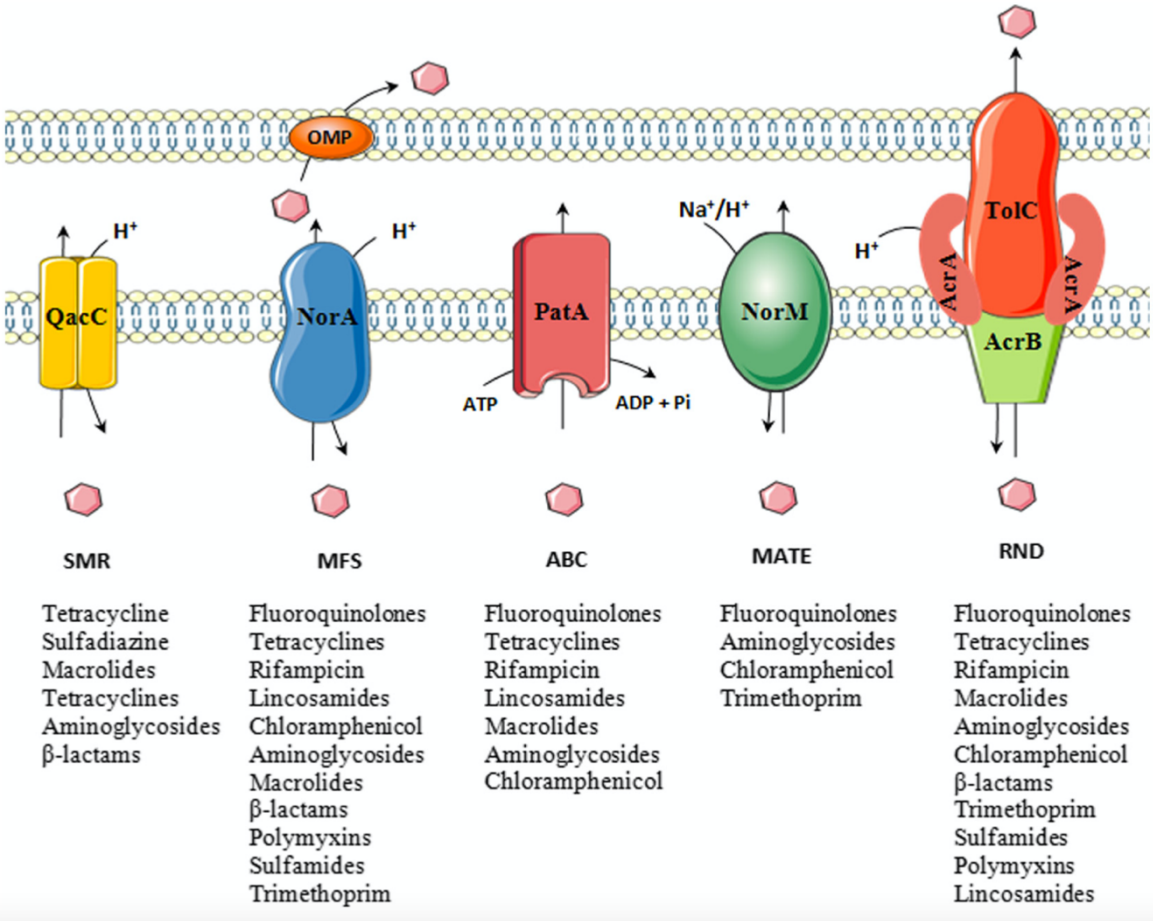


Figure 10.1.11 | Bacterial Efflux Pumps. Overview of common efflux pumps involved in antibiotic resistance mechanisms. Antibiotic classes that are substrates to these transporters are mentioned below, highlighting the importance of this well-established bacterial defense mechanism for antibiotic resistance. Taken from Yilmaz *et al.*, *Biochem. Pharmacol.* **2017**, *133*, 43-62[124].

Several antibiotic resistance mechanisms have their origin in changes in the prokaryotes' genes or gene expression. For example, one important family of transcription factors, responsible for the upregulation of efflux pumps or downregulation of transporters for cellular uptake of antibiotics, is the multiple antibiotic resistance repressor (MarR) family, whose members are able to recognize several different classes of antibiotics[129]. Other regulators of protein expression are sRNAs, which have been implicated in regulating drug uptake and efflux, as well as target modifications, cell wall specific processes, biofilm production, and transcription factors[130]. Based on the genomic richness in antibiotic resistance genes (ARGs), the term "antibiotic resistome" had been introduced, referring to ARGs in their entirety in the context of a prokaryotic organism[131]. Because of the fact that antibiotic agents naturally appear in the environment, ARGs have been found in samples as old as 30,000 years[132].

Aiding in the spread of ARGs among bacteria is their high genomic flexibility, meaning that prokaryotes are able to take up and incorporate foreign DNA *via* various mechanisms such as conjugation (with other prokaryotes), transformation (through uptake of free DNA), and transduction. Transduction can occur *via* bacteriophages – a mechanism, which is also referred to as "horizontal gene transfer"[133] – but also through mobile genes including circular DNA plasmids[134] and highly agile transposons[135]. Transposons are DNA molecules that can easily move between plasmids and genomic DNA and are key players in the spread of ARGs[135], which takes place even without the pressure of antibiotic selection[136]. Recent studies have further unveiled how bacterial communities communicate and induce increased virulence and expression of ARGs in the context of biofilms *via* quorum sensing mechanisms as part of their communal defense strategies[137]. Lastly, epigenetic processes, involving phosphorothioation of bacterial DNA, capping of bacterial mRNA, and methylation of DNA and mRNA, have been linked to adaptive acquisitions of antibiotic resistance[138].

This brief overview of resistance mechanisms and how bacteria acquire and stir resistance-managing genes or gene regulators highlight the importance of ongoing efforts in the fight of overcoming antibiotic resistance, and why it is essential to identify not only new scaffolds of antibacterial agents but also novel modes of action in targeting pathogens. In this regard, it is important to emphasize that bacteria can survive antibacterial treatments even without developing specific resistance mechanisms but instead by exhibiting tolerance, e.g. by slowing down their growth, which can lead to reduced damage caused by the antibiotic[139]. In extreme cases, bacteria can fall into a dormancy state in order to survive antibacterial exposure[139]. Hence, efforts in trying to understand these mechanisms are crucial in the development of novel and potent antibacterial compounds and treatments.

Some examples of successful campaigns trying to overcome resistance mechanisms include structure-based efforts leading to improved existing antibiotics or antibiotic target strategies[140]. In addition, antibiotic adjuvants, which do not have antibacterial functions of their own but are essential parts of co-treatment regiments with antibiotic agents to prevent resistance mechanisms, pose another successful strategy[141], as in the prominent case of the β-lactamase inhibitors[142]. Furthermore, strategies specifically targeting the spread of ARGs through inhibiting conjugation[143] or physically targeting bacteria with nanomaterials[144] are discussed. New targets within the bacterial cell that are currently investigated are, for instance, prokaryotic esterases[145], as well as the secretion pathway unique for prokaryotes[146]. In addition, approaches looking at targets outside of the

pathogen are being developed, which aim to circumvent the selectivity pressure on bacteria, and include host-directed therapies that target virulence or immunological mechanisms[147] as well as quorum sensing pathways to disrupt bacterial communication in the context of biofilms[148].

10.1.3 The Bacterial Cell Wall as a Promising Target for Antibacterial Treatments

Structure of the Bacterial Cell Wall

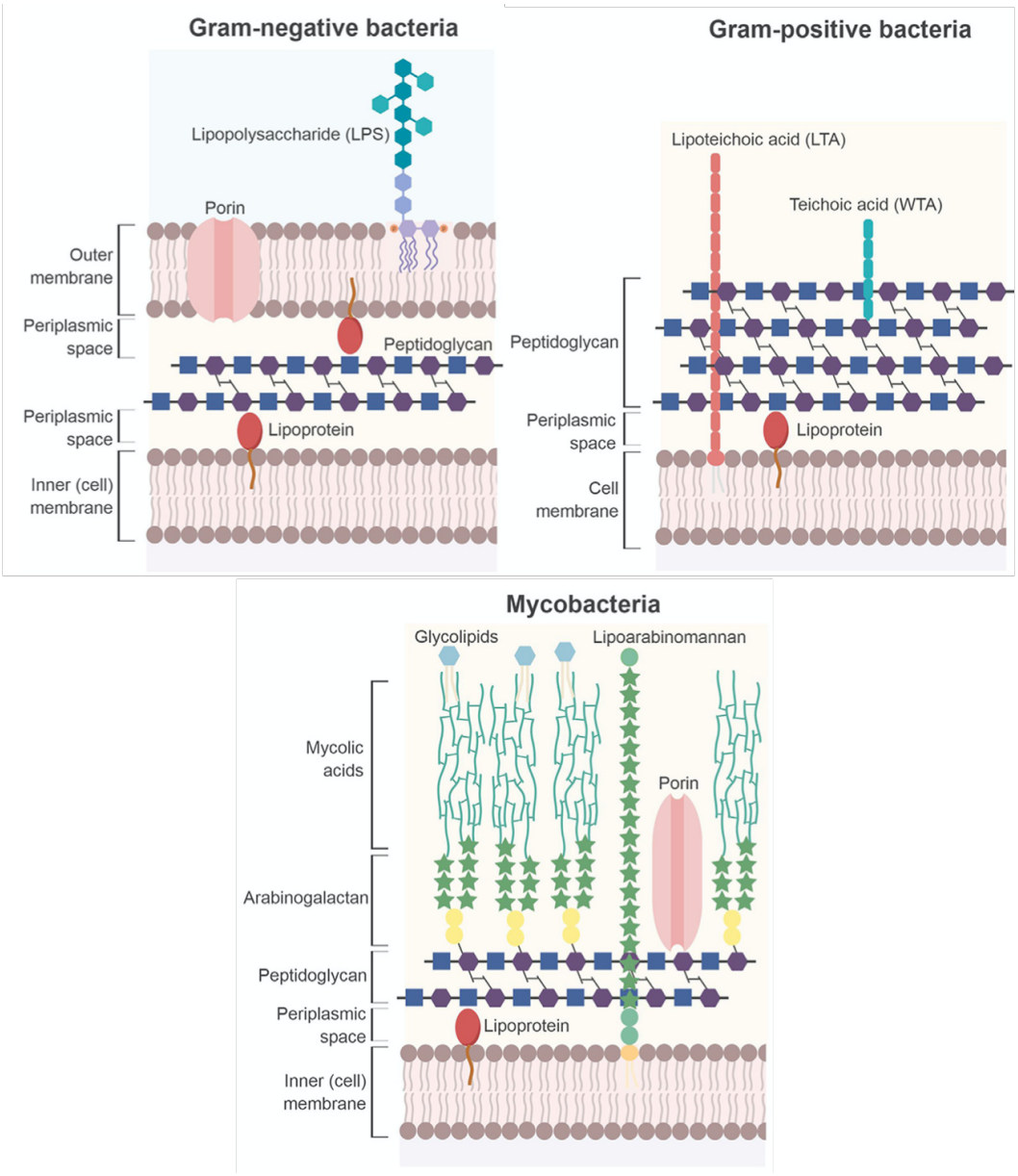


Figure 10.1.12 | General Structures of Bacterial Cell Walls. Schematic representation of the bacterial cell walls for Gram-negative, Gram-positive, and Mycobacteria. Taken from Porfirio *et al.*, *Trends Microbiol.* **2019**, *27* (7), 607-622[149].

The prokaryotic cell wall is a distinct feature of bacteria and therefore absent in eukaryotes such as human cells. Following a conserved structure (Figure 10.1.12), the main function of the bacterial cell wall is to provide cell shape and stability, while also furnishing protection and communication features. In general, anchored within the inner cellular membrane are lipoproteins, which connect to the carbohydrate backbone of peptidoglycan polymers[149]. Peptidoglycan – also known as murein – is a polymer of linear carbohydrate chains, comprised of *N*-acetylglucosamine-*N*-acetylmuramic acid (GlcNAc-MurNAc) β-1,4-connected disaccharides, which are linked to peptide side chains forming an amide bond with the lactyl moiety of MurNAc[149] (Figure 10.1.13).

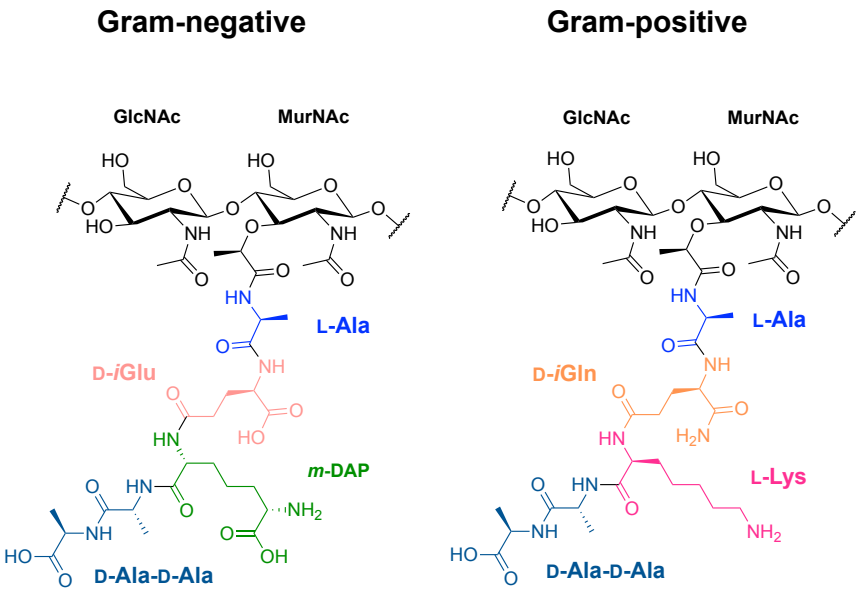


Figure 10.1.13 | Structural Composition of Peptidoglycan Repeating Units. General structure of peptidoglycan repeating units for Gram-negative and Gram-positive bacteria. Adapted from Porfirio *et al.*, *Trends Microbiol.* **2019**, *27* (7), 607-622[149].

In Gram-negative bacteria, the peptidoglycan layer is covered by a second outer membrane, which is connected to the peptidoglycan backbone *via* lipoproteins and contains pore forming proteins or cell surface carbohydrate structures such as lipopolysaccharides (LPSs)[149]. In contrast to that, Gram-positive bacteria are lacking this second lipid bilayer; however, their peptidoglycan network is much thicker than that of Gram-negative bacteria, including lipoteichoic acid and teichoic acid polymers, which are connected to the inner membrane and the peptidoglycan layer, respectively[149]. Another special form of bacterial cell wall structure is found among mycobacteria. Here, similarly to Gram-positive bacteria, they are missing a second outer membrane[149]. However, instead of a thicker peptidoglycan layer, mycobacteria form a mycolylarabinogalactan-peptidoglycan complex comprised of widely branching arabinogalactan polymers attached to the peptidoglycan layer, which carry long mycolic acid chains towards the outer layer of the cell wall[149]. Additionally, long glycolipids, such as lipoarabinomannan, stretch through the cell wall from the inner membrane and, together with the arabinogalactan layer, form a thick and low-penetrable barrier[149].

Despite the overall assembly of the bacterial cell wall, peptidoglycan structure itself – especially within the peptide crosslinkers – differs between Gram-positive and Gram-

negative bacteria (Figure 10.1.13), and can even vary between different bacteria species[149]. For example, while the peptide chains of peptidoglycans commonly consist of five amino acids, Gram-negative cell walls are generally based on L-Ala-D-*i*Glu-*m*-DAP-D-Ala-D-Ala (D-*i*Glu and *m*-DAP stand for D-*iso*-glutamate and *meso*-diaminopimelic acid, respectively), while Gram-positive cell walls have a D-*iso*-glutamine (D-*i*Gln) instead of D-*i*Glu (this is also true for mycobacteria[150]) and an L-Lys instead of *m*-DAP[149].

Except for the 4-position D-Ala, a variety of different amino acids have been reported for other positions in the pentapeptide for different species[150]. For instance, *Mycobacterium leprae* exhibits a Gly in position 1 instead of the L-Ala, and *Streptococcus pyogenes* incorporates an L-5-hydroxylysine in position 3[150]. Moreover, while Gram-negative bacteria usually exhibit direct 3-4-cross-linkage of their peptide side chains between the 3-position *m*-DAP amine and the 4-position D-Ala carboxylic acid (Figure 10.1.14), Gram-positive bacteria commonly exhibit long pentaglycine bridges, connecting their 3-position L-Lys and the 4-position D-Ala[149] (Figure 10.1.15). Notably, the cross-linked peptide chains in mature peptidoglycan lose the final D-Ala[150].

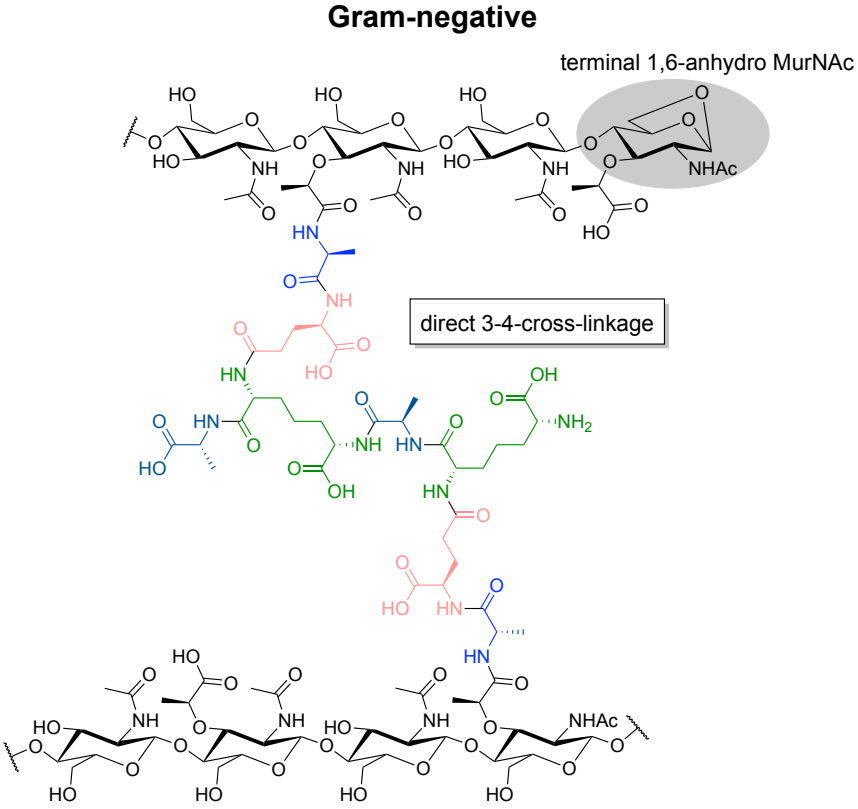


Figure 10.1.14 | Peptidoglycan Crosslinking in Gram-negative Bacteria. General structure of the direct 3-4-crosslinked peptidoglycan polymers in Gram-negative bacteria (same color code as in Figure 10.1.13). Adapted from Porfirio *et al.*, *Trends Microbiol.* **2019**, *27* (7), 607-622[149].

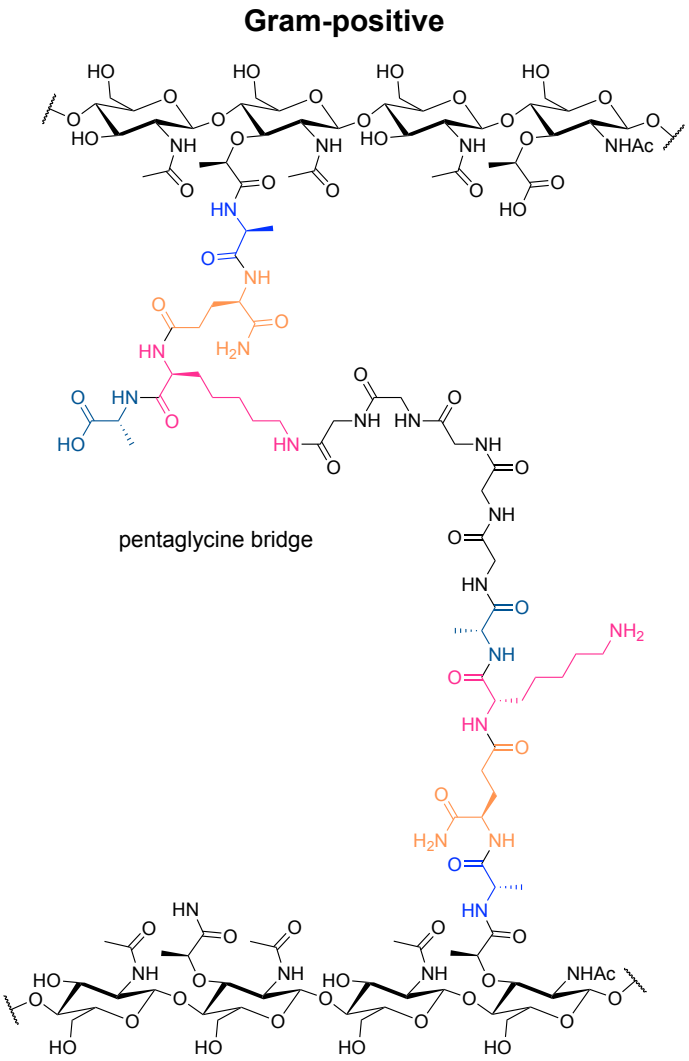


Figure 10.1.15 | Peptidoglycan Crosslinking in Gram-positive Bacteria. General structure of the peptidoglycan polymers in Gram-positive bacteria, which typically exhibit pentaglycine-bridged crosslinking patterns (same color code as in Figure 10.1.13). Adapted from Porfirio *et al.*, *Trends Microbiol.* **2019**, *27* (7), 607-622[149].

Another important difference between the disaccharide polymers of Gram-positive and Gram-negative bacteria is the typical 1,6-anhydro-MurNAc moiety (Figure 10.1.14) found exclusively in peptidoglycan strand endings of Gram-negative bacteria[149]. The incorporation of this capping moiety results in overall shorter strands and ultimately an increase in peptidoglycan elasticity, as compared to the more rigid Gram-positive cell wall structures[151]. Further reported modifications in the peptidoglycan structure include *N*-deacetylations of either one or both carbohydrate moieties, *N*-glycolylation, *O*-acetylation, and cyclic modifications such as muramic δ-lactam[149].

Other general modifications to the peptide amino acids include the cyclic imide-derivatization of the L-Lys in Gram-positive bacteria or modifications of their cross-linking, such as 2-4-cross-linkage *via* ornithine, direct 1-3- or 3-3-cross-linkage, and 3-4-cross-linkage of three pentapeptide units[149]. Some of these diversifications, for example the 5-position D-lactate found in vancomycin-resistant *Enterococci* (VRE), have been specifically associated with antibiotic resistance[150]. Similarly, changes in the crosslinking

pattern, like direct 3-3-cross-linked peptide chains that result in a denser net structure of the peptidoglycan layer, can be linked to β-lactam resistance[151].

In summary, the peptidoglycan structures of bacterial cell walls are highly complex arrangements in which the orchestration of unlinked and cross-linked peptide chains with connections to up to four other pentapeptide units[151] not only determines the stability and fluidity of the bacterial cell wall but also influences important aspects for mechanisms related to antibiotic resistance. Interestingly, peptidoglycans can further function as signaling molecules communicating antibiotic interference of bacterial cell wall biosynthesis of a fellow bacterium to adjacent bacteria, while the human innate immune response interprets these molecules as a sign of a bacterial infection *via* Nod-like receptors[152].

Biosynthesis of the Bacterial Cell Wall and Antibiotic Interference

The biosynthesis of peptidoglycan begins within the bacterial cell and is mainly performed by enzymes of the *mur* family[153] (Figure 10.1.16). In the cytoplasm, uridine diphosphate (UDP)-functionalized GlcNAc is reacted with phosphoenolpyruvate *via* the transferase MurA to yield enolpyruvyl-UDP-GlcNAc, which is subsequently reduced by the reductase MurB, resulting in UDP-MurNAc[153]. The thereby introduced acid function serves as a handle to which the respective pentapeptide chains of the different peptidoglycan units can be attached[153]. MurC-MurF are amide bond ligases that are responsible for the introduction of the pentapeptide to UDP-MurNAc[153]. As described above, these in general are L-Ala *via* MurC, D-Glu (derived from L-Glu by the racemase, MurI) *via* MurD, *m*DAP *via* MurE, and the D-Ala-D-Ala motif, synthesized from L-Ala by alanine racemase (alr) and D-Ala-D-Ala ligase (ddl), *via* MurF[153].

The resulting, still soluble, UDP-MurNAc-pentapeptide intermediate – also referred to as "Park's nucleotide" (first described by James T. Park[154]) – is then loaded onto the lipid carrier undecaprenyl phosphate (C_{55}-P) by the translocase MraY upon cleavage of a UMP unit at the inner membrane[140], which will be discussed in more detail below. Structurally, C_{55}-P is a polymerization product of isoprene units[155]. In bacteria, this is achieved through the combination of short, all-*trans* oligoprene building blocks, most commonly catalyzed by cytosolic undecaprenyl diphosphate synthase (UppS) under formation of *cis*-configured bonds[155,156] (note: describing the conformation of C=C-double bonds as "*cis*" or "*trans*" is obsolete; however, it is used here as it is still commonly seen in the context of prenyltransferases, thus is in accordance with current conventions in the respective literature). Typically, this process yields chain lengths of up to 55 carbons, although shorter lengths of C_{50} or C_{45} have also been reported[155]. Subsequent dephosphorylation of the pyrophosphate lipid carrier, which can be performed by a variety of different enzymes, results in the monophosphate intermediate needed for the following reactions in the bacterial cell wall biosynthesis[155] (Figure 10.1.16).

To the product of the linkage between C_{55}-P and Park's nucleotide – the so-called "lipid I", the glycosyltransferase MurG introduces another UDP-GlcNAc moiety, forming the disaccharide unit of the peptidoglycan molecule, thereby transforming lipid I to lipid II[157,140]. In case of Gram-positive bacteria, several acetyltransferases, such as FemX/A/B, introduce the pentaglycine bridge at this point in the biosynthesis[140]. To transport the completed peptidoglycan unit across the bacterial membrane, numerous potential enzymes have been discussed. For example, MurJ[158] and the shape, elongation, division, and

sporulation (SEDS) family member FtsW[159] are currently thought to be key flippases in this process, while other SEDS family members, such as RodA and SpoVE, are believed to perform flippase activity during elongation and sporulation, respectively[159].

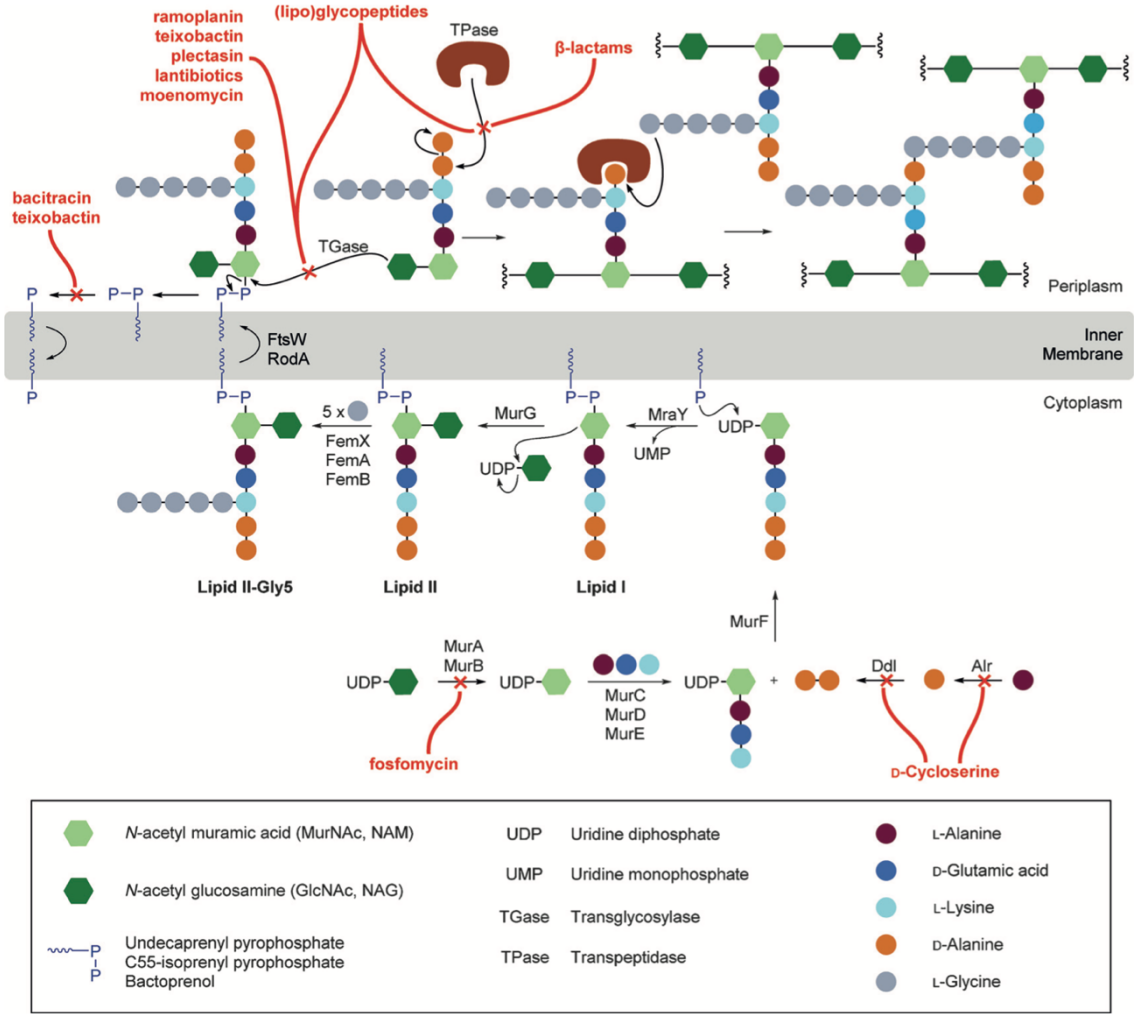


Figure 10.1.16 | Peptidoglycan Biosynthesis. The peptidoglycan biosynthesis begins in the cytoplasm, where peptidoglycan units are built by a variety of Mur enzymes. Completed peptidoglycan molecules are transported to the periplasm by flippases, where they are polymerized and crosslinked through a variety of different enzymes. Established inhibitors of pathways involved in this biosynthesis are highlighted in red. Taken from Chellat *et al.*, *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (23), 6600-6626[140].

Ultimately, class A penicillin-binding protein (aPBP) and SEDS family members, such as RodA, function as transpeptidases to build the peptidoglycan layer[160]. While aPBPs can perform the polymerization of the peptidoglycan units as well as the cross-linking of the peptides, SEDSs only function as glycosyltransferases and need to work together with class B penicillin-binding proteins (bPBPs) to build the murein layer[161]. As transpeptidases, PBPs recognize the D-Ala-D-Ala motif in peptidoglycan molecules and form amide bonds with amino acids from other units, while removing the 5-position D-Ala of the pentapeptide[161]. Other transpeptidases are specialized for L,D-amino acid connections and form connections of peptidoglycan with membrane-based lipoproteins (YbiS, ErfK, and

Ycf in *Escherichia coli*), or *m*DAP-linkages through direct DAP-DAP or 3-3-cross-linking (YnhG and YcbB in *Escherichia coli*)[162].

At the end of this process, the lipid carrier C_{55}-P stays in the cellular membrane as a pyrophosphate byproduct of the described polymerization reactions, and can be recycled to the respective monophosphate for another cycle of lipid II formation and transportation through the membrane *via* dephosphorylation by phosphatases including the membrane protein undecaprenyl pyrophosphate phosphatase (UppP), also known as BacA[163]. Figure 10.1.16 gives an overview of peptidoglycan biosynthesis, discussed in this chapter, including some established antibiotic targets with respective examples[140].

The bacterial cell wall is a crucial feature to bacterial survival and is formed in a multistep biosynthetic pathway, providing various targeting strategies for antibiotic interference. Several known antibiotics take advantage of this fact[140]. For example, inhibition of the (soluble) peptidoglycan unit biosynthesis is one successful strategy for antibacterial compounds, including covalent MurA inhibitors such as the clinically established fosfomycin[164] and other natural products such as the terreic acid, which has been found to covalently label Cys115 of this transferase in investigational studies[165]. To date, several inhibitors for Mur enzymes have been reported (MurA-G), with most inhibitors showing cross-activity with several Mur proteins[164]. For instance, phosphorylated hydroxyethyl amine derivatives exhibit inhibitory activity of MurC-MurF[166], and phenyl thiazolyl ureas and carbamates were reported as inhibitors of MurA and MurB[167]. In contrast to these multi-targeting antibiotics, selective MraY inhibitors, especially nucleoside-based natural products, have been discovered[164], and will be discussed in more detail in following chapters. Other antibiotic agents target precursors of the peptidoglycan building blocks, as in the case of the recently reported benzamide-based scaffolds, which have been shown to inhibit the UDP-functionalization of UDP-GlcNAc by targeting the uridyltransferase GlmU[168], while the clinically approved D-cycloserine prevents D-Ala-D-Ala formation by inhibiting ddl[164].

Since all these targets require the antibiotics to penetrate the bacterial cell in order to reach their targets, mechanisms involved in the polymerization and cross-linking of the peptidoglycan units have been attractive targets for antibiotic compounds to circumvent the hurdle of cellular uptake. For instance, clinically used vancomycin and non-approved lantibiotics bind lipid II thereby preventing polymerization of the disaccharide, while the investigational moenomycins are reported to directly target the glycosyltransferases[169]. Similarly, PBPs are covalently bound by clinically established β-lactams, which thereby inhibit PBP transpeptidase activity[164].

More recently, antibacterial agents targeting Gram-specific enzymes of the cell wall synthesis pathways have been reported in the literature. For instance, the NagZ glycosylase is a key player in the recycling process of peptidoglycan in Gram-negative bacteria, due to their unique 1,6-anhydromuramyl-based terminal moieties, and has been found to be inhibited by iminosaccharide analogs[170]. In Gram-positive bacteria, L-rhamnose biosynthesis is targeted by thymidylyltransferase inhibitors based on nucleotide analogs[171], while RNA-based molecules inhibit the unique pentaglycine-introducing FemX acetyltransferases[164].

Finally, recent studies have shown that bacterial features, such as biofilm function, can be influenced by cell wall biosynthesis targeting agents. For example, D-Leu is able to

interrupt anchoring mechanisms of matrix carbohydrate polymers for biofilm formation, having the advantage of avoiding direct cell killing effects[172].

MraY as a Target of Nucleoside-based Antibiotics

As mentioned above, MraY poses an attractive target for antibiotics. The murein synthetic gene cluster a-Y – short, *mraY* – has been initially described for *Escherichia coli* by Ikeda and colleagues in the early 1990s[173]. The group reported that the gene encodes for the protein responsible for attaching the MurNAc-pentapeptide peptidoglycan precursor onto the membrane-based undecaprenyl phosphate lipid carrier as part of cell wall biosynthesis[173] (Figure 10.1.16). The group further found that the hydrophobic clusters within the structure of MraY suggested that it is a membrane-based protein, which is in contrast to other enzymes encoded within this gene region, e.g. the cytosolic *mur* gene-derived proteins[173].

While Ikeda *et al.* already showed that tunicamycin inhibits MraY enzymatic activity in a dose-dependent manner (Figure 10.1.17 A)[173], it was Boyle and Donachie who validated an essential function of *mraY* for bacterial cell growth (Figure 10.1.17 B) as *mraY*-null mutant *Escherichia coli* would exhibit a distinct cell lysis phenotype[174]. In the late 1990s, Bouhss *et al.* elucidated important structural features of the MraY protein by comparing the sequence and *in silico* topological models of MraY from 25 different bacterial strains[175]. Their work supported Ikeda's initial model of MraY being an integral membrane protein[173], proposing a structure with ten transmembrane helices, five hydrophilic loops I-V facing the cytoplasm, and six periplasmic domains, including the N- and C-termini[175]. Notably, while the transmembrane and periplasmic structures of MraY show variability among different strains, their overall structural features are comparable, whereas the cytoplasmic loops are highly conserved among the orthologs[175]. Taken together, this data suggests that the hydrophilic loops facing the cytoplasm are responsible for substrate binding[175]. Moreover, it was shown that the MraY active sites of Gram-positive and Gram-negative strains bind to the same substrates, indicating that the binding mode is pentapeptide-independent[175]. Later, Al-Dabbagh *et al.* performed single amino acid mutations of 19 conserved amino acids within the five cytosol-facing loops (Figure 10.1.17 C) and uncovered that 14 of these conserved amino acids – distributed over all five hydrophilic segments – were essential for MraY function, providing evidence of a mutual interplay for the formation of the enzyme's binding site[176].

The long-time unknown structure and therefore binding mode of the MraY substrate historically prevented medicinal chemistry campaigns from rational development of MraY inhibitors. In 2013, Chung and co-workers reported the first crystal structure of MraY isolated from *Aquifex aeolicus* strains ($MraY_{AA}$), which validated studies carried out by Al-Dabbagh and Bouhss[177]. The crystal structure shows ten transmembrane helices, with the ninth exhibiting a distinct bend (Figures 10.1.18 A and B), as well as five conserved loops – referred to as loops A-E, facing the cytosol[177]. Moreover, the structure reveals two MraY molecules dimerizing to form a hydrophobic pore in their interface[177] (Figure 10.1.18 B).

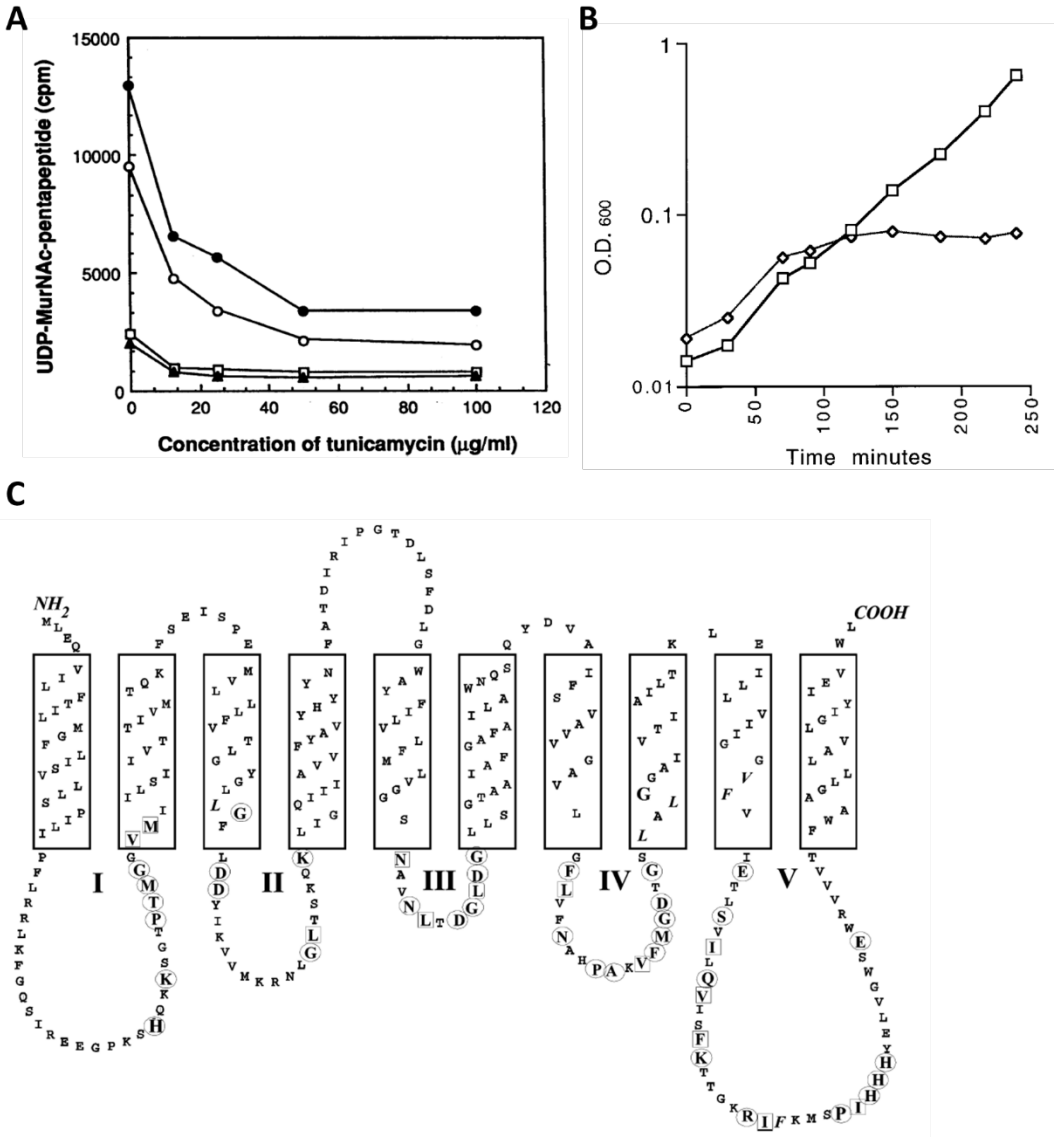


Figure 10.1.17 | MraY – An Integral Membrane Protein Essential for Bacterial Cell Growth. **A.** Tunicamycin inhibits MraY (from *E. coli* mutants containing different plasmids either including *mraY* (circles) or not (squares)) catalytic activity in a dose-dependent manner. Taken from Ikeda *et al*., *J. Bacteriol.* **1991**, *173* (3), 1021-1026[173]. **B.** Bacterial cell growth for mutant *E. coli* strains with suppressed (small diamonds) or induced (big squares) *mraY* expression. Taken from Boyle *et al*., *J. Bacteriol.* **1998**, *180* (23), 6429-6432[174]. **C.** Proposed topology of MraY with conserved regions marked as identical (circle) or similar (square). Taken from Al-Dabbagh *et al*., *Biochemistry* **2008**, *47* (34), 8919-8928[176].

Chung *et al*. identified Asp117/118/265 and His324 as crucial amino acids in the catalytic mechanism of MraY *via* mutation-based experiments[177]. Furthermore, it was reported that Asp265 is potentially involved in the complexation of a Mg^{2+}-ion, and that Asp117 may bind the C_{55}-P phosphate[177] (Figure 10.1.18 C). An adjacent long hydrophobic pocket supports this hypothesis, since it could serve as a potential binding site for the C_{55}-P polyisoprenyl chain[177] (Figure 10.1.18 D). However, it was not until the publication of a co-crystal structure of $MraY_{AA}$ with the inhibitor muraymycin D2 (Figure 10.1.19) that the binding mode of small molecules to the cytosolic loops was vaguely understood and therefore useful for drug development purposes[178].

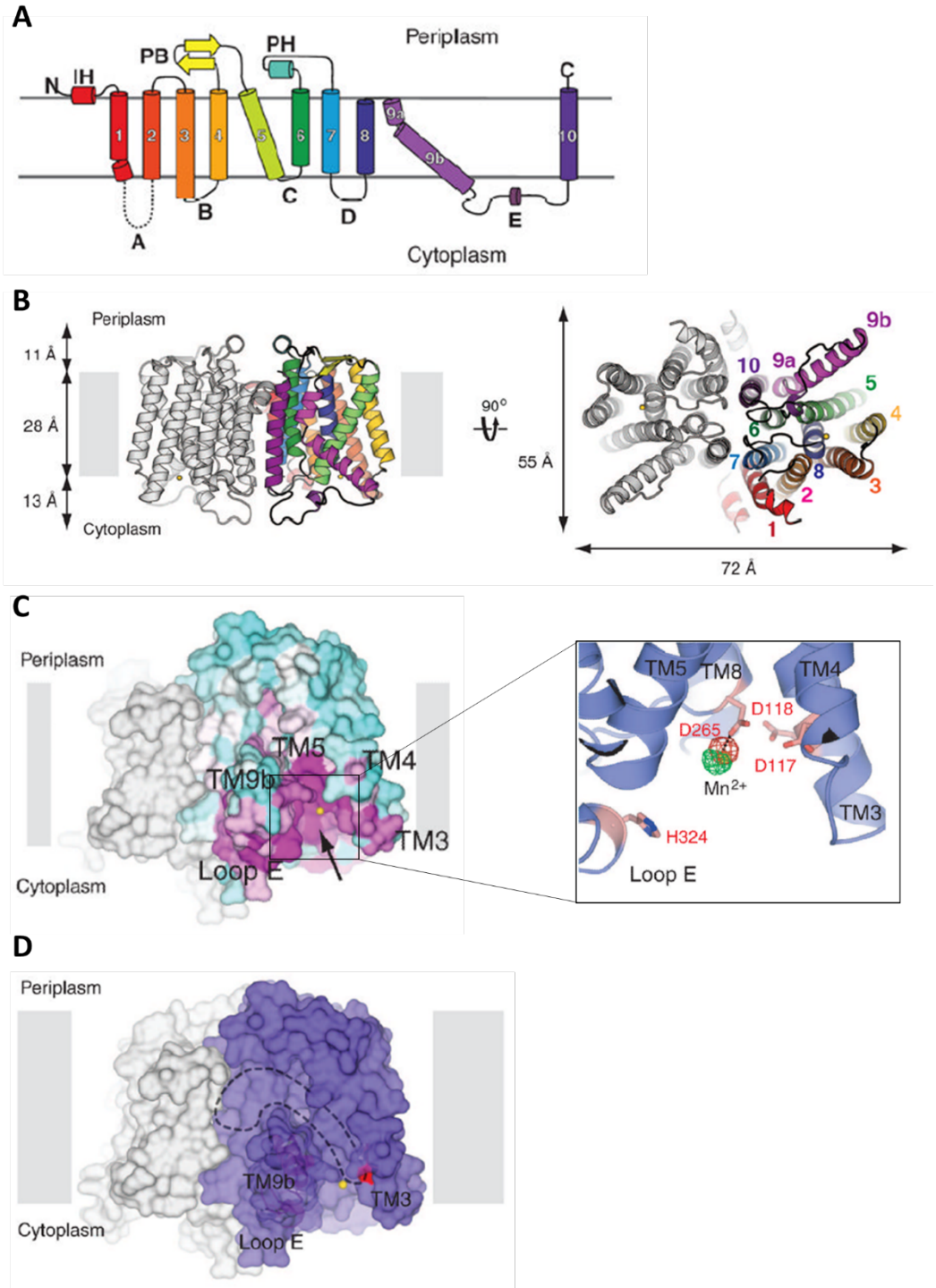


Figure 10.1.18 | Crystal Structure of MraY$_{AA}$[a]. A. Principle topology of MraY$_{AA}$. **B.** Crystal structure of MraY$_{AA}$ dimer (one homolog in gray, the other shown in the same color-code as in A). **C.** Sequence conservation of MraY$_{AA}$ with colors ranging from magenta (absolutely conserved) to cyan (least conserved), including a zoom-in on the proposed substrate binding site as well as important amino acids (highlighted in red). **D.** A hydrophobic pocket adjacent to the nucleotide binding site is depicted in dashed lines. Mg^{2+} and Asp117 are highlighted as yellow and red shapes, respectively.

[a] Figures adapted from Chung *et al.*, *Science* **2013**, *341* (6149), 1012-1016[177].

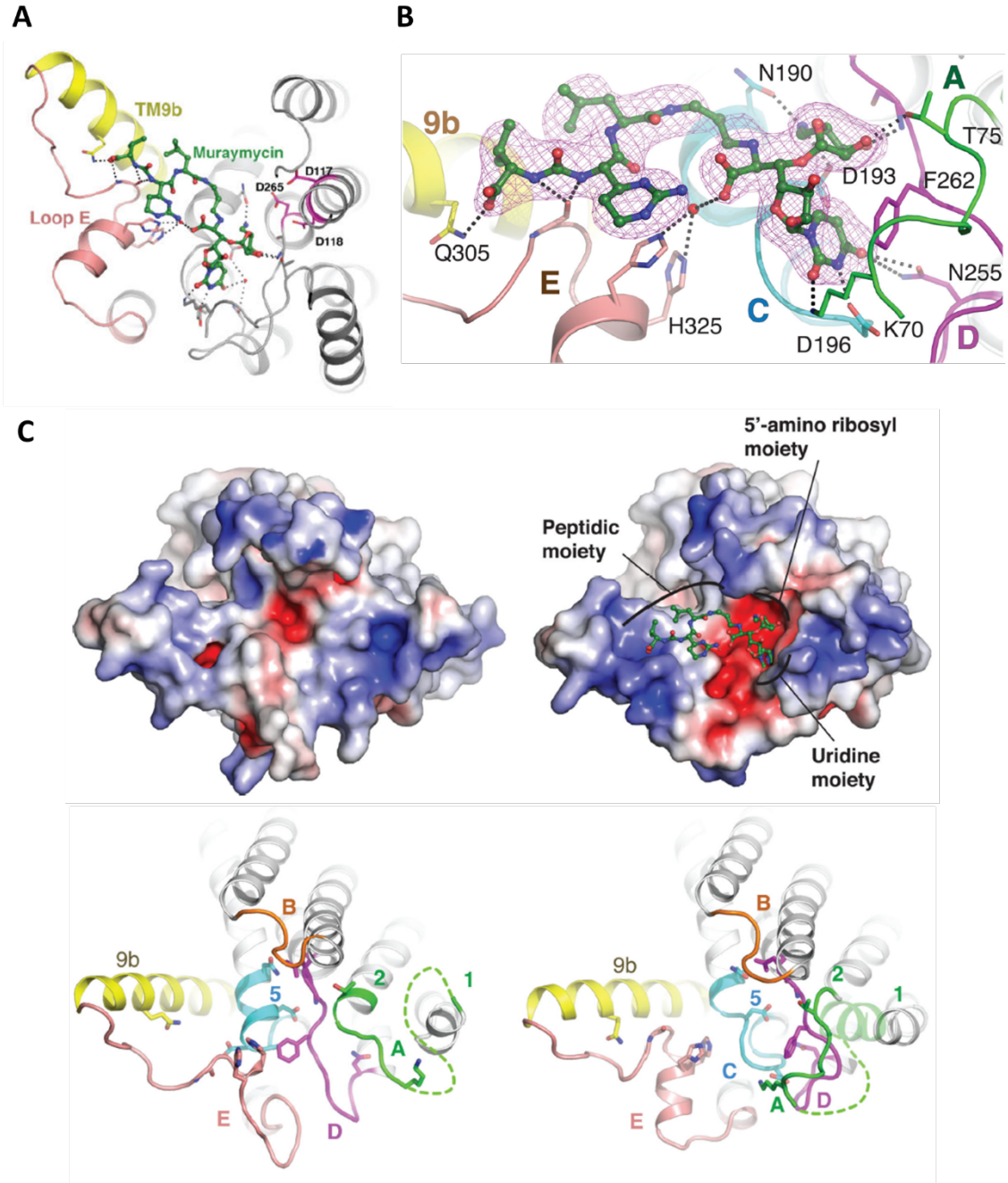


Figure 10.1.19 | Co-crystal Structure of MraY$_{AA}$ in Complex with Muraymycin D2[a]. A. Muraymycin D2 bound to MraY$_{AA}$. View from cytosol. **B.** Close-up on muraymycin D2 interactions with the MraY$_{AA}$ binding site. Important amino acids are highlighted (labeled). **C.** Comparison of the structure of apoMraY$_{AA}$ (top and bottom, left) and the co-crystal structure of MraY$_{AA}$ in complex with muraymycin D2 (top and bottom, right).
[a] Adapted from Chung *et al.*, *Nature* **2016**, *533* (7604), 557-560 [178].

Interestingly, it was shown that the conserved loops undergo impressive conformational changes upon ligand binding, especially within the E loop that interacts with the peptide chain of the inhibitor[178] (Figure 10.1.19 C). This flexibility might contribute to the ability of MraY to bind peptidoglycan precursors carrying different pentapeptide moieties, as well as to the variety of inhibitor scaffolds that are targeting MraY[178]. Other co-crystal structures with different nucleoside-based antibiotics that have been reported since then

support this hypothesis and further add to our understanding of MraY function and inhibition, and aid in ongoing efforts to develop potential antibacterial agents[179,180]. For example, structures of MraY from *Clostridium bolteae* strains crystallized in complex with tunicamycin[179], as well as structures of $MraY_{AA}$ bound to nucleoside-based, naturally-occurring antibiotics (caprazamycin, capuramycin, and mureidomycin)[180] reveal the chemical logic for binding of these distinct natural products, highlighting the characteristic flexibility of the nucleoside binding site through the interplay of the intracellular loops (Figure 10.1.20).

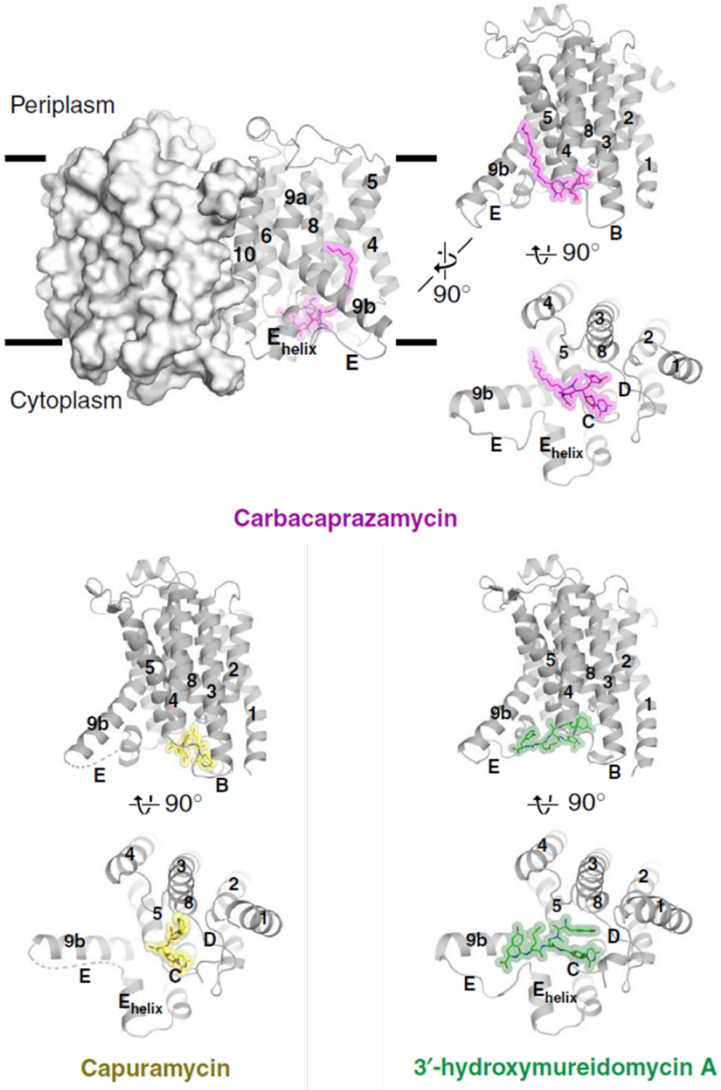


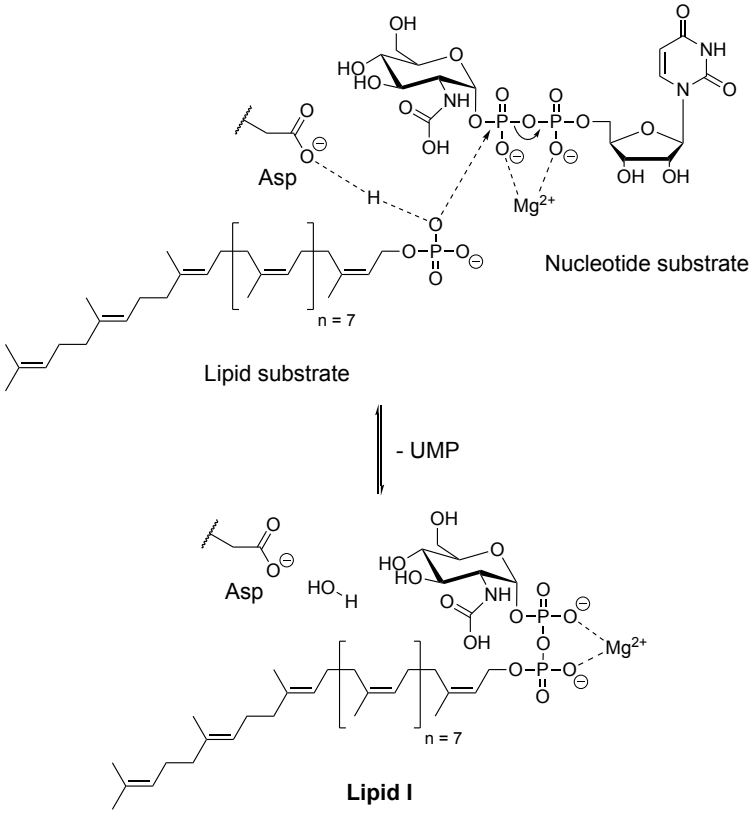
Figure 10.1.20 | Comparison of Co-crystal Structures of $MraY_{AA}$. Co-crystal structures of $MraY_{AA}$ in complex with a caprazmycin (magenta), capuramycin (yellow), and a mureidomycin (green). Adapted from Mashalidis *et al.*, *Nat. Commun.* **2019**, *10* (1), 2917[180].

Finally, kinetic studies using a fluorescent substrate probe further deepened our understanding of the catalytic mechanism of MraY[181]. The results of these studies suggest that the binding of both C_{55}-P and Park's nucleotide is crucial for MraY activity[181]. It is further believed that His289 facilitates the nucleophilic attack of the C_{55}-P phosphate either through direct binding of the phosphate or activation by deprotonation *via* its activity as a

base, yielding bond formation with the MurNAc-pentapeptide substrate under cleavage of UMP in a reversible S_N2-type reaction[181]. However, it remained unclear whether MraY forms a covalent complex with MurNAc-pentapeptide first, resulting in a two-step mechanism, or whether C_{55}-P gets deprotonated to directly attack the nucleotide, which would be a one-step catalytic mechanism. Studies that employed radiolabeled MraY substrates support the one-step mechanism in which the C_{55}-P phosphate is deprotonated and subsequently reacts with the Park's nucleotide[182]. Interestingly, these studies also suggest that conserved aspartate residues within the active site are responsible for the phosphate deprotonation rather than His289[182] (Scheme 10.1.3).

Based on these catalytical studies and due to mutual lipid substrates, MraY is considered part of the polyprenyl-phosphate *N*-acetylhexosamine 1-phosphate transferase (P2HPT) superfamily, whose enzymes are all sharing similar catalytic mechanisms[182]. Notably, some prokaryotic enzymes, such as WecA, TagO, WbcO, WbpL and RpG (exhibiting substrate specificity for their carbohydrate substrates), but also eukaryotic enzymes, such as glutamate pyruvate transaminase 1 (GPT), are part of this family[182].

Scheme 10.1.3 | Proposed One-Step Mechanism for the Catalytic Activity of MraY[a]



[a] Adapted from Al-Dabbagh *et al*., *Biochimie* **2016**, *127*, 249-257[182].

Several efforts in the development of MraY inhibitors as novel antibiotic agents have been described. To date, six different classes of nucleoside antibiotics – most of which were isolated from *Streptomyces* strains – have been identified to have MraY inhibitory activity, namely, caprazamycins, capuramycins, mureidomycins, muraymycins, tunicamycins[180,183], and sphaerimicins[184] (Figures 10.1.21 and 10.1.22 B). Even though

these classes all target a distinct nucleoside binding site due to their conserved uridine motif, their unique structures exhibit otherwise distinguishable binding modes as a result of the overall flexible binding site of MraY[180] (Figure 10.1.20). Notably, while these nucleoside antibiotics are generally selective for the prokaryotic MraY enzyme, tunicamycins also bind the related eukaryotic GPT protein, leading to off-target toxicity[180,183]. Interestingly, other nucleoside-based natural products, including nikkomycins and polyoxins, have been shown to inhibit the fungal chitin biosynthesis – a crucial building block of the cell wall found in fungi[183].

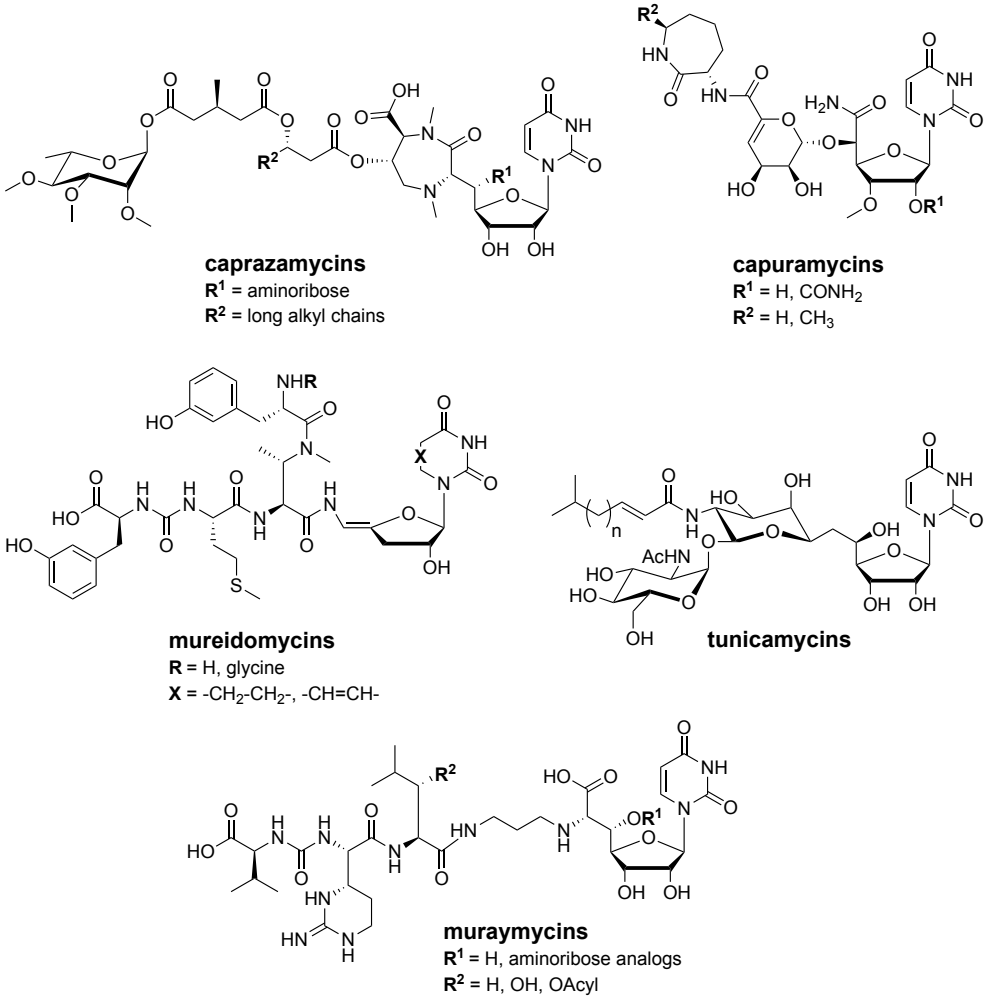


Figure 10.1.21 | Structures of Natural Product-based MraY Inhibitors. Overview of well-established, nucleoside-based MraY inhibitors. Shown are the general scaffolds of caprazamycins, capuramycins, mureidomycins, muraymycins, and tunicamycins. The structure of sphaerimicin – another structurally related MraY inhibitor – is shown in Figure 10.1.22 B.

Another reported MraY inhibitor with a completely different binding mode is the bacteriophage-based lysis protein E, which is proposed to bind to helix 9 *via* specific π-π-interactions with amino acids close to the periplasmic side of MraY[185] (Figure 10.1.22 A). Some studies further suggest that certain nucleoside antibiotics are potentially also binding to this site as part of their mode of action[186]. Interestingly, strategies to develop MraY inhibitors from closer substrate mimetics yielded disappointing results[187], while genetic

approaches and library screenings led to the discovery of novel nucleoside-based MraY inhibitor scaffolds (sphaerimicin)[184], and the isoquinoline alkaloid-related michellamine B[188], respectively (Figure 10.1.22 B). Recent studies have further added MraY-specific monoclonal antibodies as potential novel antibacterial therapeutic strategies, showing good antibacterial activity in mouse models[189]. As the muraymycin class of antibiotics is part of this work, they will be discussed in more detail in the following chapter.

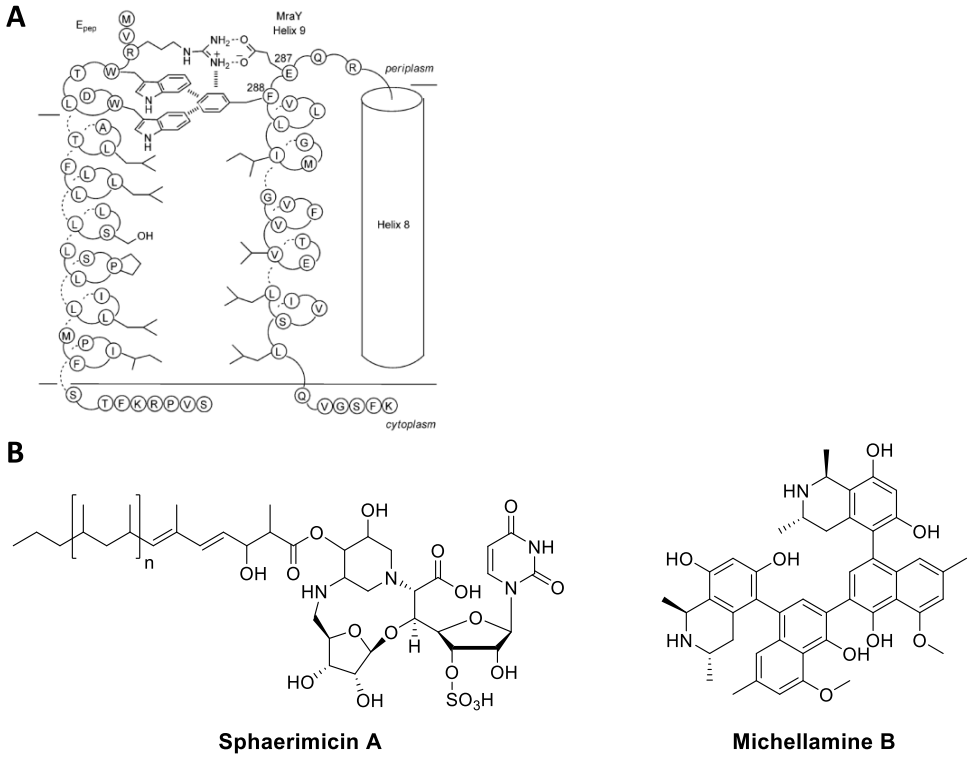


Figure 10.1.22 | Additional Structures of Reported MraY Inhibitors. A. Proposed binding mechanism of MraY-inhibiting lysis protein E of bacteriophage ϕ X174. Figure taken from Rodolis et al., *ChemBioChem.* **2014**, *15* (9), 1300-1308[185]. **B.** Recently reported MraY inhibitors, sphaerimicin A (nucleoside antibiotic) and michellamine B (alkaloid).

10.1.4 Muraymycin-based Nucleoside Antibiotics

While nucleoside antibiotics can generally affect a variety of different processes within a bacteria, such as pseudouridimycin-based inhibition of RNA polymerases[190] or albomycin-based inhibition of seryl-tRNA synthases[191], uridine-derived natural products are an important class of potent and selective MraY inhibitors, which disrupt peptidoglycan biosynthesis. The uridine motif shared by MraY-inhibiting nucleoside antibiotics is also found in the natural MraY substrate UMP-MurNAc-pentapeptide[180] (Figures 10.1.21 and 10.1.22 B). As already described, these nucleoside antibiotics bind to the intracellular substrate binding site of the membrane-bound translocase[180] (Figure 10.1.20).

There are several subclasses of uridine-derived MraY-inhibiting nucleoside antibiotics, including (Figure 10.1.21):

(1) *Tunicamycins* (including tunicamycins, streptovirudins, and corynetoxins), which have a distinct tunicamine carbohydrate structure, including the sugar moiety of the uridine nucleoside, as well as a GlcNAc and a fatty acid group[192].

(2) *Mureidomycins* (such as mureidomycins, pacidamycins, napsamycins, and sansanmycins), which contain a unique 3'-deoxyuridine core, which is connected to a short peptide motif *via* an unsaturated 4',5'-enamide linkage[192]. Notably, reduced analogs retain MraY inhibitory activity[192]. The peptide commonly includes an *N*-methylated 2,3-diaminobutyric acid motif with an aromatic amino acid – mostly *meta*-tyrosine – in the 3-position and a methionine in the 2-position, followed by another *meta*-tyrosine connected *via* a urea linker[192].

(3) Liponucleoside antibiotics, such as the *liposidomycin* group (including liposidomycins and caprazamycins), commonly contain the typical uridine base linked to an aminoribofuranoside carbohydrate moiety in the 5'-position, followed by a distinct diazepanone ring system[192]. Attached to this aliphatic heterocycle are long fatty acids carrying a 3-methylglutaryl group, which, in case of the caprazamycins, includes another L-rhamnose sugar substituent[192].

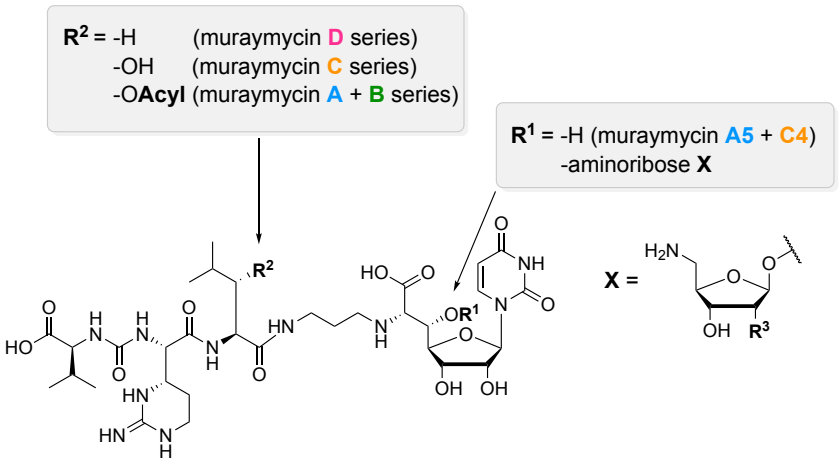
(4) Finally, the *capuramycins* (comprised of capuramycin and A-500359A (and analogs)), which have a uronic acid moiety attached to the 5'-position of the uridine core, which is decorated with a carbamate and methyl group at the 2'- and 3'-positions, respectively[192]. The uronic acid forms an amide linkage with an aminocaprolactam ring or is replaced by a simple methyl ester in some A-500359 analogs[192].

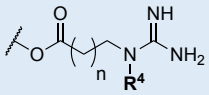
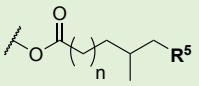
This chapter will focus on muraymycin antibiotics, as this is the main focus of this project as part of this thesis. In 2002, McDonald and co-workers first reported the isolation and MraY inhibitory activity of 19 muraymycin analogs extracted from *Streptomyces* species[193]. Based on structural features, muraymycin antibiotics can be divided into four groups, A-D [193] (Table 10.1.1).

Generally, muraymycins have a 5'-hydroxy group at the uridine nucleoside core, which is functionalized with an aminoribofuranoside moiety; this is putatively cleaved off during purification processes, leading to a free hydroxy group for analogs A5 and C4[192,193]. 5'-Deoxy analogs, which can be more easily accessed synthetically, were also shown to be active MraY inhibitors[194]. A diaminopropyl linker connects the nucleoside core to a short peptide sequence, which consist of the characteristic amino acids L-(hydroxy)leucine,

epicapreomycidine (a cyclic guanidino amino acid), and an L-valine, which is linked by a urea motif[192].

Table 10.1.1 | Structures of Naturally Occurring Muraymycins[a]



Muraymycin		R^3	Acyl	n	$R^{4/5}$
A	**1**	$-OCH_3$		11	-OH
	2	$-OCH_3$		9	-OH
	3	$-OCH_3$		11	-H
	4	-OH		11	-OH
	5	N/A		11	-OH
B	**1**	$-OCH_3$		5	$-CH_3$
	2	$-OCH_3$		5	-H
	3	$-OCH_3$		3	$-CH_3$
	4	$-OCH_3$		4	-H
	5	-OH		4	-H
	6	$-OCH_3$		3	-H
	7	-OH		3	-H
	8	$-OCH_3$		11	-H
	9	$-OCH_3$		9	-H
C	**1**	$-OCH_3$	N/A	-	-
	2	-OH		-	-
	3	-H		-	-
	4	N/A		-	-
D	**1**	$-OCH_3$	N/A	-	-
	2	-OH		-	-
	3	-H		-	-

[a] N/A, not applicable.

The (hydroxy)leucine functionalization determines the muraymycin classification into sub-Groups A-D (Table 10.1.1). Compounds of Group A, which most potently inhibit MraY, contain guanidinylated or hydroxyguanidinylated fatty acids at this position, while Group

B compounds have unfunctionalized alkyl fatty acid chains[193]. Group C and D muraymycins do not carry any lipophilic side chain; instead, Group C compounds are characterized by a free hydroxy group at the central leucine moiety, while the group D compounds consist of an unfunctionalized leucine[193]. Recent studies have added additional members to the B- and C-Groups of muraymycins, raising the number of isolated muraymycin analogs to a total of 21[195].

In 2011, Cheng *et al.* identified the gene cluster responsible for muraymycin biosynthesis in the *Streptomyces* strain NRRL 30471, which allowed them to propose the first biosynthetic route for muraymycin antibiotics[196]. While the exact mechanism is still being investigated, several Mur proteins are reported to orchestrate muraymycin biosynthesis starting from uridine monophosphate (UMP) units[196,197]. Furthermore, it was found that the biosynthetic buildup of the uridine-based muraymycin core was found to be comparable to the biosynthesis of other nucleoside antibiotic core structures, including those of liposidomycins, caprazamycins, and capuramycins[198]. Notably, in the same gene cluster, Cui and colleagues have identified Mur28 and Mur29, which phosphorylate or adenylate the aminoribosylated uridine core, respectively, yielding compounds with no significant MraY inhibitory activity[199]. This mechanism is essential to protect nucleoside antibiotics-producing bacteria from interfering in their own cell wall biosynthesis[199].

Antibacterial Activity of Nucleoside Antibiotics

As already mentioned, the A-Group among the muraymycins – especially muraymycin A1 – exhibit the most potent antibacterial effects of these nucleoside antibiotics[193]. Muraymycin A1 shows strong potency against Gram-positive bacteria, including *Staphylococci* and *Enterococci* with minimum inhibitory concentration (MIC) values of 2-16 and 16->64 µg/mL, respectively[200]. In addition, muraymycin A1 also exhibits potency against certain Gram-negative strains with MIC values varying from 8->64 µg/mL[200]. Notably, cellular uptake of the drug was likely an issue in strains that muraymycin exhibited little to no potency in, as the MIC was under 0.03 µg/mL in mutant *E. coli* with an improved membrane permeability[200].

In comparison, other nucleoside-based MraY inhibitors have shown different profiles. For example, tunicamycins are mainly active against *Bacillus* strains (MIC = 0.1-20 µg/mL; Gram-positive). The structurally related streptovirudins and corynetoxins show a similar profile, including toxicity effects due to mutual off-targets in mammals[200]. Liposidomycins were found to have cellular uptake issues, as their biochemical MraY activity differs significantly from their cellular activity[200]. Here, extensive SAR studies have shown that only analogs with long fatty acid chains are active against, for example, *S. aureus* (MIC = 2.5 µg/mL; Gram-positive)[200].

In contrast, mureidomycins are active against the Gram-negative *Pseudomonas aeruginosa* including β-lactam resistant strains, with MIC values of 1.5-12.5 and 0.1-1.56 µg/mL, respectively[200]. Interestingly, some mureidomycin analogs were even found to be potent against *M. tuberculosis* with a MIC_{90} of 12.5 µg/mL[200]. Other examples for nucleoside antibiotics active on Gram-negative bacteria are the mureidomycin-related pacidamycins, which are potent against *P. aeruginosa* (MIC = 4-16 µg/mL), napsamycins, being active against *P. aeruginosa* and *M. tuberculosis* (MIC = 8-20 µg/mL), and sansanmycins, with MIC values of 2 and 8 µg/mL for *M. tuberculosis* and *P. aeruginosa*, respectively[200]. Caprazamycins are highly potent against mycobacteria with a MIC value

of 3.13 µg/mL against *M. tuberculosis*[200]. Notably, muraymycin D1 was also found to inhibit *M. tuberculosis*, with a MIC_{50} of 1.56 µg/mL[201].

Interestingly, some synthetic analogs of caprazamycins exhibit big shifts in their activity profile upon minor changes in their scaffold, leading to active compounds against *S. aureus*, *E. faecalis*, and/or lost activity on *M. tuberculosis*[200]. Antibiotics of the class among the A-series, such as A-94964, show similar profiles[200]. Capuramycins and analogs, as well as relatives from the A-series (e.g. A-500359A), exhibit antibacterial activity against *M. smegmatis* (3 µg/mL)[200]. Here, more lipophilic analogs, carrying additional fatty acids, yielded in an improved activity profile towards Gram-negative and -positive species[200], potentially indicating better cellular uptake. A summary of the data presented in this chapter is given in Table 10.1.2.

Table 10.1.2 | Antibacterial Activity of Selected Nucleoside Antibiotics[a]

Nucleoside Antibiotic	MIC				
	Gram pos.		Gram neg.		*M. tuberculosis*
	[µg/mL]	strain	[µg/mL]	strain	[µg/mL]
muraymycin A1	2 - 16 8 - >64	*Staphylococci* *Enterococci*	8 - >64 <0.03	various *E. coli* (mutant)	N/A
muraymycin D1	N/A	N/A	N/A	N/A	1.56
tunicamycins	0.1 - 20	*Bacillus*	N/A	N/A	N/A
liposidomycins	2.5	*S. aureus*	N/A	N/A	N/A
mureidomycins	N/A	N/A	1.5 - 12.5 (0.1 -1.56)	*P. aeruginosa* (resistant strain)	12.5 (MIC_{90})
pacidamycins	N/A	N/A	4 - 16	*P. aeruginosa*	N/A
napsamycins	N/A	N/A	8 - 20	*P. aeruginosa*	8 - 20
sansanmycin	N/A	N/A	8	*P. aeruginosa*	2
caprazamycins	N/A	N/A	N/A	N/A	3.13
capuramycins	N/A	N/A	N/A	N/A	3 (*M. smegmatis*)

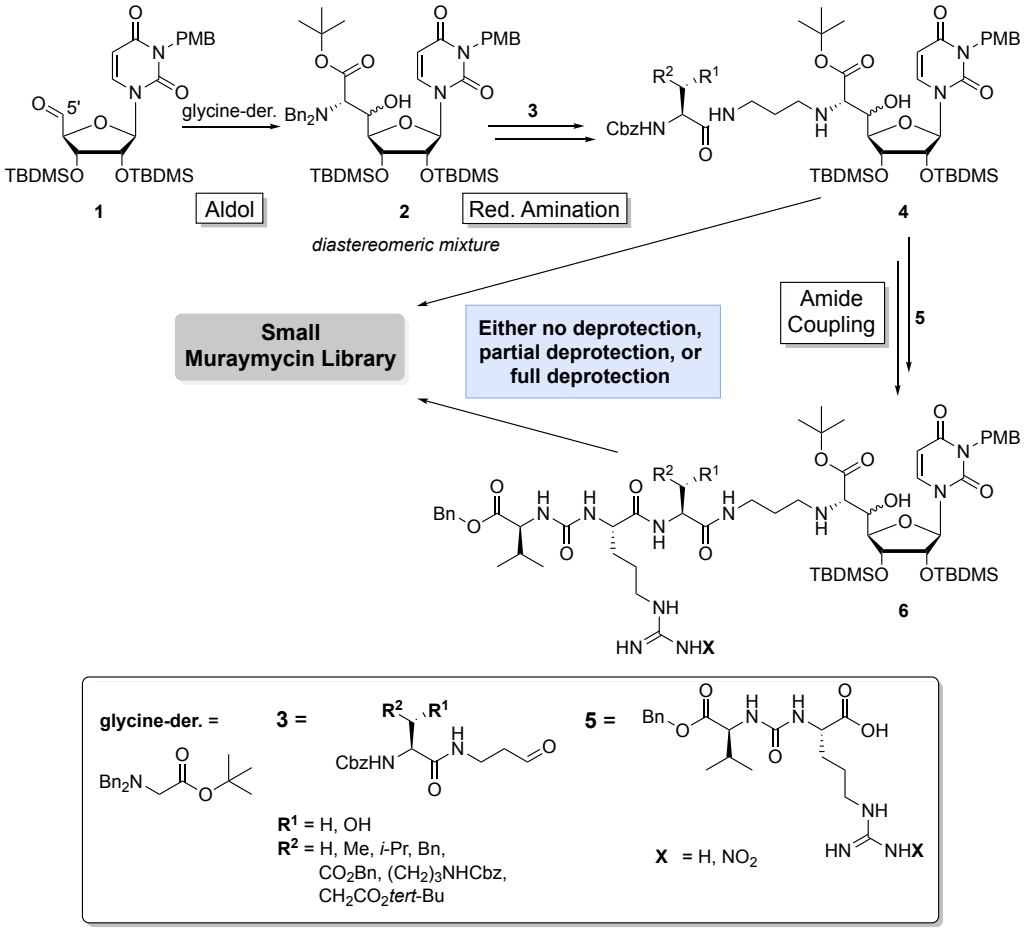
[a] N/A, not available.

Note: This table is not a comprehensive list and merely functions as a summary of the data provided in this chapter.

Synthesis of Muraymycins and Their Analogs

In order to study the SAR and antibacterial profiles of muraymycin analogs, as well as to achieve more drug-like molecules, efficient and affordable synthetic strategies to access their core scaffolds were developed. The first synthetic approaches to yield the muraymycin core structure were reported by Yamashita *et al.*[202] (Scheme 10.1.6). Similar to the reported biosynthetic route, the group started from a protected uridyl-5' aldehyde (**1**), which was converted with a protected glycine in an aldol reaction, followed by separation of the resulting diastereomers (**2**)[202]. Subsequently, reductive amination reactions with different linkers (**3**), amide couplings with simplified short peptides (**5**), and different deprotection strategies of intermediates **4** and **6** yielded a small library of some structurally simplified muraymycin analogs[202]. Notably, these analogs all lack the aminoribose group in the 5'-position, as well as the fatty acid moiety and the distinct cyclic epicapreomycidine in the peptide chain of naturally occurring muraymycins[202].

Scheme 10.1.6 | First Synthetic Approaches towards Muraymycin Analogs[a]

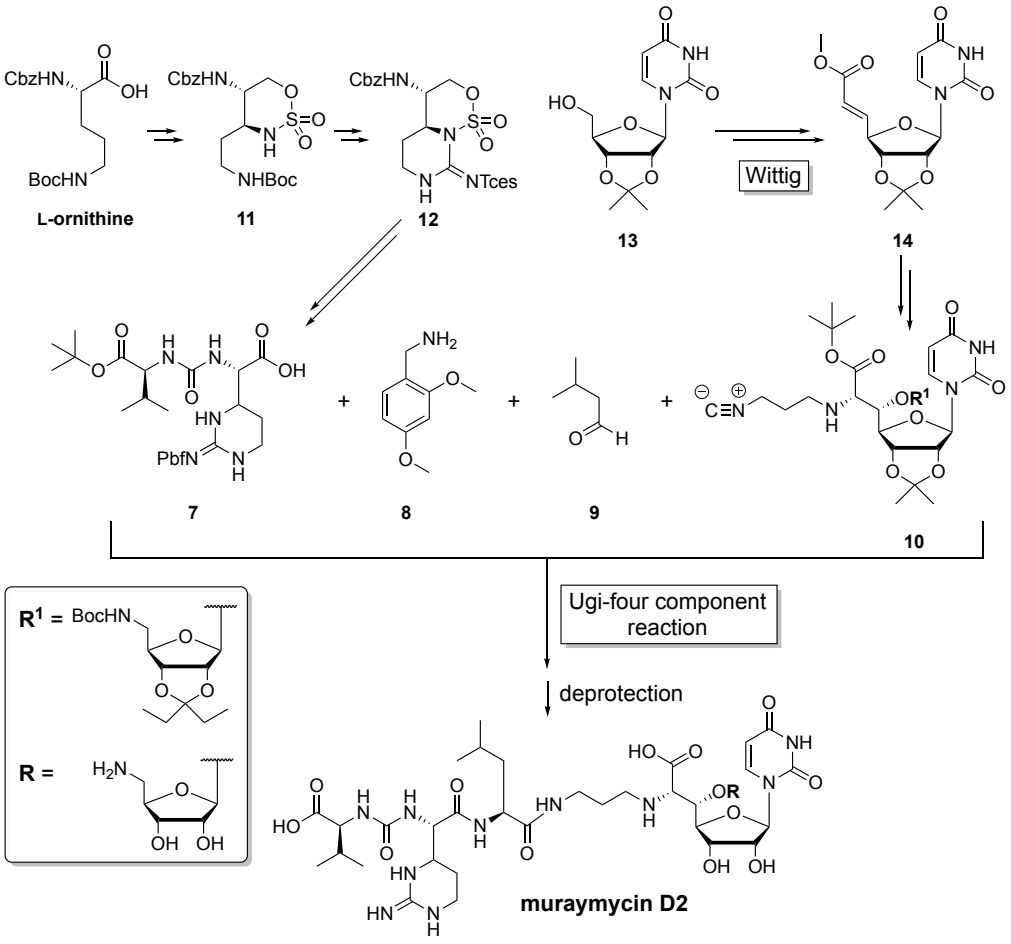


[a] Based on Yamashita *et al.*, *Bioorg Med. Chem. Lett.* **2003**, *13* (19), 3345-3350[202].
Bn, benzyl; Cbz, benzyloxycarbonyl; der., derivative; *i*-Pr, isopropyl; PMB, *p*-methoxybenzyl; red., reductive; TBDMS, *tert*-butyldimethysilyl ether; tert-Bu, *tert*-butyl.

In 2008, Tanino and co-workers developed a synthetic route for muraymycin analogs based on the Ugi-four component reaction method[203]. In their approach, a nucleoside core

isonitrile, a leucine-derived *aldehyde*, a benzylic *amine*, and a urea dipeptide *carboxylic acid* were coupled in a one-pot reaction to yield analogs of the D-Group muraymycins[203] (the four components of the Ugi-four component reaction are highlighted in *italic*). While these analogs included the aminoribose substituent at the nucleoside core, they only contained a ring-opened version of the epicapreomycidine amino acid[203] (not shown). Two years later, the group had developed a synthetic approach to yield the cyclic L-epicapreaomycidine amino acid, which enabled them to report the first total synthesis of muraymycin D2[204] (Scheme 10.1.7).

Scheme 10.1.7 | Total Synthesis of Muraymycin D2 *via* Ugi-four Component Reaction[a]



[a] Based on Tanino *et al*., *Nucleic Acids Symp. Ser.* **2008**, (52), 557-558[203], Hirano *et al*., *J. Org. Chem.* **2008**, *73* (2), 569-577[205], and Tanino *et al*., *J. Org. Chem.* **2010**, *75* (5), 1366-1377[204]. Cbz, benzyloxycarbonyl; Boc, *tert*-butyloxycarbonyl; Tces, 2,2,2-trichloroethoxysulfonyl.

For this, they combined the established synthetic route for the generation of the aminoribosylated uridine core, in which a uridine 5'-aldehyde is converted using Wittig reaction conditions, Sharpless aminohydroxylation, and subsequent β-selective aminoribosylation to yield key intermediate **10**[205], with their synthetic route for the epicapreomycidine building block (**7**)[204]. The latter was synthesized starting from L-ornithine *via* a C-H amination and the formation of sulfamate intermediate **11**, followed by cyclization of the desired isomer resulting in guanidine **12**[204]. The corresponding

intermediates were then taken forward with their Ugi-four component approach to yield muraymycin D2[204] (Scheme 10.1.7).

Similarly, another total synthesis route was developed in 2016 by Mitachi *et al.* for the muraymycin analog D1 (not shown) – only differing from muraymycin D2 in a methylated hydroxy moiety[201] (Table 10.1.1). While the group here also used a β-selective ribosylation reaction to introduce the additional carbohydrate moiety to the uridine core, the 3-aminopropyl linker was added *via* a stereoselective Strecker reaction[201]. Finally, the peptide was combined with the nucleoside core *via* an amide coupling, and the use of almost exclusively acid cleavable protecting groups simplified the endgame of this synthetic approach to one global acidic deprotection step[201].

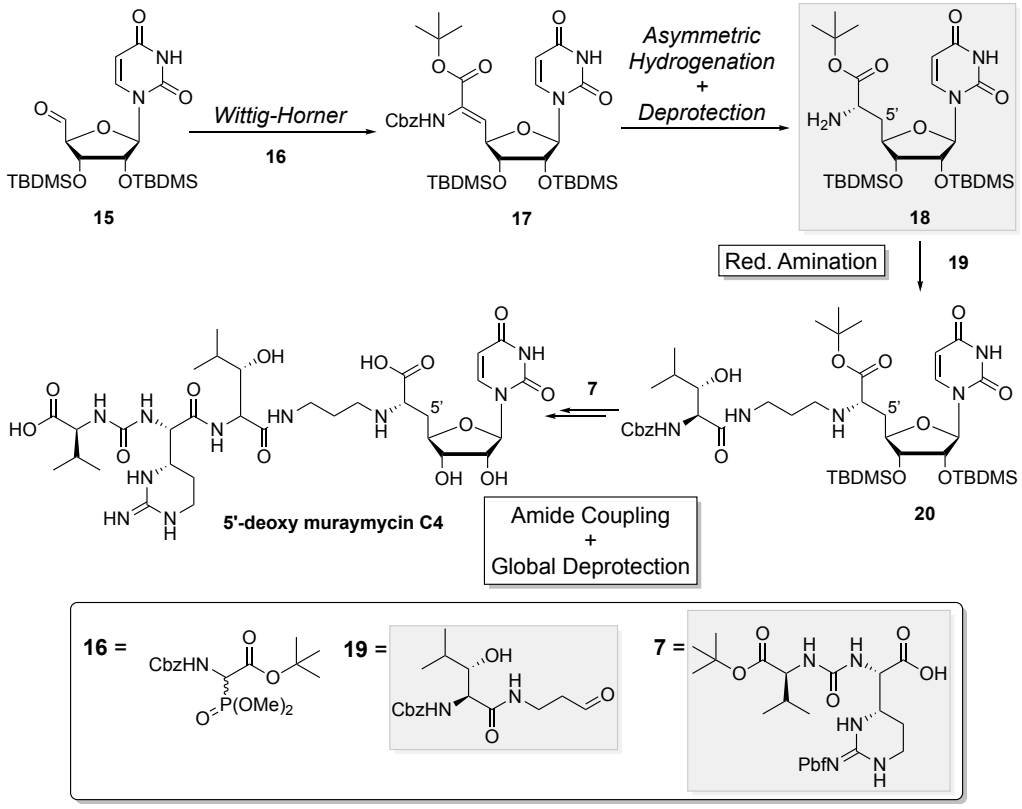
In the early 2010s, Spork *et al.* developed a tripartite synthetic approach for simplified 5'-deoxy analogs of C-Group muraymycins that carry a hydroxyleucine residue in the center[194,206,207] (Scheme 10.1.8). Starting from the precursor **15** (uridine-5'-aldehyde), the carboxylic acid motif was introduced to the nucleoside core *via* a Wittig-Horner reaction with phosphate **16**, followed by an asymmetric hydrogenation and a benzyloxycarbonyl (Cbz) deprotection step to yield key intermediate **18** in a stereoselective manner[207]. In the following, linker aldehyde **19**, including the stereoselective hydroxyleucine amino acid, was introduced using reductive amination conditions[207]. In the final steps, urea dipeptide analog **7** was reacted with intermediate **20** in an amide coupling and the resulting product was deprotected in an acidic global deprotection step, resulting in 5'-deoxy muraymycin C4 as the final compound[207].

Building block **19** was synthesized following a novel protocol described by Ries *et al.*[208] (not shown), and the urea dipeptide analog **24** was accessed following an optimized synthetic route to build the epicapreomycidine moiety (Scheme 10.1.9)[206]. For this, Garner's aldehyde was used to form a terminal imine (structure not shown), which was alkylated using a Grignard reaction to yield amine **21** with great diastereoselectivity[206]. The introduced alkene moiety was then transformed *via* ozonolysis and another sequence of imine formation and reduction to the respective amine to yield intermediate **22**[206]. The two primary amines of deprotected intermediate **23** were then cyclized using a specifically prepared guanidinylation reagent (**24**), resulting in the desired functional group[206]. The epicapreomycidine precursor (**25**) was further transformed to the respective amino alcohol, converted to the urea dipeptide (**27**), and oxidized to yield the desired carboxylic acid (**7**)[206]. Since this synthetic route was used for the synthesis of the muraymycin analogs part of this thesis, more detailed schemes and discussions are provided in Chapter 10.3.

Based on the described synthetic route, Spork and coworkers also attempted to access 5'-hydroxylated congeners such as muraymycin C4 (Scheme 10.1.10)[206]. For this, precursor **15** was used to form a uridine-epoxide intermediate (**29**), enabling the stereospecific formation of a bromohydrin (**30**), which was subsequently transformed into an azide by S_N2-type nucleophilic substitution[206]. The intermediate was eventually reduced to the respective amine, resulting in the nucleosyl precursor **31**, which was the first key building block for the synthesis of 5'-hydroxylated muraymycin analogs in no more than four additional steps, including the final global acidic deprotection step[206]. Importantly, while it takes 8 steps to form compound **31** from starting material **15**, it only takes a total of 3 steps to arrive at the respective analog (**18**) in the case of 5'-deoxy analogs. Surprisingly, the total synthesis of muraymycin C4, using this synthetic approach, did not succeed due to the chemical instability of intermediates **31** and **33** under acidic conditions[206].

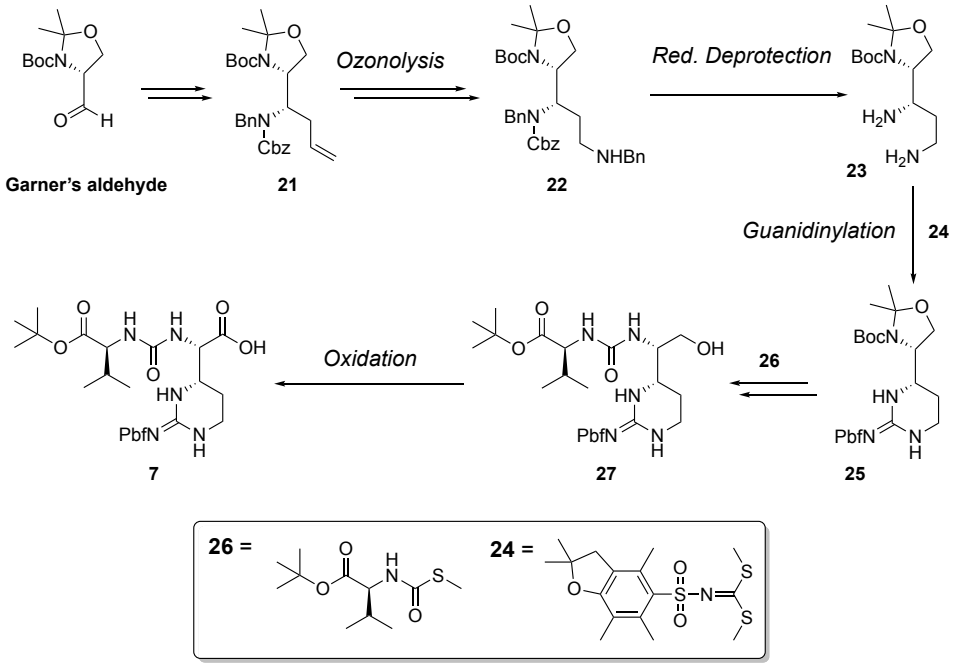
In summary, the three main building blocks of the tripartite approach developed by Spork and colleagues – uridine **18**, linker **19**, and dipeptide moiety **7** – are synthetically accessible intermediates with manageable and flexible synthetic routes for the introduction of chemical changes to the muraymycin core scaffold. Hence, *via* three synthetic steps – a reductive amination, an amide coupling, and final global deprotection – muraymycin core scaffolds can be built[206], making this synthetic route the most efficient and adaptable for building compound libraries and probes.

Scheme 10.1.8 | Tripartite Synthesis of 5'-Deoxy Analog of Muraymycin C4[a]



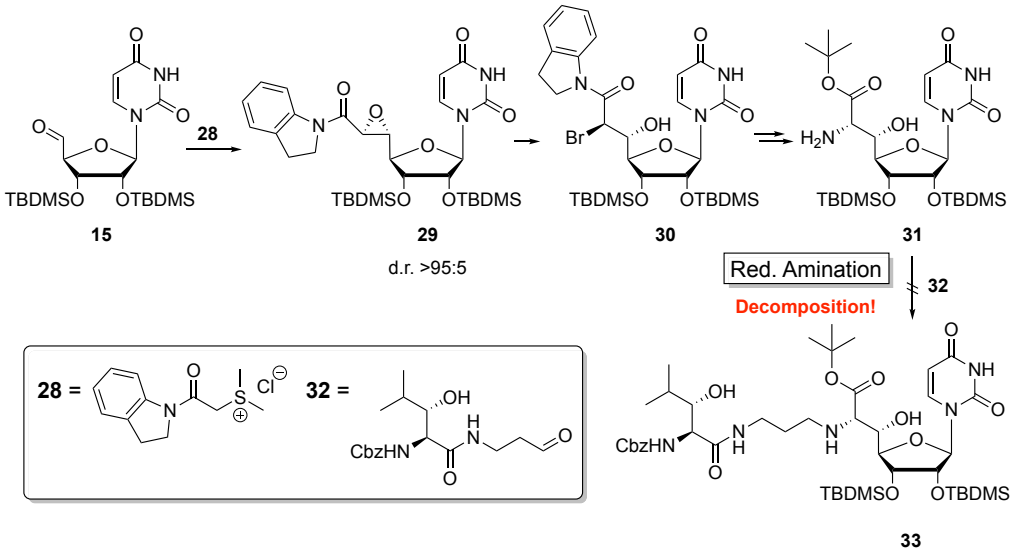
[a] Based on Spork *et al.*, *J. Org. Chem.* **2011**, *76*, 10083-10098[207].
The three key intermediates for the construction of the muraymycin core are highlighted in gray boxes. The final assembly of these building blocks can be achieved in three overall steps (highlighted in framed boxes).

Scheme 10.1.9 | Synthetic Strategy to Access Urea Dipeptide Intermediate 7[a]



[a] Based on Spork *et al.*, *J. Org. Chem*. **2011**, *76*, 10083-10098[207].
Red., reductive.

Scheme 10.1.10 | Attempted Total Synthesis of Muraymycin C4[a]



[a] Based on Spork *et al.*, *Chem. Eur. J.* **2014**, *20* (47), 15292-15297[206].
d.r., diastereomeric ratio.

Structure-Activity Relationship of Muraymycins as MraY Inhibitors

Before comprehensive synthetic routes for the muraymycin core were developed, Yamashita *et al.* reported antibacterial activity of some truncated muraymycin analogs, targeting exclusively Gram-positive bacteria[202]. Comparing naturally occurring muraymycin analogs in a fluorescence-based biochemical MraY assay, Koppermann *et al.* later showed that the fatty acid moieties in muraymycins from the A- and B-Groups potentially affect cellular uptake, rather than on-target activity[209]. Thus, these findings could provide an explanation for the lack of activity against Gram-negative bacteria of the truncated muraymycin analogs in Yamashita's studies. However, muraymycin analogs with a substituted leucine were generally also found to be more potent than D-Group muraymycins in *in vitro* MraY assays[209], pointing towards some importance of this moiety for on-target inhibition.

In the past decade, due to increasing advances in synthetic approaches to access muraymycin analogs, more information regarding the SAR of these promising nucleoside-based natural antibiotics has been reported. For instance, Tanino and coworkers demonstrated that several simplifications of the muraymycin core scaffold are tolerated, including the truncation of the dipeptide, exchange of the epicapreomycidine to a simpler ornithine moiety, as well as a direct conjugation of the central leucine function and long alkyl chains, with activities of these compounds against antibiotic-resistant strains, such as methicillin-resistant *S. aureus* (MRSA) and VRE[210]. Notably, truncations of the dipeptide moiety were generally not tolerated in our hands[211]. More importantly, full-length muraymycin core structures seem to be essential for MraY inhibitory activity in the context of 5'-defunctionalized analogs[212].

Other studies focused their investigations on specific parts of the muraymycin scaffold. For example, Heib *et al.* showed that changes in the uracil unit were generally not tolerated, while certain changes in the nucleoside moiety, such as 2'-defunctionalization, were well tolerated[213]. Similarly, analogs in which the aminoribose group was directly linked to the uracil ring lost MraY inhibitory activity[214]. Strategies to link the aminoribose moiety to the 6'-*N* of the nucleoside core also led to inactive compounds; however, smaller substituents at the 6'-*N* were decently tolerated[214]. Importantly, muraymycin analogs with a missing aminoribose lost biochemical potency by about 6,000-fold, while *O*-methylations on the aminoribose moiety had no effect on MraY activity[209]. Nonetheless, despite the significant loss of MraY inhibitory activity of muraymycin analogs lacking the aminoribose moiety (such as the 5'-defunctionalized muraymycin analogs described above), such analogs can still be decent MraY inhibitors, thus making them excellent compounds for SAR studies and the development of MraY-specific probes due to their superior chemical accessibility[209], as further discussed in the course of this work.

In another study, Spork *et al.* indicated that the stereochemistry at the 5'- and the 6'-position of the uridine motif as well as the central leucine amino acid might stir MraY inhibitory activity, with the most potent analogs of this study exhibiting *S*-configuration for the 6'-position (particularly pronounced for the 5'-deoxy analogs) and leucine, and *R*-configuration (*anti*) at the 5'-position of the uridine[215]. However, these effects were rather mild and the native muraymycin configuration ((5'-*S*,6'-*S*)-*syn* for the uridine moiety) was not included in this data set due to challenges in synthetically accessing respective analogs[215]. Regarding the configuration at the leucine, Tanino and coworkers also found

that *S*-configured leucines are generally more active than respective *R*-configured analogs[216].

The group further showed that longer alkyl chains, including analogs with incorporated aromatic ring systems within the chains, did not improve potency on MraY inhibitory activity[216]. Nevertheless, these analogs had improved antibacterial activity[216], which again points to a cellular uptake enhancing function of the fatty acid chain in muraymycins. Tanino *et al.* further demonstrated in another study that truncated analogs that lack the complete peptide motif were inactive in cells[210]; however, the resulting free carboxylic acid in combination with the smaller molecular size of the compound probably caused cellular uptake issues. Thus, this data is likely not informative about the biochemical MraY inhibitory activities of these compounds.

The SAR data for the MraY inhibitory activity of muraymycins presented in this chapter is summarized in Figure 10.1.23.

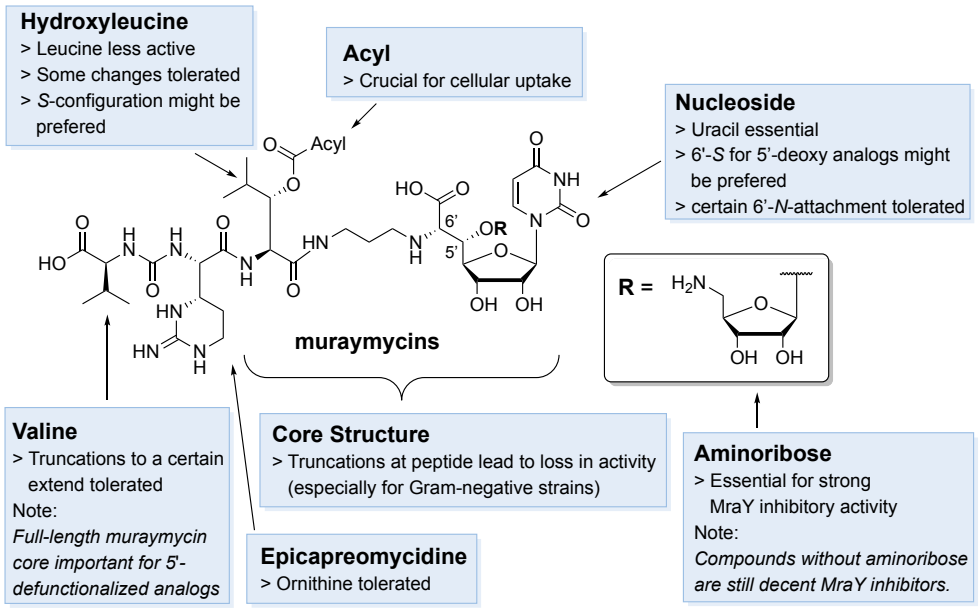


Figure 10.1.23 | Structure-Activity Relationship for MraY Inhibition by Muraymycin Antibiotics. SAR summary of data provided within this chapter.

10.2 Aim of this Project

When this project was started, co-crystal structures of MraY confirming the binding site of nucleoside antibiotics, such as muraymycins, were not available and more detailed studies of muraymycin analogs were still needed to fully understand their binding mode, SAR, as well as selectivity. In particular, information about binding affinity to other target proteins was missing and it was hypothesized that knowledge of these interaction partners could enlighten cellular uptake or efflux mechanisms of muraymycin antibiotics. Furthermore, other secondary targets in prokaryotic cells, and also potential off-targets in human cells, remained unknown. This is partially due to the lack of efficient tool compounds to study and detect muraymycin binding, thus being able to provide information on their binding profile across the human and bacterial proteomes.

In this work, ABPP strategies were applied to design a chemical probe that would allow a more in-depth analysis of muraymycins' mode of action, and which could be used to perform chemical proteomics studies in the human as well as the bacterial cell. Although co-crystal structures of several nucleoside antibiotics – including muraymycin D2[178] – in complex with MraY have been reported since the initiation of this project[178-180], comprehensive selectivity studies of this class of nucleoside antibiotics are still needed.

10.2.1 Design of Muraymycin-based A*f*BPs

Since A*f*BPs are supposed to provide information about a molecule of interest, the overall goal was to keep the muraymycin core structure mainly intact; meaning that the photoactive warhead structure as well as the bioorthogonal handle moiety were supposed to be as small as possible, as well as most efficiently incorporated into the parent muraymycin scaffold. The following considerations and designs are summarized in Figure 10.2.1.

In a first step, a diazirine was chosen as a photoaffinity warhead. The muraymycin scaffold does not naturally carry aryl groups, which would be necessary for the introduction of other PAL warheads such as aryl azides or benzophenones. In contrast, alkyl-based diazirines are established reactive groups for chemical proteomic studies, as discussed in chapter 10.1.1. Other advantages of diazirines include their long excitation wavelength of ~365 nm, as well as the high reactivity of the carbenes resulting from UV irradiation, which provide optimal conditions for cellular treatments with a diazirine probe and potentially fewer unspecific labeling events as compared to other warheads.

Based on available SAR data at that time, it was already known that truncated muraymycin scaffolds, lacking the terminal L-valine amino acid in the dipeptide moiety[210], or compounds with different alterations at the central leucine motif, with or without fatty acid groups[216], were still inhibiting MraY in biochemical activity assays. Consequently, it was envisioned that the valine or the fatty acid were ideal attachment points for the introduction of the photoactive warhead – here, a diazirine – and an alkyne group as the bioorthogonal handle. Diazirine-decorated long alkyl chains have been reported for example in the context of quorum sensing studies[62], and were here applied to the fatty acid chain of the muraymycin scaffold. Additionally, terminal alkyne groups within these diazirine-decorated alkyl chains have the advantage of combining both functional groups of A*f*BPs in one structural motif[62].

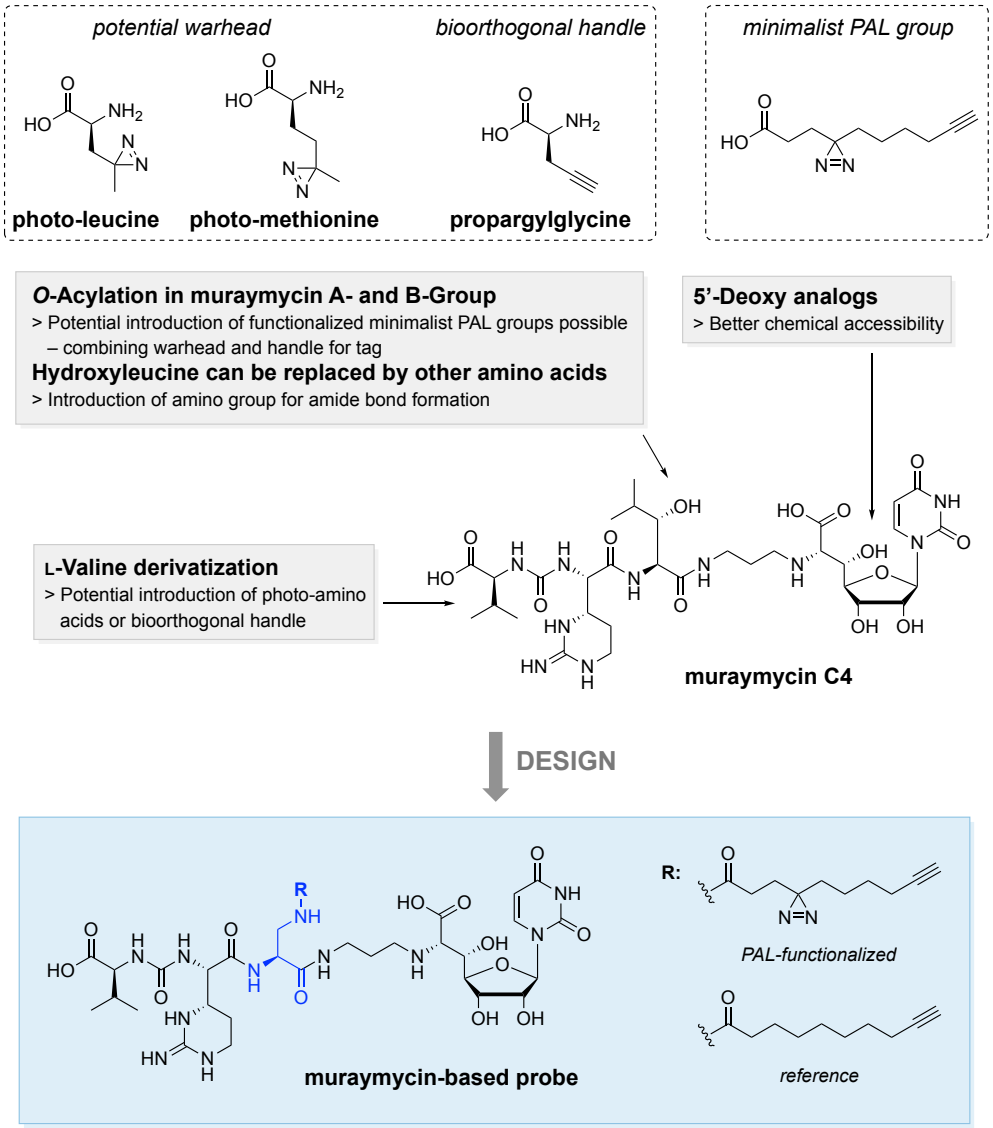


Figure 10.2.1 | Design of Muraymycin-based Chemical Probes. Considerations in the probe design based on muraymycin C4 analogs. 5'-Deoxy analogs simplify the chemical synthetic route accessing the main muraymycin core, while still exhibiting MraY inhibitory activity[194,206]. Even though introduction of the warhead or bioorthogonal handle are potentially possible within the L-valine moiety, a minimalist PAL group – combining both functional groups in one fatty acid moiety – would be most efficient. Shown is the first design of a muraymycin-based probe with either PAL-functionalized fatty acid or reference moiety. This design allows for the core scaffold to stay mostly unchanged, with only hydroxyleucine being exchanged for 2,3-diaminopropionic acid (blue), in order to shorten the synthetic route.

Another potential strategy to incorporate the diazirine group into the muraymycin scaffold could be the introduction of photo-amino acids, such as photo-leucine or photo-methionine[217], in place of the terminal valine, or incorporation of alkyne-carrying amino acids such as propargylglycine. Even though this feature of the peptide moiety in the muraymycin scaffold is not considered essential for MraY binding, in order to avoid unnecessary changes within the muraymycin core scaffold, the minimalist strategy, including both PAL groups incorporated into one fatty acid moiety, was deemed superior. Moreover, an established simplification in the muraymycin scaffold, leading to the 5'-

deoxy analogs, was considered (Figure 10.2.1). These analogs, missing the second sugar moiety including the respective hydroxy group at the nucleosyl unit, are much more synthetically accessible[194], while still exhibiting potent MraY inhibition[206].

Considering that changes in the leucine moiety were well tolerated[216], it was further envisioned that a primary amine in this position could serve as a flexible and easily accessible hinge for the introduction of different functionalized or unfunctionalized fatty acids. Amide couplings are common reactions carried out under mild basic conditions, which would not interfere with the chemical stability of the overall acid-labile protecting groups of the muarymycin core structure. In addition to that, there are a vast variety of respective amino acid derivatives, carrying a primary amine in the side chain while offering several different protecting groups and protecting group combinations, (commercially) available. This allows for convenient protecting group strategies for selective deprotection of this central building block, when introducing the fatty acid groups, while leaving the other protection groups of the core structure intact.

Finally, even though the changes made to the muraymycin scaffold are relatively minor, it is necessary to synthesize a respective analog (reference probe) without the diazirine moiety. This reference probe can be used to provide a *proof of concept* for the described probe design, while being synthetically more easily accessible than the respective diazirine analog. For this, biochemical MraY inhibitory activity of the probe is validated in our fluorescence-based MraY activity assay. Since the photosensitive diazirine warhead might not be stable under these assay conditions, it might also be problematic to determine MraY inhibitory activity of the final probe directly.

10.2.2 Retrosynthetic Considerations for the Synthesis of Muraymycin-based Chemical Probes

The general synthetic strategy to assemble the 5'-deoxy muraymycin core structure has been previously reported[194,206,207]. Four building blocks are connected to form the core structure of this nucleoside antibiotic (Figure 10.2.2). Starting from the 5'-deoxy nucleosyl amino acid (**BB1**), Cbz-deprotection, and subsequent reductive amination with aldehyde **BB2** would lead to the introduction of the 3-aminopropyl linker. Following another Cbz-deprotection, the leucine-replacing 2,3-diaminopropionic acid (**BB3**) could be introduced *via* amide coupling.

Since **BB3** has two amino groups, which have to be deprotected at different stages of the synthetic route and should not require acidic deprotection conditions with respect to the acid-cleavable protection groups of the core scaffold, Cbz and the base-labile fluorenylmethyloxycarbonyl (Fmoc) group were chosen. In a last step, following Cbz-deprotection and another amide coupling, the urea dipeptide unit (**BB4**) would be introduced, thus completing the muraymycin core scaffold. Under basic conditions, Fmoc, an amine protecting group well-known from solid-phase peptide synthesis, would be cleaved, allowing for the amide coupling with the respective PAL-functionalized fatty acid (**BB5-PAL**) or unfunctionalized reference fatty acid (**BB5-ref**), which would represent the fifth and last key building block of this retrosynthetic approach.

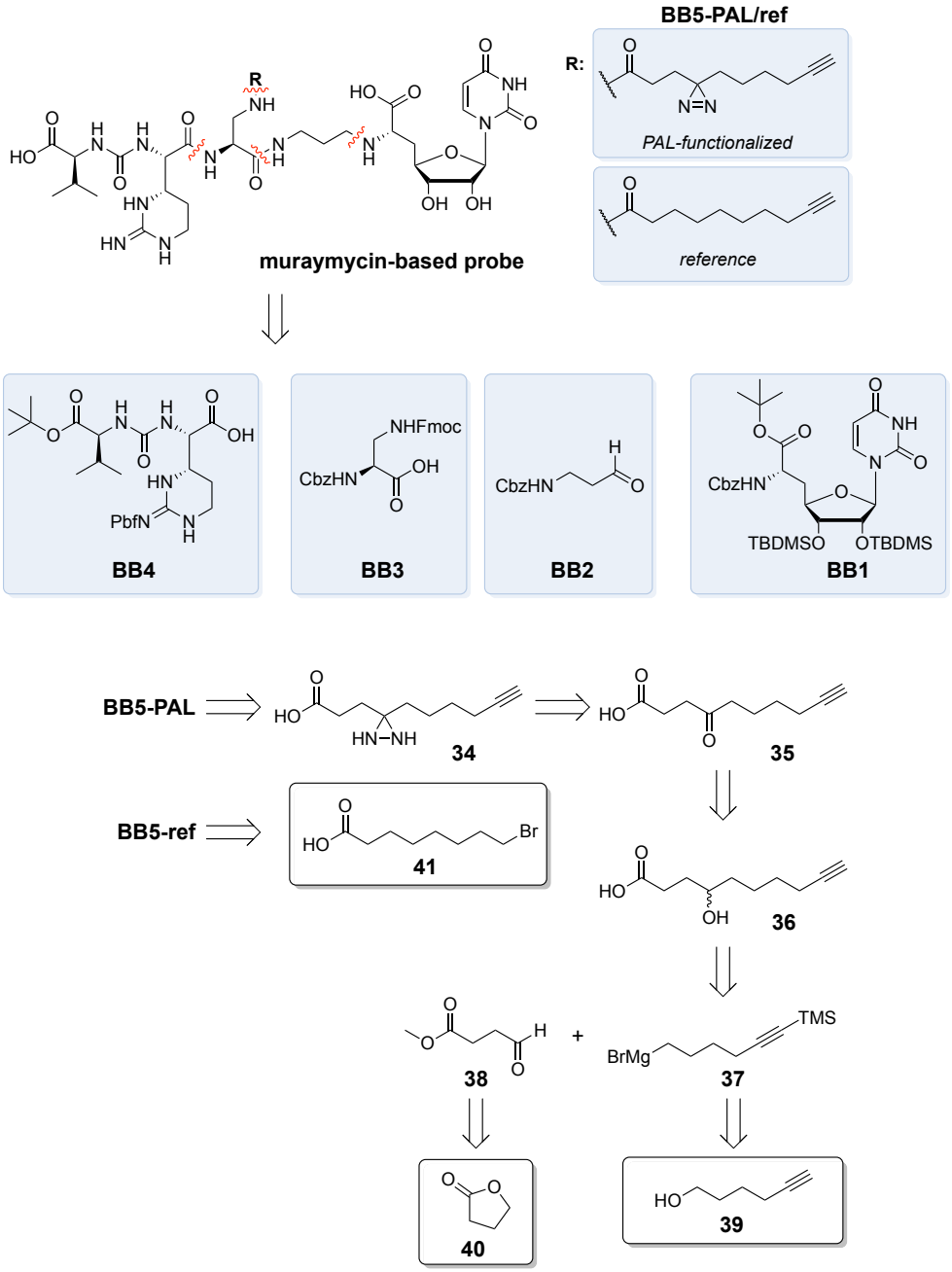


Figure 10.2.2 | Retrosynthetic Considerations to Access Muraymycin-based Probes. Five main building blocks (**BB1-5**) are necessary to form the muraymycin-based A*f*BP probe. While synthetic strategies to access **BB1**, **BB2**, and **BB4** have already been developed within our lab[194,206,207], additional retrosynthetic approaches for the synthesis of **BB5-PAL** and **BB5-ref** are given, starting from commercially available starting materials (shown in white boxes). **BB3** was commercially available.

The synthetic routes towards the uridinyl unit (**BB1**), the urea dipeptide intermediate (**BB4**), and the propionaldehyde analog (**BB2**) had already been developed within our lab[194,206,207] and protected diamino acid analogs (**BB3**) are readily available. In contrast, retrosynthetic approaches to access the fatty acid moieties (**BB5-PAL/ref**) had to be investigated. For this, a ketone precursor (**35**) could serve as the building block to access

the diazirine moiety (**BB5-PAL**) *via* diaziridine intermediate (**34**), for which several different reaction conditions have been described[88,91-93].

It was envisioned that the carbonyl-containing fatty acid (**35**) could be obtained from the respective alcohol (**36**) *via* oxidation using, for instance, 2-iodoxybenzoic acid (IBX). In a Grignard reaction, alkyne-containing Grignard reagent **37** could be converted with aldehyde **38** to form the desired C-C-bond, while also introducing a hydroxyl group at the coupling site. For this, the protected short carboxylic acid could be synthesized from γ-butyrolactone (**40**) under ring-opening with methanol and subsequent IBX-oxidation to the respective aldehyde. On the other side, the Grignard reagent could be potentially accessed through commercially available alcohol **39** that carries a terminal alkyne group. Alcohols can be transformed to the corresponding alkyl bromides using *N*-bromosuccinimide (NBS). Due to the potential reactivity of Grignard reagents with terminal alkynes, protection of the alkyne moiety with a trimethyl silyl (TMS) group was considered and could be achieved by silylation of the hydroxy and alkyne groups in a first step and subsequent selective deprotection of the resultant silyl ether under mild acidic conditions in a second. It was further envisioned that reference fatty acid **BB5-ref** could be synthesized starting from a respective ω-bromo-functionalized fatty acid (**41**), into which the desired alkyne moiety would be introduced *via* an alkylation reaction.

10.3 Results and Discussion

In the following chapter, the detailed synthesis of each building block for the synthesis of muraymycin-based A*f*BPs is described. The chapter is hence divided into three major parts. First, the synthesis of the two main intermediates **BB1** and **BB4** is presented in detail, since these building blocks make up a big part of this work – even though their synthetic routes had already been developed and reported by our group[194,206,207]. In the second part, the synthesis of the fatty acid moieties (**BB5-PAL/ref**), carrying either a photoactive diazirine warhead or being unfunctionalized, is discussed in detail, including some chemical stability investigations of the photoactive intermediate. With these building blocks in hand, the chapter ends in the description of the overall assembly yielding the full muraymycin core structure, while also providing first biochemical data of the synthesized reference probe.

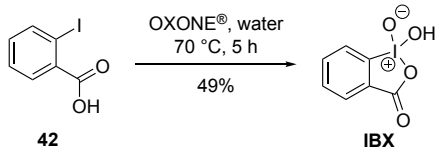
10.3.1 Synthesis of the Main Building Blocks

Synthesis of Reagent IBX

Some of the discussed reactions below are in need of a reagent to transform a hydroxy function into a carbonyl. A very useful and mild reagent for this purpose is IBX, which was synthesized in accordance with reported procedures from the commercially available precursor 2-iodobenzoic acid using potassium peroxymonosulfate ($2KHSO_5 \cdot KHSO_4 \cdot K_2SO_4$), also known as OXONE®[218]. Besides its reliable and versatile applicability for chemical synthesis, IBX is easy to synthesize and use, which is in part due to its poor solubility in many organic solvents[218,219].

The reaction was carried out in large scale as this reagent is part of a general supply within our lab. For this, 2-iodobenzoic acid (**42**) and OXONE® were stirred in water at 70 °C (Scheme 10.3.1). Since IBX has an extremely low solubility in water, the reaction mixture is cooled to 0 °C after completion, enabling simple filtration of the precipitated product, which is thus obtained as a pure white solid in good yield of 49%. This feature is highly convenient when using this reagent for oxidation reactions as well, as a quick filtration can also serve as a sufficient work-up for IBX-mediated oxidation reactions.

Scheme 10.3.1 | Synthesis of IBX



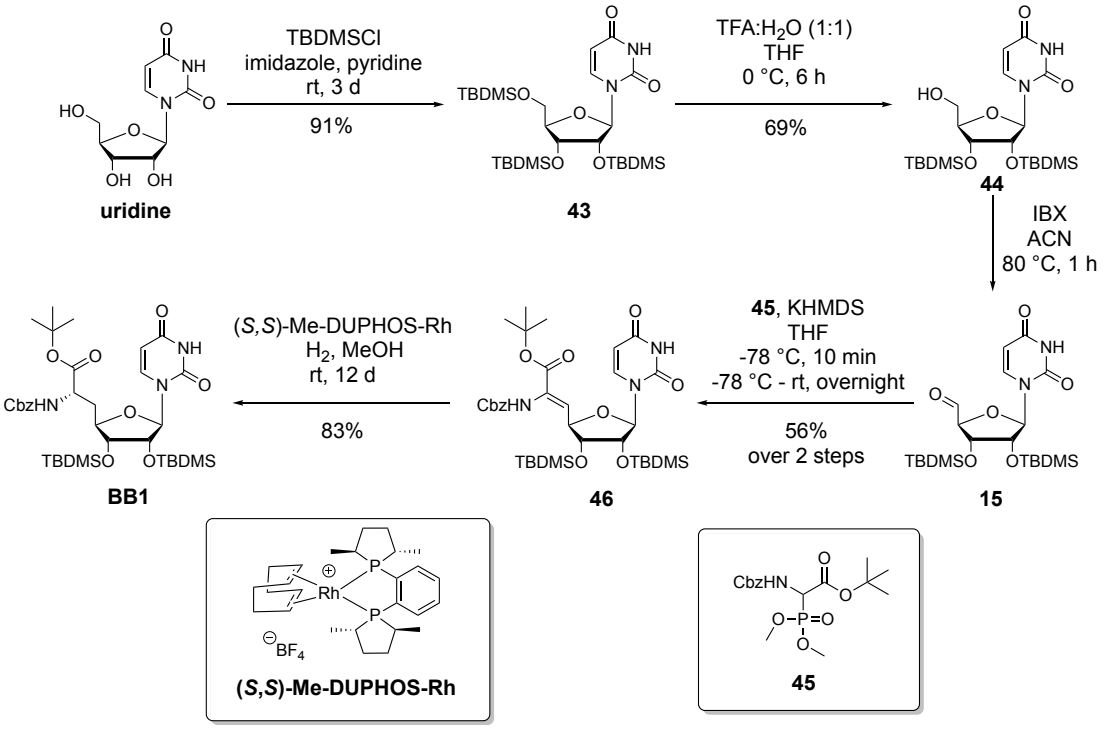
42 **IBX**

*Synthesis of the Uridine Building Block (**BB1**)*

The uridine moiety was synthesized following optimized protocols that had been developed within our lab (Scheme 10.3.2). According to procedures described in A. Spork's Ph.D. thesis[211], the three free hydroxyl groups in uridine were protected with TBDMS groups using TBDMSCl in pyridine with addition of imidazole. In a first step, uridine was co-evaporated with pyridine, re-dissolved in pyridine, and all reagents were added. The reaction mixture was stirred at room temperature for a few days until full protection of the nucleoside, which was obtained in 91% yield as a white foam upon aqueous work-up and column chromatography. A subsequent selective deprotection of the 5'-hydroxy group was

necessary to enable the following functionalization step. This reaction was carried out under careful monitoring of the reaction progress *via* thin-layer chromatography (TLC) control. Here, the tris-TBDMS-protected uridine intermediate (**43**) was dissolved in tetrahydrofuran (THF) and cooled to 0 °C. The reaction was quenched at the first appearance of double-deprotection products, resulting in the desired 2',3'-*O*-bis(*tert*-butyldimethylsilyl)uridine product (**44**) as a white foam in good yield of 69% after work-up and purification.

Scheme 10.3.2 | Synthesis of the Uridine Building Block (BB1)



The following two steps had to be performed in a successive manner due to the instability of the intermediate aldehyde (**15**). For this, alcohol **44** was oxidized using IBX in acetonitrile (ACN) under standard conditions known from the literature[219]. Upon heating of the reaction mixture to 80 °C, IBX dissolves just enough to perform the oxidation of the primary alcohol to the respective aldehyde function, while cooling and subsequent filtration of the resulting suspension led to pure product **15** without the need of further purification steps. Notably, the reaction progress was monitored by proton nuclear magnetic resonance (^{1}H NMR) spectroscopy, allowing for fast detection of the characteristic aldehyde proton signal. However, even with immediate and fast work-up, decomposition of the aldehyde moiety was already detectable in the NMR spectrum, highlighting the sensitivity of this intermediate.

Hence, this reaction was immediately followed by the conversion of the aldehyde in a Wittig-Horner reaction, as described by Spork *et al.*[194]. Here, aldehyde **15** was dissolved in THF and added dropwise to a flask with phosphonate **45** in THF and potassium bis(trimethylsilyl)amide (KHMDS) at -78 °C. Phosphonate **45** was prepared in large quantities in our lab and was kindly provided for this project. After addition of a slight excess of aldehyde **15**, the reaction mixture was allowed to warm to room temperature

overnight. Notably, it was previously shown by our group that the reaction yielded almost exclusively the *Z*-isomer of **46** using ^{1}H-NOE-NMR experiments[194]. Intermediate **46** was obtained after aqueous work-up and column chromatography purification as a white foam in 56% yield over two steps from **44**, which was in accordance to reported yields of around 66-67% for protected uridine analogs starting from the respective aldehyde intermediates[194], thereby not including potential loss of product due to decomposition of the aldehyde. Diastereomeric purity of the product was confirmed by ^{1}H NMR analysis.

In a last step, intermediate **46** was reduced using stereoselective hydrogenation conditions to yield the (6'*S*)-epimer **BB1** as the final uridine building block for further assembly of the muraymycin core structure. This strategy had been developed and validated by A. P. Spork[194,211]. For this last step, intermediate **46** was dissolved in methanol (MeOH) and degassed with argon gas. To control the stereochemical course of the reaction, the hydrogenation was performed in the presence of the chiral (*S*,*S*)-Me-DUPHOS-Rh catalyst, which is highly sensitive towards oxygen and thus requires strict inert gas atmosphere conditions at all times. The reaction mixture was stirred under slight overpressure of hydrogen gas (grade 6.0), which was re-applied every day for several days until completion of the reaction was observed. For this, a small sample was taken, worked up in a miniature scale work-up, and analyzed by NMR spectroscopy every other day. Column chromatography purification ultimately yielded the final product (**BB1**) as a white foam in a good yield of 83%.

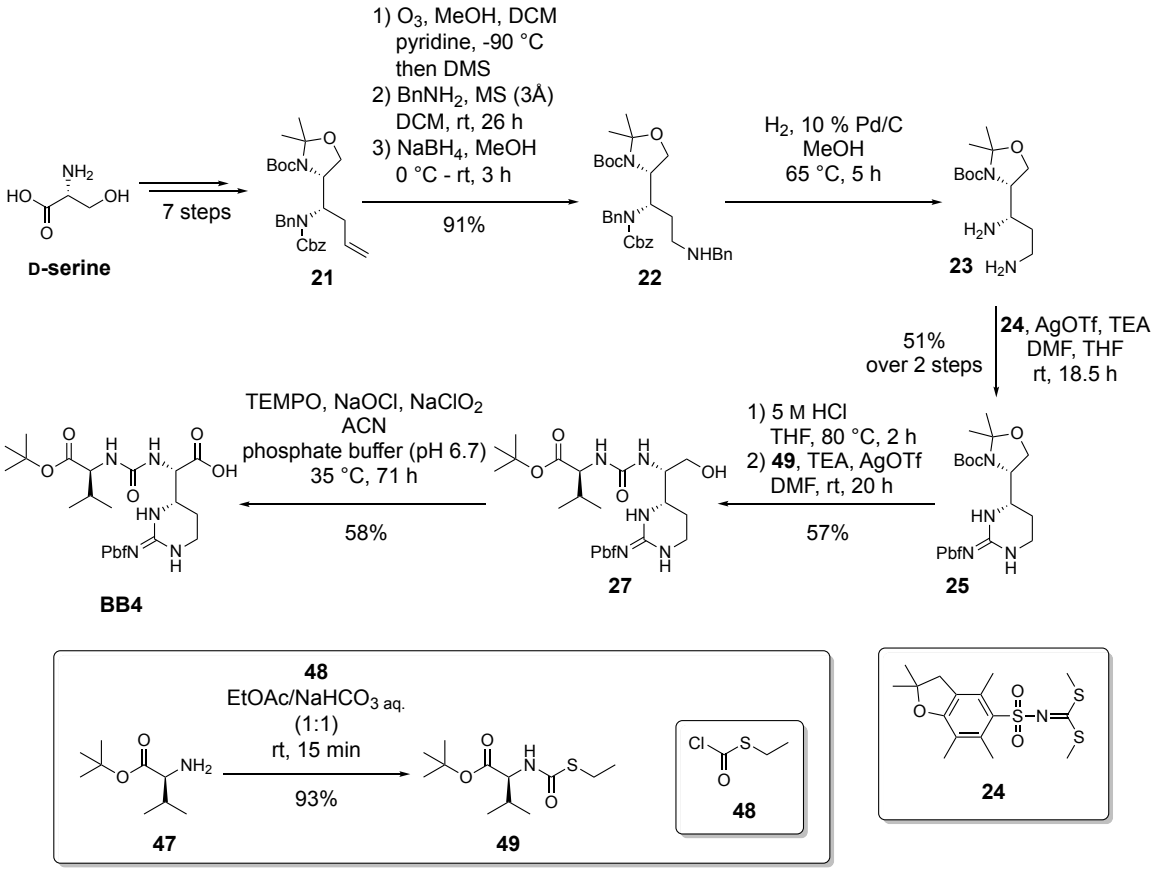
*Synthesis of the Urea Dipeptide Building Block (**BB4**)*

For the synthesis of the urea dipeptide intermediate (**BB4**), optimized synthetic routes developed in our lab were used (Scheme 10.3.3) and mainly based on efforts by M. Büschleb[220,221]. In a first step, protected valine **47** was functionalized with a thiocarbamate moiety. For this, ester **47** was dissolved in a 1:1-mixture of ethyl acetate (EtOAc) and a saturated aqueous solution of sodium bicarbonate ($NaHCO_3$). Commercially available thioethyl chloroformate (**48**) was added at room temperature and the reaction mixture was stirred for a few minutes. After completion, the reaction was quenched and the desired product **49** was obtained after aqueous work-up and column chromatography as a colorless oil in an excellent yield of 93%.

Intermediate **49** could then be directly coupled with the epicapreomycidine moiety. Important precursors of this moiety had been synthesized in large quantities within our lab and were kindly provided for this project. Starting from D-serine, Garner's aldehyde (not shown) was synthesized over five steps. Another two steps led to alkene intermediate **21**, which was kindly provided by K. Leyerer. The alkene was cleaved in an ozonolysis reaction with subsequent treatment of the reaction mixture with dimethyl sulfide (DMS), resulting in aldehyde formation. The reaction mixture was dried and the crude was re-dissolved in dichloromethane (DCM) including molecular sieve (pore size = 3 Å) to avoid moisture in the reaction mixture. Benzylamine was added and the mixture was stirred at room temperature for about one day, yielding an imine intermediate (not shown). The solvent was again removed *in vacuo* and the crude was re-dissolved in MeOH and cooled to 0 °C. Reduction of the imine with sodium borohydride ($NaBH_4$) yielded the respective benzyl-protected amine **22** after aqueous work-up and column chromatography as a light-yellow oil in 91% yield over three steps.

For the subsequent ring formation, the benzyl and Cbz protecting groups were cleaved. Thus, intermediate **22** was dissolved in MeOH and palladium on charcoal (Pd/C; 10 wt.% loading) was added. The reaction was stirred under a hydrogen atmosphere (1 bar; balloon) at 80 °C for a few hours until its completion, as observed by TLC control. The mixture was cooled to room temperature, filtered, and the solvent was removed *in vacuo* to yield pure product **23** as a clear, yellow oil. Without further purification steps, compound **23** was dissolved in a 1:1-mixture of THF and dimethyl formamide (DMF). Guanidinylation with dimethyl-2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-ylsulfonylcarbonimidodithioate (**24**) – which was available in large quantities in the lab and was kindly provided for this work – was achieved after addition of triethylamine as a base and silver trifluoromethanesulfonate (AgOTf) at room temperature and overnight stirring of the mixture, resulting in the formation of the 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf)-protected cyclic guanidine moiety. Filtration and subsequent aqueous work-up, followed by column chromatography, led to the desired product **25** in 51% yield over two steps from **22**.

Scheme 10.3.3 | Synthesis of Urea Dipeptide Building Block (BB4)[a]



[a] aq., aqueous; Bn, benzyl; MS, molecular sieve.

Acidic cleavage of the acetonide and Boc protecting groups in a 5 M HCl/THF mixture at 80 °C furnished the epicapreomycidine precursor for the subsequent urea formation with valine **49**. For this, the dried crude product from the deprotection step was re-dissolved in DMF and triethylamine was added as a base. AgOTf and **49** were added and the mixture was stirred at room temperature overnight. Filtration, aqueous work-up, and column

chromatography purification afforded product **27** as a white solid in a moderate yield of 57% over two steps from **25**.

In a final step, the primary alcohol was oxidized to the respective carboxylic acid using combined 2,2,6,6-tetramethylpiperidine 1-oxyl (TEMPO)-catalyzed and Pinnick reaction conditions, resulting in the formation of the amino acid motif of the desired building block[222]. Here, alcohol **27** was dissolved in acetonitrile and phosphate buffer (pH = 6.7). After addition of catalytic amounts of TEMPO reagent as well as sodium chlorite ($NaClO_2$) in excess, the mixture was slightly warmed to 35 °C. Aqueous sodium hypochlorite (NaOCl) solution was added dropwise to the reaction mixture, which was stirred at this temperature until completion of the reaction. Following aqueous work-up and column chromatography purification, the desired urea dipeptide building block **BB4** was obtained as a white solid in 58% yield.

10.3.2 Synthesis of the Fatty Acid Building Blocks

Essential to the development of the discussed muraymycin-based probes was the incorporation of both a photoactive group (here: diazirine) and a bioorthogonal handle (here: alkyne) into a fatty acid-like moiety. The minimalist fatty acid PAL group was derived from the *O*-acylated muraymycins of the A- and B-Groups, which are naturally decorated with long fatty acids at the hydroxyleucine moiety (Table 10.1.1). Hydroxy-functionalized alkyl chains with ω-alkyne groups were commercially available and alkyl-based diazirines are commonly accessed from respective ketones[88]. Thus, we sought to synthesize minimalist PAL fatty acids, with an alkyne group at the ω-position and a ketone within the chain, as well as strategies for converting the ketone into a diazirine group.

In the first step, synthetic routes to build the fatty acid core were investigated. Here, it was critical to introduce a ketone moiety, which would serve as the starting point for the conversion into a diazirine moiety, into the alkyl chain. Subsequently, the synthesized diazirines were used to interrogate the chemical stability of the probe. Lastly, since a reference compound with the same fatty acid as the photoactive probe but lacking the diazirine warhead was required for biochemical evaluation of the probe as an MraY inhibitor, synthetic routes of respective unfunctionalized ω-alkyne fatty acids are described.

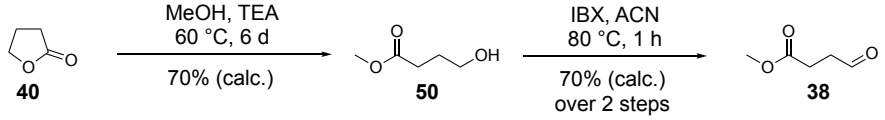
Synthesis of the Fatty Acid Core

To access a fatty acid with a ketone moiety within its chain, reaction of a Grignard reagent with an aldehyde could result in the desired long alkyl chain, forming a hydroxyl group at the coupling site, which could subsequently be oxidized to yield the respective ketone intermediate (Figure 10.2.2). As similar reactions between Grignard reagents and aldehydes had been reported[223-225], we began with the synthesis of short alkyl chains with a protected carboxylic acid on one end and an aldehyde moiety on the other. An elegant and affordable way of synthesizing such an aldehyde is through ring-opening of γ-butyrolactone using MeOH, which would yield the respective methyl ester and a 4-hydroxy group, followed by alcohol oxidation (Scheme 10.3.5).

Using previously reported conditions[223], lactone **40** was dissolved in MeOH and stirred at 60 °C in the presence of triethylamine as a base. Although yields of over 90% after 15 hours are reported in the literature[223], overnight conversion only gave about 25% of ring-opening product, as determined by ^{1}H NMR analysis, in our hands. After another 6 days of heating the reaction mixture, this ratio could be shifted to only 20% of cyclic

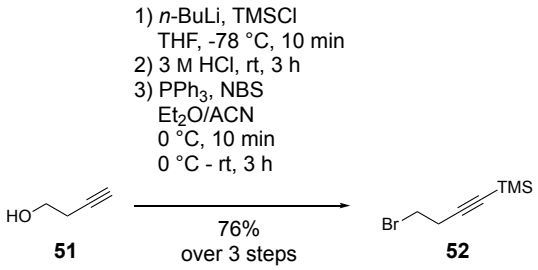
starting material (**40**). The dried crude product was obtained as a yellow oil with about 70% calculated product yield, considering a 1:4-ratio of product (**50**) to starting material (**40**). Since the starting material would not interfere with following reactions, the crude could be used without further purification. Using the prepared IBX reagent[219] enabled the transformation of the primary hydroxy group into the desired aldehyde. For this, incubation at 80 °C in acetonitrile for as little as one hour was sufficient to complete the reaction and generate desired intermediate **38** in 70% calculated yield over two steps from lactone **40** (^{1}H NMR analysis showing a 1:4-ratio of **40** to **38**). Interestingly, the resulting aldehyde exhibited excellent stability and was still intact after several weeks of storage at -20 °C.

Scheme 10.3.5 | Synthesis of Aldehyde 38



The synthesis of the alkyne motif – the second part of the fatty acid structure – began with commercially available long-chain (C_6) alkyne alcohols. Since shorter alkyl chains (C_4) were more affordable than the relatively expensive long chains, the below-described synthetic strategy was first established using short chain analogs and later applied to longer alkyl chains. Here, the primary alcohols could be transformed into bromides following a previously reported synthetic route[226], which was successfully employed to generate the desired Grignard precursors (Scheme 10.3.6). For this, TMSCl was used to protect both the alcohol and the alkyne groups. To do so, starting material **51** was dissolved in THF and cooled to -78 °C. Under nitrogen atmosphere, about two equivalents of the strong base *n*-butyllithium (*n*-BuLi; 2.5 M in *n*-hexanes) were carefully added, while making sure to maintain a temperature of approximately -78 °C. This mixture was stirred for a few minutes, leading to deprotonation of the hydroxy and alkyne groups. TMSCl was then added, and the reaction mixture was allowed to warm to room temperature, thereby generating the dual-TMS-protected intermediate (not shown).

Scheme 10.3.6 | Synthesis of Bromide 52



In the next step, the primary alcohol was selectively deprotected by addition of a 3.0 M aqueous HCl solution. After stirring the reaction mixture at room temperature for about three hours, an aqueous work-up followed to give the crude intermediate alcohol (not shown). Lastly, the crude product was dissolved in a 1:1-mixture of diethyl ether (Et_2O) and acetonitrile. After addition of triphenylphosphine (PPh_3), the mixture was cooled to 0 °C and NBS was slowly added. Upon addition of the bromination reagent, the mixture

was allowed to warm to room temperature and was continuously stirred for a few hours until completion of the reaction was confirmed by TLC control. The resulting suspension was diluted with petroleum ether (PE), filtered, and purified using a very short silica column, since the only impurities in the crude would separate very well using PE/EtOAc (99:1) as the mobile phase, resulting in a convenient and fast work-up procedure. The product (**52**) was obtained as a slight-yellow oil with a yield of 76% over three steps from **51**.

In order to connect intermediates **38** and **52** to yield the desired fatty acid, bromide **52** was converted into a Grignard reagent. For this, commonly used conditions that had been reported for structurally related compounds involved stirring the respective bromide compound in dry THF in the presence of magnesia turnings at room temperature[227,228]. Furthermore, addition of catalytic amounts of iodine was reported to serve as an excellent activator for Grignard formation[228]. Since this reaction could not be monitored by TLC, LC-MS, or NMR analysis, the putative Grignard reagent was immediately used in the subsequent coupling reaction with aldehyde **38** to confirm Grignard formation.

Several different conditions for reactions between Grignard reagents and aldehydes are reported, commonly employing Et_2O or THF as a solvent and using temperatures ranging from -78 °C to room temperature[223,225,227,229]. Consequently, THF as a solvent was chosen so that the solvents for Grignard reagent formation and subsequent coupling were the same (Scheme 10.3.7). Moreover, -30 °C was chosen as an initial reaction temperature because the reactivity of the Grignard reagent was not yet known. However, as it was not certain whether those coupling conditions would work in this case, the Grignard formation reaction itself was carefully monitored. Certain indicators, such as self-sustained heating of the mixture, local boiling of the solvent, or dissolving of the magnesia turnings, could indicate a positive Grignard reagent formation. Conditions used for this step are summarized in Table 10.3.1.

Scheme 10.3.7 | Development of the Synthetic Route to Access Fatty Acid 57

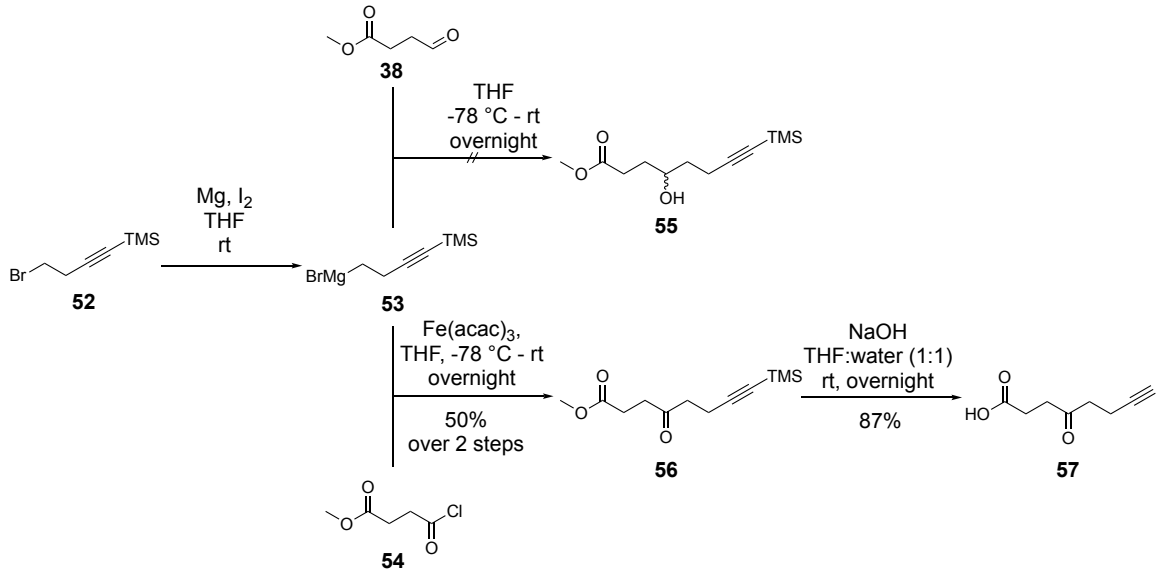
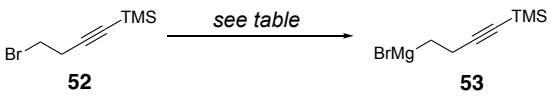


Table 10.3.1 | Optimization of the Formation of Grignard Reagent 53



52 **53**

Attempt	52 [eq.]	c(52) [M]	Mg [eq.]	Mg Batch	Heated	Outcome
1	1.0	0.26	3.4	old, untreated	no	**no conversion**
2	1.0	0.24	1.5	old, crushed	yes	**no conversion**
3	1.0	0.16	1.7	old, polished	yes	**no conversion**
4	1.0	0.87	1.5	new, untreated	yes	**conversion**

In the first attempt, **52** in THF was slowly added to magnesia turnings, and a few crumbs of iodine were added to help start the reaction. However, no signs of Grignard formation, such as warming of the reaction mixture, was observed and subsequent treatment of aldehyde **38** with the potential Grignard reagent did not lead to the desired product, even after the reaction had been allowed to warm to room temperature. Moreover, TLC control showed the formation of several side products, the identities of which could not be determined by NMR spectroscopy. In a second attempt to synthesize Grignard reagent **53**, the mixture was carefully heated using a heat gun to help initiate the reaction. Additionally, the relatively old magnesia turnings were crushed to increase unoxidized surface area. Nonetheless, the reaction did not warm up by itself or exhibited other signs of Grignard formation. Due to the old age of the magnesia turnings, the presence of an oxidized coating might have hindered this reaction. As a result, the magnesia turnings were polished using emery paper, but no conversion was observed.

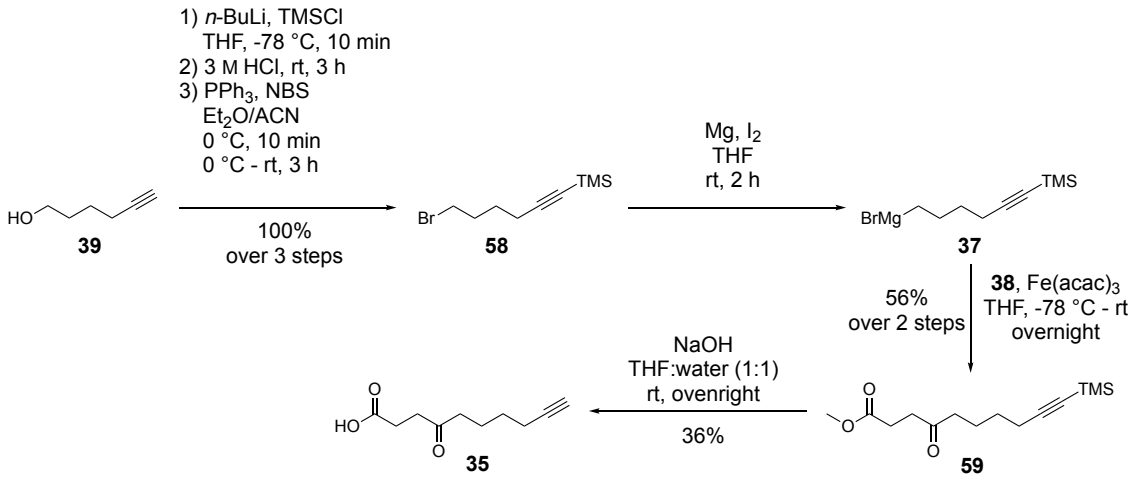
Different reaction conditions in the literature[223,225,227] commonly report the use of rather concentrated (~0.67-0.87 M) solutions of the bromide reagent, whereas so far, only concentrations of up to 0.26 M of bromide **52** in THF had been used in the aforementioned test reactions. Consequently, the reaction was repeated using a fresh batch of magnesia turnings to avoid time-consuming polishing of this reagent and a 0.87 M solution of bromide **52** in dry THF. After addition of a small portion of bromide reagent, the mixture was gently heated with a heat gun and autonomous heating of the reaction mixture could be observed, indicating initiation of the now self-sustained Grignard formation. After complete addition of **52** in THF to the reaction mixture, the resulting Grignard reagent was immediately mixed with aldehyde **38** (Scheme 10.3.7). For this, **53** was slowly added to a stirred solution of **38** in dry THF at -78 °C. However, several side products, but not the desired product, were observed by TLC control and NMR analysis, indicating that a change in the synthetic strategy was needed.

Returning to the literature, a recent paper had reported a synthetic route for minimalist PAL-functionalized fatty acids[230]. In this approach, the Grignard reagent was converted

with a carboxylic acid chloride in the presence of iron (III) acetylacetonate ($Fe(acac)_3$), leading not only to the desired fatty acid chain but also to the formation of the ketone function needed for diazirine synthesis. Notably, iron-catalyzed cross coupling reactions have a wide range of applicability, and iron salts usually exhibit low toxicity and cost[231]. Conveniently, methyl 4-chloro-4-oxobutyrate (**54**) was commercially available, resulting in an overall reduction of steps needed in the synthesis of the fatty acid building blocks, including the fact that this strategy directly results in the desired ketone moiety instead of a hydroxyl group. Following this procedure, acid chloride **54** and $Fe(acac)_3$ were dissolved in THF and freshly prepared Grignard reagent **53** was added dropwise at -78 °C. The mixture was allowed to warm to room temperature overnight, and a distinct product was observed *via* TLC control. After aqueous work-up and column chromatography, NMR spectroscopy verified that the desired product (**56**) had been obtained as a light-yellow oil in moderate yield of 50% over two steps from **52**.

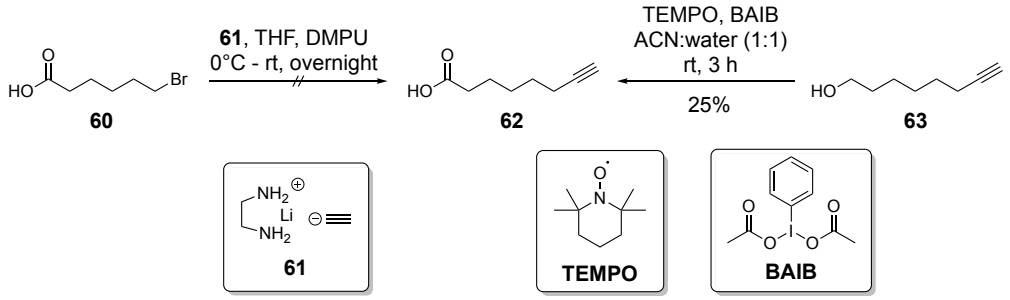
In a last step, intermediate **56** was deprotected with sodium hydroxide as a base to yield free fatty acid **57**[231] in a good yield of 87% (Scheme 10.3.7). This optimized synthetic scheme was further successfully applied to synthesize the longer fatty acid analog **35** (Scheme 10.3.8) in 20% overall yield (6 steps). Here, starting from the 6-carbon alkyl chain precursor alcohol **39**, addition of TMSCl under highly basic conditions resulted in the formation of TMS protection of the primary alcohol moiety and the alkyne. Subsequent acidic treatment of the intermediate selectively deprotected the primary alcohol, which was then converted to a bromide using PPh_3 and NBS to form Grignard precursor **58** in 100% yield over three steps. As described above, bromide **58** was then converted to Grignard reagent **37** using a fresh batch of magnesia turnings and a few crumbs of iodine to initialize the reaction. The resulting Grignard reagent was directly converted with aldehyde **38** in a $Fe(acac)_3$-containing THF solution at -78 °C, which was allowed to warm to room temperature overnight. Product **59** was obtained in 56% yield over two steps from **58**. After final deprotection of the carboxylic acid moiety under basic conditions, acid **35** was isolated in 36% yield.

Scheme 10.3.8 | Synthesis of Fatty Acid 35

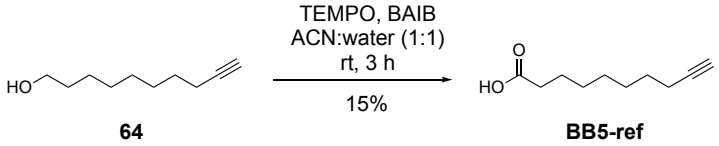


Fatty acids for the reference compound lacking the photoactive diazirine moiety were accessible from commercially available ω-bromo fatty acid analogs (Scheme 10.3.9). First, using previously reported conditions[232], a mixture of lithium acetylide and ethylenediamine (commercially available, 90% purity; **61**) was dissolved in dry THF under inert gas atmosphere, while bromide **60** was dissolved in *N*,*N*'-dimethylpropyleneurea (DMPU). The latter was added to the THF solution of **61** at 0 °C and was allowed to warm to room temperature overnight. As before, this synthetic strategy was first tried with a more affordable shorter alkyl chain analog (C_6). However, no conversion was observed. Hence, another strategy, starting from alkyne-functionalized long chain alcohols which could be oxidized to the respective carboxylic acids, was investigated (Scheme 10.3.9). Inspired by protocols developed by A. Lemke in our lab[233], commercially available alcohol **63**, TEMPO, and bis(acetoxy)iodobenzene (BAIB) were dissolved in a 1:1-mixture of acetonitrile and water. The mixture was stirred at room temperature for a few hours and showed full conversion *via* TLC control. After aqueous work-up and column chromatography, pure product **62** was obtained in 25% yield. Notably, the reaction likely yielded quantitative conversion; however, some fractions had an unknown impurity that was challenging to remove by column chromatography, so only highly pure fractions were collected and used for subsequent reactions. This synthetic route was applied to the longer fatty acid analog (C_{10}), yielding **BB5-ref** as a key intermediate in 15% yield from alcohol **64** (Scheme 10.3.10).

Scheme 10.3.9 | Development of the Synthetic Route to Access Fatty Acid 62



Scheme 10.3.10 | Synthesis of Reference Fatty Acid (BB5-ref)

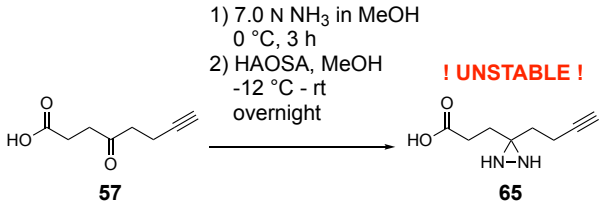


Synthesis of the Diazirine Function

With ketone-functionalized fatty acids in hand, the next step was to synthesize the diazirine moiety *via* a diaziridine intermediate (Scheme 10.3.11). Based on reported procedures for similar molecules[231], ketone **57** was dissolved in dry MeOH and cooled to 0 °C. A 7.0 N solution of ammonia in MeOH was added. The mixture was stirred for a few hours and a solution of HAOSA in MeOH was slowly added at this temperature. The reaction was allowed to warm to room temperature and TLC control confirmed conversion. Subsequently, the crude was filtered and dried to yield a yellow oil. However, NMR spectra

of the crude product were difficult to interpret, and subsequent TLC analysis suggested instability of the diaziridine intermediate. Multiple other attempts of this synthetic step – all of which resulted in decomposition of the potential product (**65**) in the course of preparing or handling the molecule during NMR analysis, or even during work-up steps – confirmed this instability. Decomposition was mostly indicated by a sudden precipitation of a white solid, which could not be identified, as well as multiple spots in TLC analysis. Due to these stability issues, it was decided to immediately transform diaziridine **65** into the respective diazirine in a follow-up reaction.

Scheme 10.3.11 | Synthesis of the Diaziridine Intermediate 65



For optimal handling of this sensitive intermediate (**65**), additional reported work-up conditions for the diaziridine were investigated (Table 10.3.2). For example, the reaction mixture was dried under a stream of nitrogen gas instead of in a rotary evaporator, resuspended in MeOH, filtered, concentrated *in vacuo*, and then diluted with MeOH to continue with the next step[234] (**Method A**); or alternatively, the mixture was filtered and concentrated to about 0.5 mL after addition of triethylamine, then diluted with MeOH and more triethylamine to directly continue with the oxidation step to the diazirine moiety[235] (**Method B**).

Table 10.3.2 | Optimization of Work-Up Conditions for Diaziridine Intermediates

Attempt	1st Step	2nd Step	3rd Step	4th Step	5th Step	Outcome
1	filtration	drying *in vacuo*	oxidation	-	-	**random decomposition**
2 (Method A)	drying in N_2 stream	re-dissolving in MeOH	filtration	drying *in vacuo*	oxidation	**~stable intermediate (!not reliable!)**
3 (Method B)	filtration	addition of TEA	concentration to ~0.5 mL	addition of TEA	oxidation	**stable intermediate**

Subsequently, crude products from both Methods A and B were treated with an iodine solution in MeOH at 0 °C (Scheme 10.3.12). The reaction was performed in the dark until its completion was observed as indicated by the persistence of the yellow-brown color of the iodine solution for longer than 1 min. TLC control confirmed formation of only one product for each method, which showed decomposition after UV irradiation (365 nm) of the TLC spots prior to development of the TLC plate (Figure 10.3.1). For both test reactions, the solvent was removed *in vacuo* and the crude products were re-dissolved in

water and acidified with a 0.1 M HCl solution. The aqueous layers were extracted with EtOAc, and crude products were obtained after drying of the combined organics. TLC analysis after work-up showed only one spot, indicating sufficient purity of product **72** to continue further reactions. This was preferred as this would avoid unnecessary handling of this potentially sensitive intermediate. While crude NMR spectra were still difficult to interpret, infrared (IR) analysis revealed a characteristic peak at ~1,586 cm^{-1} (Figure 10.3.2) as the known signal for the diazirine moiety[236], thus confirming its successful synthesis.

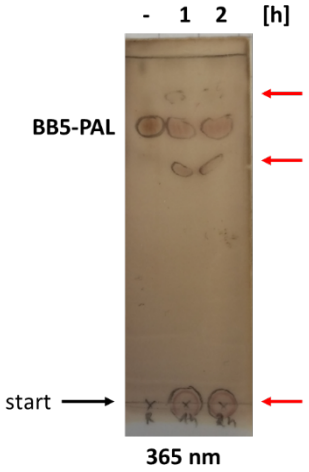
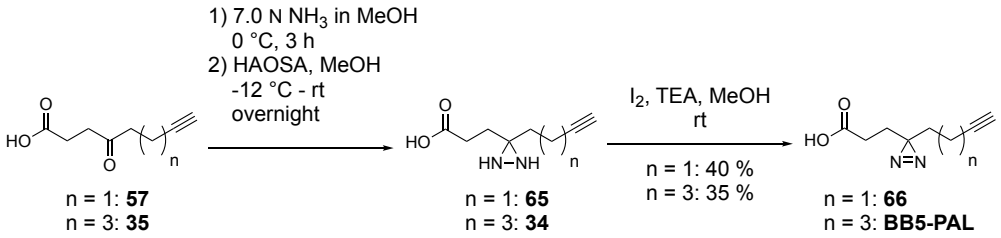


Figure 10.3.1 | UV Irradiation of Diazirine-functionalized Fatty Acid BB5-PAL. A solution of **BB5-PAL** in MeOH was used for TLC analysis. The spots at the start line were irradiated using 365 nm UV light (<5 cm distance) for 0 h, 1 h, and 2 h. After irradiation, the TLC plate was developed using DCM:MeOH (9:1) and dyed with vanillin/sulfuric acid. Obvious decomposition (red arrows) of **BB5-PAL** was observed after UV irradiation with a strong spot at the start line.

Scheme 10.3.12 | Synthesis of Diazirine-functionalized Fatty Acids



In summary, due to the easier diaziridine work-up of Method B, as well as a higher chance for potential decomposition of intermediate **65** during work-up in Method A, Method B was chosen as a reliable standard work-up for diaziridine intermediates. As seen with other methods, this synthetic route was applied to the longer fatty acid analog **35**, resulting in similar yields of 35% over two steps of pure diazirine (**BB5-PAL**) after column purification, compared to 40% overall yield for the short fatty acid analog (**66**). The purified analogs gave cleaner NMR spectra, which could now be used to further confirm diazirine formation. Hence, only purified functionalized fatty acids were used for the synthesis of muraymycin-based probes. While column purification of diazirine-functionalized

molecules turned out to be rather challenging and often resulted in product decomposition (see below), the crude was commonly used for first test reactions and to explore the general handling of these molecules. An overview of the developed synthetic route to access the diazirine-labeled intermediates is summarized in Scheme 10.3.12.

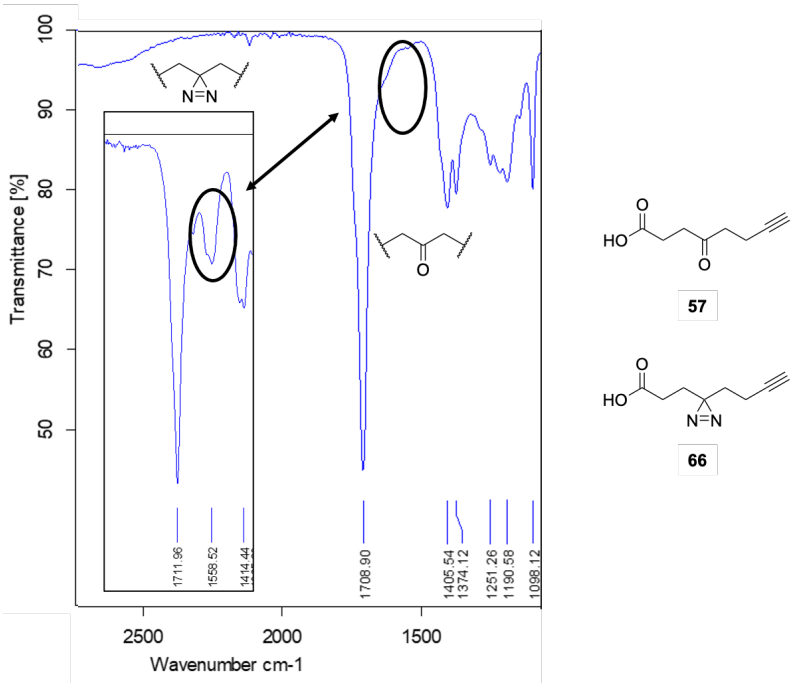


Figure 10.3.2 | Comparison of IR Spectra of Compounds 57 and 66. Excerpts of the IR spectra of compounds **57** and **66** are given. Shown is the region of the carbonyl band, present in both spectra, with the additional diazirine band highlighted in a black circle, which is missing in the reference spectrum of compound **57**.

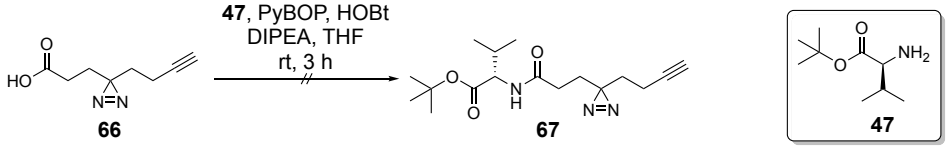
Stability of Diazirine-labeled Fatty Acids

Diazirines are popular photoaffinity warheads for many reasons, one of which is their reasonable stability under common laboratory conditions, including chemical stability within a variety of standard coupling reactions[94,95]. To ensure that the designed functionalized fatty acids could be incorporated into the muraymycin scaffold, they would have to be stable under both amide coupling and strong acidic conditions of the final global deprotection step. While coupling of ketone **35** with the muraymycin core and subsequent diazirine formation after global deprotection could potentially be a feasible strategy, it is likely that the iodine in the diaziridine oxidation step would cause side reactions, e.g. with the uracil ring. Hence, optimization of amide coupling conditions with the diazirine-labeled fatty acid were performed, and treatments of amidated products under global deprotection conditions were monitored. Moreover, other observations regarding the stability of the diazirine-functionalized fatty acids and handling suggestions for this probe are provided.

Even though there are numerous amide coupling reagents commercially available, two methods were most commonly used within our lab as part of the synthetic route towards different muraymycin analogs, namely benzotriazole-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl). Thus, couplings were first attempted using these reagents. For this, diazirine **66** was dissolved in THF. The flask used was either covered or made from

brown glass to avoid unintended decomposition due to UV light exposure of the intermediate. PyBOP, hydroxybenzotriazole (HOBt), and *N*,*N*-diisopropylethylamine (DIEA) were added and the mixture was cooled to 0 °C (Scheme 10.3.13). A valine derivative (**47**) was chosen as a first coupling partner, which was dissolved in THF and slowly added to the reaction mixture. After stirring for about 1.5 hours at 0 °C, the reaction was allowed to warm to room temperature and stirred for another three hours. TLC control showed a shift in the main spot, potentially indicating conversion of the starting materials. However, due to the rather small size of the product (molecular weight; 321 g/mol) and lack of UV activity, LC-MS analysis was difficult to interpret. Preliminary UV irradiation (365 nm) of the TLC plate of the reaction product nicely confirmed decomposition only after irradiation, which indicated that the diazirine motif was still intact. Notably, column chromatography purification of the reaction mixture resulted in complete decomposition of the probe, even though the work-up was performed in the dark, which prevented further analysis of the reaction mixture. While this suggested a higher sensitivity of the probe to diffuse light in the environment of silica gel, we were unfortunately not able to confirm conversion of fatty acid **66** to amide **67**.

Scheme 10.3.13 | Attempted Amide Coupling with Diazirine-functionalized Fatty Acids[a]

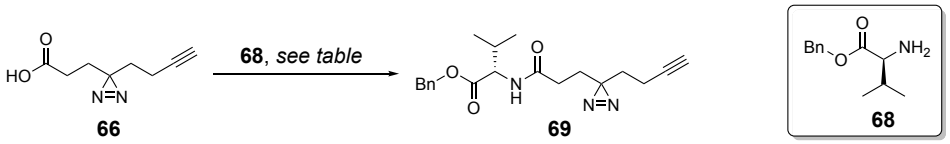


[a] Conversion could not be confirmed. Product decomposed before identity could be verified.

Since the reaction control *via* TLC was not conclusive in the reaction described above and valine derivative **47** as well as fatty acid **66** are not UV-active, bigger, and preferably aromatic coupling partners would be better for these test couplings, which would also allow for reaction monitoring by LC-MS. Hence, a benzyl-protected valine derivative (**68**) was chosen to repeat the PyBOP-mediated amide coupling. However, again, neither TLC, nor LC-MS could confirm conversion of the starting materials. Performing the reaction under inert gas conditions and with dry solvents, as well as allowing for longer reaction periods still failed to give the desired amide product (Table 10.3.3).

Consequently, EDC·HCl coupling conditions were investigated. For this, fatty acid **66**, amine **68**, and 1-hydroxy-7-azabenzotriazole (HOAt) were dissolved in DCM. At room temperature, EDC·HCl was added and the reaction was stirred for several hours. After three hours, product formation was observed by TLC. However, both starting materials were still found in the reaction mixture and overnight incubation led to decomposition of the product. Repetition of this step with a reaction period of about three hours yielded 18% of product **69** over three steps (including diazirine formation, starting from the respective ketone **57**), following column chromatography purification. Complete coverage of the silica column, as well as working in the dark was necessary for this purification step. Successful amide formation could be confirmed by high-resolution MS (HRMS; (electrospray ionization; ESI) calculated mass-to-charge-ratio (m/z calc.) for $C_{20}H_{25}N_3O_3$ $[M+H]^+$ 356.1969; found 356.1990). Consequently, EDC·HCl amide coupling conditions were chosen for the following assembly of the muraymycin analogs (Table 10.3.3).

Table 10.3.3 | Optimization of Amide Coupling Conditions for Diazirine-functionalized Fatty Acids



Attempt	Coupling Reagent	Base	Solvent	t [h]	Dry Conditions	Outcome
1	PyBOP HOBt	TEA	THF	3	no	**no conversion**
2	PyBOP HOBt	TEA	THF	24	yes	**no conversion**
3	EDC·HCl HOAt	-	DCM	24	no	**conversion, but decomposition overnight**
4	EDC·HCl HOAt	-	DCM	3	no	**18% yield over 3 steps**

The resulting product **69** was further treated under highly acidic conditions, as would be needed for the final global deprotection step of the muraymycin core scaffold. For this, the diazirine-containing compound was dissolved in an 80% solution of TFA in water and was stirred at room temperature for several hours. The reaction was monitored by TLC control and no visible decomposition of the probe was noticed even after 24 hours of incubation time, suggesting a high stability of the diazirine motif under acidic conditions. Thus, the diazirine intermediate should be chemically stable during incorporation into the muraymycin scaffold using EDC·HCl as a mediating reagent for the amide coupling reaction as well as during the global deprotection step. Notably, the diazirine compound was also stable while being handled in most ambient lighting conditions, although column chromatography using silica gel should be avoided or performed only in a strictly dark environment and covered column. These findings align with what is commonly reported for the (chemical) stability of diazirines[94]. It was also observed that diazirine-decorated intermediates decomposed after longer storage (several weeks) at -20 °C, suggesting that the fatty acid intermediates, as well as chemical probes, should be prepared for immediate use to ensure optimal biochemical analysis.

10.3.3 Synthetic Endgame

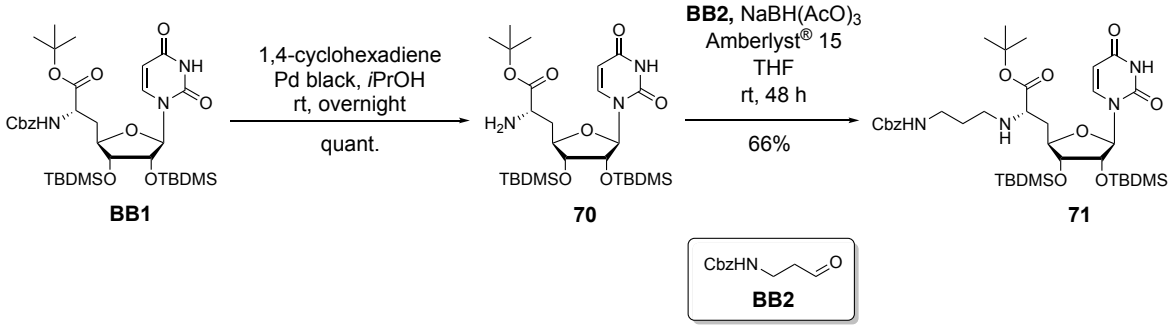
In the final step, the building blocks synthesized in the previous parts of this chapter were assembled to build the complete muraymycin core structure, ending in the attachment of the PAL-functionalized fatty acid (**BB5-PAL**), as well as a respective control fatty acid (**BB5-ref**), which is missing the photoactive warhead.

The assembly started with reaction conditions optimized within our lab, particularly by K. Leyerer and C. Schütz[237]. Starting from the uridine intermediate **BB1**, the primary

amine in 6'-position had to be deprotected. For this, 1,4-cyclohexadiene and Pd black were added to an isopropanol solution of **BB1** and stirred overnight at room temperature to yield complete conversion. The product (**70**) was obtained after filtration of the reaction mixture as a white foam in quantitative yield.

Without further purification, the free amine was transformed in a reductive amination in order to introduce the 3-aminopropyl linker **BB2**. This step had been optimized by A. Spork[211] and proceeded as follows. Free amine **70** was dissolved in dry THF including molecular sieve (pore size = 4 Å), followed by the addition of aldehyde **BB2**. The mixture was stirred at room temperature overnight and the formed imine was subsequently reduced, using sodium triacetoxyborohydride ($NaBH(OAc)_3$) as reducing agent and Amberlyst®15 as a proton source. The reaction was stirred at room temperature for another 24 hours. Ultimately, product **71** was obtained after aqueous work-up and column chromatography as a white foam in 66% yield (Scheme 10.3.14).

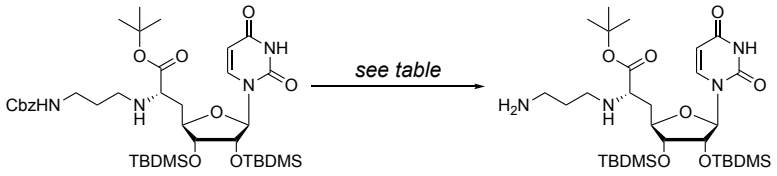
Scheme 10.3.14 | Synthesis of Intermediate 71



Another Cbz-deprotection was performed to deprotect the 3"-amino group of the aminopropyl linker. Similar to what has been described above, nucleosyl intermediate **71** was dissolved in isopropanol, and 1,4-cyclohexadiene and Pd black were added. However, no conversion could be observed. Changing the solvent to MeOH instead led to Cbz deprotection, but also the formation of unidentified side products. Consequently, less harsh conditions, using Pd/C (10 wt.% loading) as a catalyst in MeOH, were chosen. The mixture was stirred at room temperature for two hours and product **72** was yielded after filtration as a white foam with quantitative conversion (Table 10.3.4).

For this particular probe, the aim was to introduce an amino acid with an amino function instead of a hydroxyl group, in place of the naturally occurring hydroxyleucine of muraymycins. Thus, a commercially available L-2,3-diaminopropionic acid derivative (**BB3**), carrying a Cbz-protected and an Fmoc-protected amine in the 2- and 3-positions, respectively, was coupled with nucleosyl intermediate **72**. In a first attempt, PyBOP and HOBt were used as coupling reagents in a manner similar to what has been described above (Scheme 10.3.13). However, no conversion was observed under these conditions. Switching to EDC·HCl and HOAt instead, using conditions described in the previous chapter (10.3.2), afforded good conversion to product **73** with moderate yields of 30% in a small scale (<10 mg) and good yields of 74% in a larger scale (50-100 mg) (Table 10.3.5).

Table 10.3.4 | Cbz-deprotection Optimization to Yield Intermediate 72

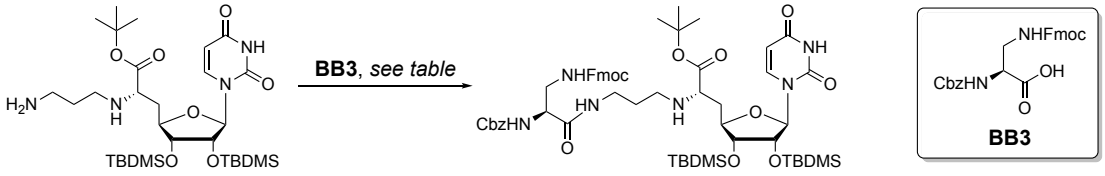


71 72

Attempt	Catalyst	1,4-cyclo-hexadiene	Solvent[a]	Outcome[b]
1	Pd black	10 eq.	*i*PrOH	**no conversion**
2	Pd black	10 eq.	MeOH	**side product formation**
3	Pd/C (10%)	10 eq.	MeOH	**quant. yield**

[a] *i*PrOH = isopropanol. [b] quant. = quantitative.

Table 10.3.5 | Optimization of Amide Coupling Conditions to Yield Intermediate 73



72 73

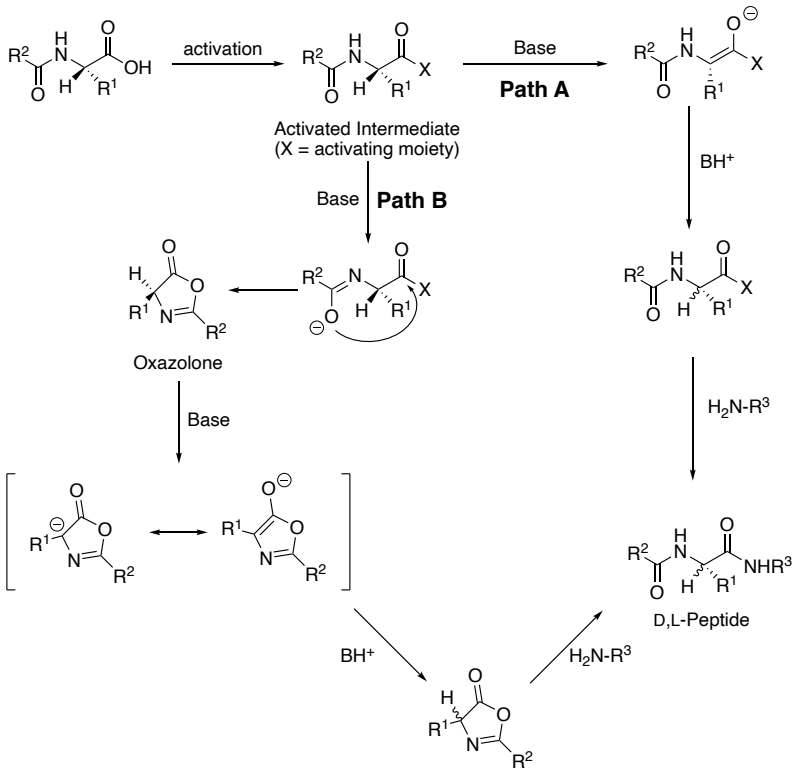
Attempt	Coupling Reagent	Base	Solvent	t [h]	Outcome[a]
1	PyBOP HOBt	TEA	THF	24	**no conversion (small scale)**
2	EDC·HCl HOAt	-	DCM	24	**30% yield (small scale)**
3	EDC·HCl HOAt	-	DCM	24	**74% yield (up-scale)**

[a] Small-scale reactions were carried out with <10 mg starting material. Up-scale reactions were performed with 50-100 mg.

Notably, since epimer formation is a known issue of peptide coupling reactions[238], for this and following amide coupling reactions, configurational integrity of the coupling products was investigated by NMR analysis. In general, racemization during peptide coupling can occur most commonly through oxazolone formation or as a result of the activation of the carboxylic acid moiety[238] (Scheme 10.3.15). In both cases, the carboxylic acid moiety is

activated by formation of a good leaving group with the amide coupling reagent[238]. This in turn increases acidity of the α-proton, which can promote ozaxolone-mediated or immediate racemization of the peptide product[238]. For compound **73**, no epimerization was observed under the described peptide coupling conditions using ^{1}H NMR analysis.

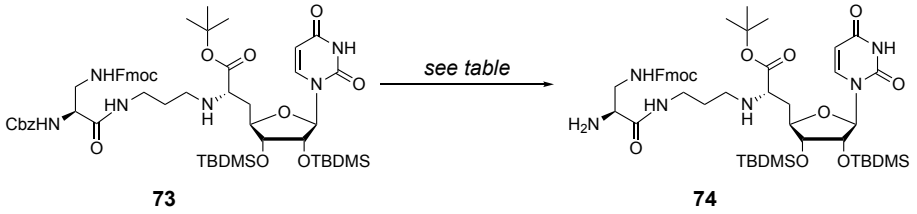
Scheme 10.3.15 | Mechanism of Racemization During Peptide Coupling Reactions[a]



[a] Adapted from Al-Warhi et al., *J. Saudi Chem. Soc.* **2012**, *12*, 97-116[238].

In order to attach the urea dipeptide building block **BB4**, a third Cbz-deprotection had to be performed. In a first attempt, the milder conditions used in the last coupling step were tried, applying 1,4-cyclohexadiene and Pd/C (10 wt.% loading) in MeOH. However, only partial or no conversion to product **74** was observed. Changing the catalyst to Pd black led to almost full conversion in small scale reactions and quantitative yields in larger scale set-ups (Table 10.3.6). For the following amide coupling with **BB4**, conditions using EDC·HCl and HOAt were first applied and led to the desired muraymycin core structure in a good yield of 68% (Scheme 10.3.16). However, conformational integrity of the α-proton of the epicapreomycidine moiety could not be confirmed in this step as the NMR analysis of the fully protected molecule **75** was difficult to interpret. Because of this, the ^{1}H NMR spectra of the deprotected final compound was used to retrospectively determine (and confirm) epimerization within this amide coupling step.

Table 10.3.6 | Optimization of the Cbz-deprotection to Yield Intermediate 74

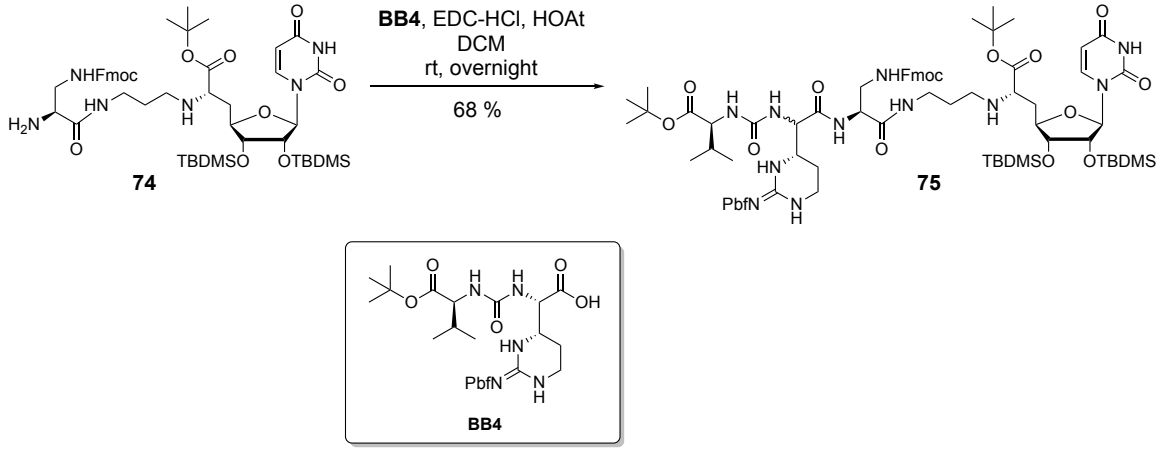


73 74

Attempt	Catalyst	1,4-cyclo-hexadiene	Solvent	Outcome[a]
1	Pd/C (10%)	10 eq.	MeOH	**no conversion (small scale)**
2	Pd black	10 eq.	MeOH	**~50% yield (small scale)**
3	Pd black	10 eq.	MeOH	**quant. yield (up-scale)**

[a] Small-scale reaction are carried out with <10 mg starting material. Up-scale reactions are performed with 50-200 mg.

Scheme 10.3.16 | Synthesis of the Protected 5'-Deoxy Muraymycin Core Structure 75



74 75 BB4

Another critical step in the synthesis of these muraymycin-based A*f*BPs was the cleavage of the Fmoc group at the central propionic acid moiety to allow for amide coupling with the functionalized fatty acids. Fmoc groups are favorable protection groups in solid phase supported peptide synthesis, due to their basic cleavage conditions[239]. Fluoren-9-ylmethoxycarbonyl – or Fmoc – is cleaved forming carbon dioxide and dibenzofulvene upon basic cleavage[240]. The latter reacts with the excess of cyclic amines, e.g. piperidine, which are commonly used bases for Fmoc-deprotection reactions[240] (Scheme 10.3.17). The yielded adducts are often found to be difficult to separate from the desired deprotection product[237]. Thus, in solid phase supported synthesis, this problem is circumvented due to the fixation of the desired deprotected amine product on the solid support that is easily

washed to remove the undesired side products. In this work, solution-based approaches for the removal of the Fmoc-protection group were investigated.

In a first attempt, the protected muraymycin core structure **75** was dissolved in DCM, and piperidine was added at room temperature to reach a 1:3-ratio of piperidine to DCM. After about 30 min, deprotection of the Fmoc-protected amine could be observed *via* TLC control. The reaction was then diluted with chloroform and the organic layer was washed with a saturated sodium carbonate (Na_2CO_3) solution. After purification by column chromatography, product **76** was obtained as a white foam. However, residues of piperidine base or respective dibenzofulvene addition product turned out to be difficult to remove (Table 10.3.7). Though it was not clear whether the side product observed on TLC controls was the free base or the piperidine adduct of dibenzofulvene, impurities would disturb the subsequent amide coupling with the fatty acids. Because of this, the Fmoc deprotection reactions and following amide couplings are discussed together.

Scheme 10.3.17 | Proposed Dibenzofulvene-Piperidine Adduct Formation

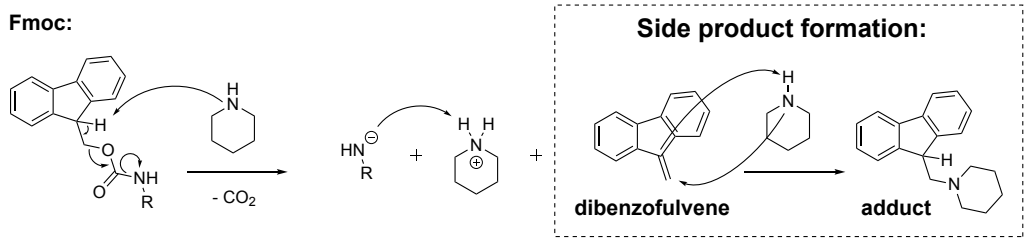
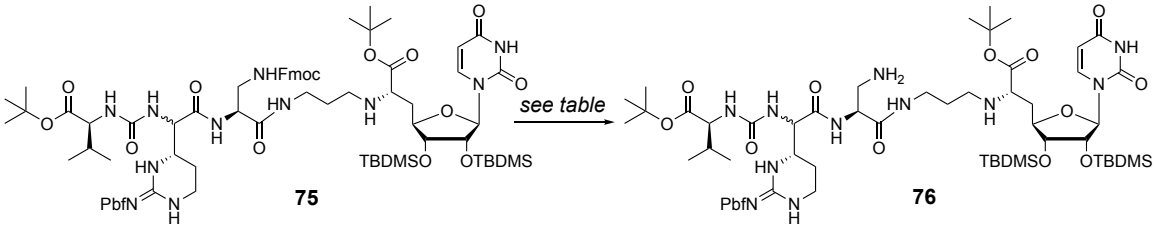


Table 10.3.7 | Optimization of the Fmoc-deprotection Step to Yield Intermediate 76



Attempt	Base	c_{base} [%]	Solvent	t [h]	Outcome
1	piperidine	25	DCM	1.5	**38% yield (side product tedious to remove)**
2	piperidine	10	DCM	1.5	**23% yield (side product tedious to remove)**
3	DIEA	50	DCM	72	**no conversion**
4	diethylamine	50	DCM	3	**mainly side product formation**
5	poymer-bound piperazine	(1.0-2.0 mmol/g)	DCM	24	**no conversion**

Attempts to remove excess piperidine by evaporation gave mixed results. Since repeated co-evaporation with toluene or subsequent co-evaporation with acetonitrile and DCM did not yield the desired removal of piperidine or side product, resulting in failure of the following amide couplings, lyophilization of the sample was tested. After two successive lyophilization efforts and column chromatography purification, product **76** was obtained in 49% yield. Subsequent amide coupling, using EDC·HCl as described in previous sections, worked well and the resulting amide (**77**) was obtained with 59% yield (Scheme 10.3.18).

In an attempt to further optimize the reaction conditions and simplify the work-up, less piperidine (10%) was used, but this did not lead to an improved yield. Exchanging the base to a sterically hindered base, for example DIEA in DCM (1:1-ratio), was thought to help avoiding adduct formation with the dibenzofulvene side product. However, treatments with DIEA or diethylamine (DEA) were not able to cleave the Fmoc group. Lastly, another approach related to the solid phase synthesis techniques was investigated. Here, polymer-bound base was envisioned to cleave the protection group, while potential side products formed with the base should be easily separable by filtration. This would be in contrast to the classical solid phase peptide synthesis approach, in which the product is fixed on the solid phase instead. Yet, this method also failed to furnish Fmoc deprotection (Table 10.3.7).

Going back to piperidine-based deprotection conditions (Attempt 1, Table 10.3.7), simplifications of the general work-up were investigated to help improve yields (Table 10.3.8). The aim was to optimize the workflow of this step, since these conditions showed the best efficacy in Fmoc-cleavage. Skipping the aqueous work-up led to problems with the following column chromatography purification step, due to the huge amount of base left in the organic layer. Likewise, not performing a final column chromatography purification step, even after thorough drying of the crude mixture, also led to problems with subsequent amide coupling reactions. Hence, only the combination of aqueous work-up, thorough drying steps, and chromatographic purification yielded product **76** in sufficient purity for follow-up reactions. Notably, extra lyophilization steps were not superior to simple successive co-evaporation with toluene, acetonitrile, and DCM prior to column chromatography purification, which resulted in a respectable yield of 67% of pure product **76**, and no issues for following amide coupling reactions. Surprisingly, the extra lyophilization step actually led to more loss of compound, and thus to lower overall yields of 49%. Thus, conditions of Attempt 4 (Table 10.3.8) were applied for all synthesized probes, and subsequent amide coupling reactions with all fatty acids, including diazirine labeled analogs, worked well.

As a first final muraymycin analog, the reference compound for the chemical probe was synthesized, introducing a fatty acid with an ω-alkyne function, but missing the photoactive diazirine warhead of the photoactive probe. Here, the previously synthesized dec-9-ynoic acid (**BB5-ref**) was coupled with the muraymycin core structure, resulting in final intermediate **77** in 59% yield. That reaction followed the global deprotection of the acid-cleavable protecting groups of the two carboxylic acids, the epicapreomycidine, as well as the ribose unit. Final compound **79** was obtained as a white solid after preparative HPLC in 36% yield. This final compound is the direct analog of the desired diazirine-functionalized probe. Hence, it poses an essential compound to determine biochemical activity of the designed A*f*BP, since the diazirine-carrying molecule is likely not stable in the conditions used for the fluorescence-based MraY activity assay. Notably, final NMR

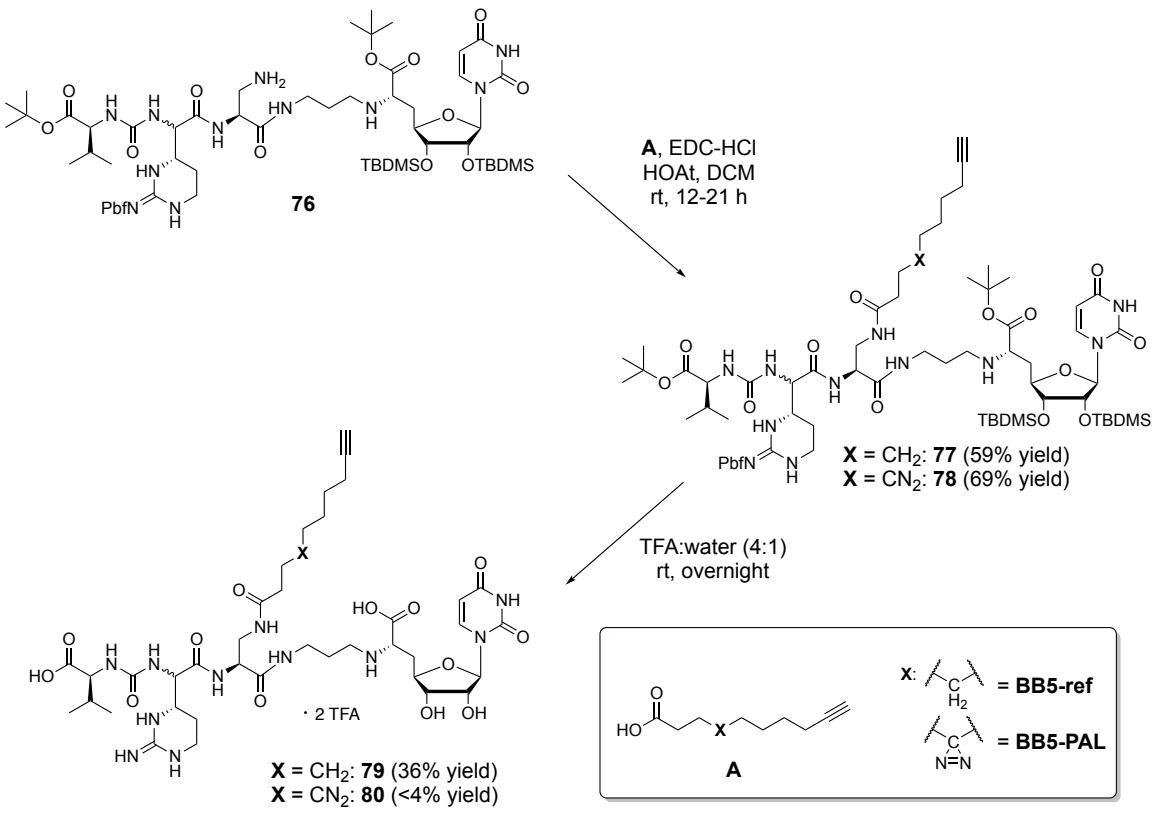
analysis revealed that compound **79** was obtained as a diastereomeric mixture (*d.r.* ~ 1.2:1), indicating that the amide coupling reaction introducing the epicapreomycidine moiety must have caused epimerization at the α-position of the epicapreomycidine, as already indicated above.

Table 10.3.8 | Optimization of Work-up Conditions for the Fmoc-deprotection Step[a]

Attempt	1st Step	2nd Step	3rd Step	4th Step	5th Step	Outcome
1	drying *in vacuo*	co-evaporation (toluene – ACN – DCM)	column chrom.	-	-	**No separation**
2	aqueous work-up	drying *in vacuo*	co-evaporation (toluene – ACN – DCM)	-	-	**Subsequent amide coupling failed**
3	aqueous work-up	drying *in vacuo*	co-evaporation (toluene – ACN – DCM)	freeze drying (2x)	column chrom.	**49% yield (pure)**
4	drying *in vacuo*	co-evaporation (toluene – ACN – DCM)	aqueous work-up	drying *in vacuo*	column chrom.	**67% yield (pure)**

[a] chrom., chromatography.

Scheme 10.3.18 | Synthesis of Final Probes 79 and 80



Finally, a first coupling with a diazirine-labeled fatty acid was performed. For this, freshly prepared diazirine fatty acid **BB5-PAL** was coupled with the muraymycin core structure (**76**) using brown glass ware. After column chromatography purification, product **78** was obtained in good yield of 69%. However, after global deprotection and HPLC purification, decomposition was observed, and amounts of isolated product **80** were insufficient for further analysis (<1 mg; <4% yield).

In summary, the envisioned strategy of building the muraymycin analogs (**79** and **80**) from the discussed building blocks was a good start for the development of chemical probes for these nucleoside antibiotics. Most of the steps are robust amino deprotection and amide coupling steps with decent yields. However, Fmoc deprotection in solution caused issues for work-up and follow-up reactions and different amide coupling conditions to introduce the epicapreomycidine moiety are needed to avoid epimerization in this step. Even though a diazirine-labeled probe was not obtained in sufficient quantities for follow-up biological investigations, the presented synthetic route showed that photoactive warheads are sufficiently stable under discussed conditions and gave insights into their handling. Nevertheless, further optimization of critical steps such as the global deprotection and HPLC purification of the diazirine-labeled probe would be essential.

10.3.4 Biological Evaluation

In vitro activity of the synthesized probe was determined in a fluorescence-based activity assay of MraY obtained from *S. aureus* – developed and previously reported by our lab[241]. Here, dansylated Park's nucleotide (fluorescent) that reacts with MraY to form dansylated lipid I exhibits an increase in fluorescence[241]. That is, an increase in fluorescence is a measure of formed lipid I and therefore MraY enzymatic activity[241]. As mentioned, a reference probe (**79**) had to be synthesized since the described assay uses excitation wavelengths of 355 nm, which would lead to decomposition of the photoactive chemical probe. As a *proof of concept*, reference probe **79** was tested and exhibited an IC_{50} value of 0.99 ± 0.14 µM (Figure 10.3.3). This is comparable to other 5'-deoxy analogs tested in this assay[206,213], thus validating that the designed muraymycin analog retains reasonably potent MraY inhibitor activity.

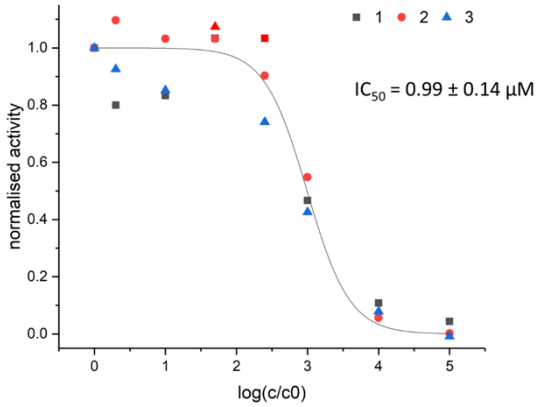


Figure 10.3.3 | Biochemical IC_{50} Determination of Reference Probe 79. Reference probe **79** was incubated with MraY obtained from *S. aureus* (as crude membranes) and conversion of fluorescent-labeled substrate was analyzed in a dose-dependent manner. Shown are the results of three independent experiments. c_0 = 1 nM.

10.4 Conclusion

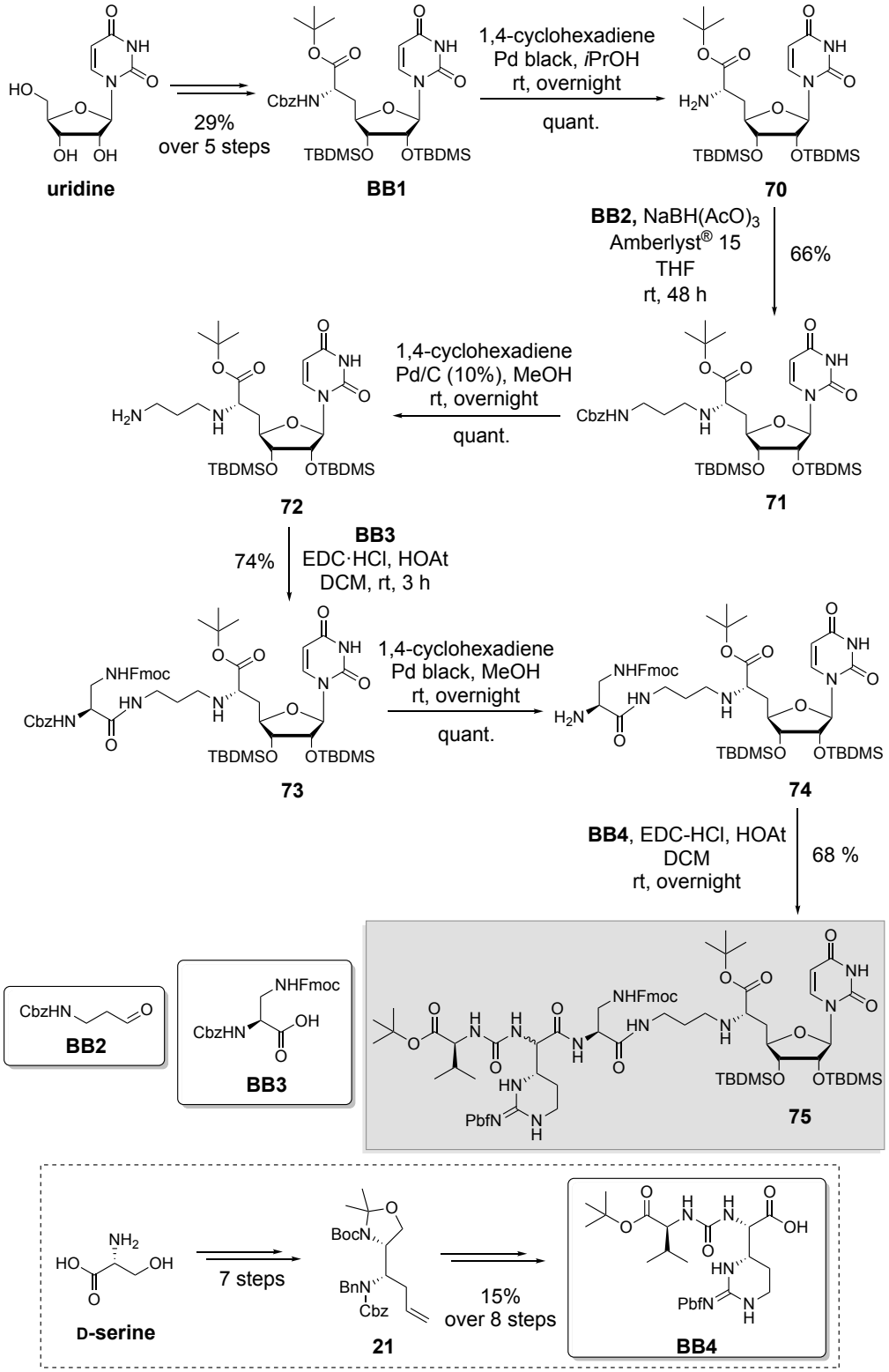
In this project, the goal was to develop a photoactive chemical probe for studies on a novel class of nucleoside-based antibiotics – the muraymycins. Whilst the muraymycins are known MraY inhibitors, not much is known about their potential off-targets within the human or bacterial cell. We proposed to address this by developing ABPP strategies, incorporating reactive groups into the muraymycin core scaffold. For this, fatty acid moieties, naturally occurring in muraymycins of the A- and B-Groups, were functionalized with a photoactive diazirine warhead as well as an alkyne motif for subsequent tag attachment. This new motif had to be merged with the muraymycin scaffold and respective synthetic efforts were explored to establish a successful synthetic route to these new derivatives. To provide further *proof of concept*, a reference compound without the photoactive warhead was synthesized, which was needed to determine biochemical activity of the probe in comparison to other analogs of this naturally occurring scaffold while also being much easily chemically accessible than the respective diazirine-decorated probe.

Synthesizing the muraymycin core structure (Scheme 10.4.1) required the synthesis of two main building blocks in the course of this project, namely, the uridine core **BB1** with a free amine for linker (**BB2**) attachment, as well as the urea dipeptide moiety **BB4** including a free carboxylic acid for coupling with the central amino acid **BB3**, which would ultimately serve as the hinge to connect the fatty acids (**BB5-PAL/ref**) to the main muraymycin core structure. The general synthetic routes for these two building blocks were developed and refined by several people within the lab, including A. P. Spork, K. Leyerer, and others[194,206,211,237,242]. The nucleosyl intermediate **BB1** was synthesized starting from uridine and was obtained over five steps and an overall yield of 29% from uridine. The urea dipeptide intermediate (**BB4**) was built from key intermediate **21**, synthesized over seven steps from D-serine, and was obtained after another eight steps with an overall yield of 18% starting from **21**.

The subsequent assembly of the nucleoside inhibitor core structure (Scheme 10.4.1) began with a Cbz-deprotection and reductive amination of the primary amine in the nucleosyl unit (**BB1**) with aldehyde **BB2**, which was a key intermediate available in stock within our lab. Intermediate **71** was obtained in 66% yield. Subsequent Cbz-deprotection and introduction of the essential hinge amino acid **BB3** led to intermediate **73** in 74% yield. This amino acid incorporates two amino groups, which can be ultimately used to connect to the urea dipeptide moiety **BB4** on one side and the (functionalized) fatty acids (**BB5-PAL/ref**) on the other. To do that, a special protection group pattern had to be chosen. A commercially available amino acid (**BB3**), carrying a Cbz protection group for the primary amine in the 2-position as well as an Fmoc group on the primary amine in the 3-position, was therefore selected. This protection group strategy posed an excellent choice since the Cbz group is cleaved under reductive conditions, while the Fmoc group requires strong basic conditions. Acidically cleavable protection groups could not have been used for this moiety, since those conditions would also lead to deprotection of protection groups on the nucleosyl and urea dipeptide moieties. **BB3** was hence coupled with the 3-aminopropane linker of intermediate **72** using EDC·HCl and HOAt as mediating reagents, which showed to be superior to PyBOP and HOBt. **73** was obtained in good yields of 74% over two steps from **71**. Another sequence of Cbz-deprotection and EDC·HCl/HOAt-mediated amide

coupling with **BB4** finally led to the complete muraymycin core structure **75** in 68% yield over two steps, or 33% overall yield starting from the nucleosyl building block **BB1**.

Scheme 10.4.1 | Synthesis of the Protected Muraymycin Scaffold 75



While the ω-alkyne fatty acid **BB5-ref** could be synthesized in a one-step TEMPO/BAIB-oxidation reaction from a respective alcohol starting material (**64**) (Scheme 10.4.2), the corresponding analog carrying a ketone for diazirine synthesis required more effort. To access this key intermediate (**41**), a first strategy was to react an aldehyde with a Grignard reagent forming a hydroxy group at the coupling site, which could be subsequently oxidized to the respective ketone. The aldehyde moiety was synthesized *via* ring opening with MeOH from γ-butyrolactone, followed by IBX oxidation with an overall yield of 82%. The Grignard reagent was built from alcohols of different lengths (**57** and **45** for the shorter and longer alkyl chain, respectively). Even though longer alkyl chains were generally more desirable due to the resemblance with naturally occurring fatty acids in muraymycin antibiotics, shorter chains were more affordable and served as test compounds for optimization of this synthetic route (therefore not included in schemes within the Conclusion chapter).

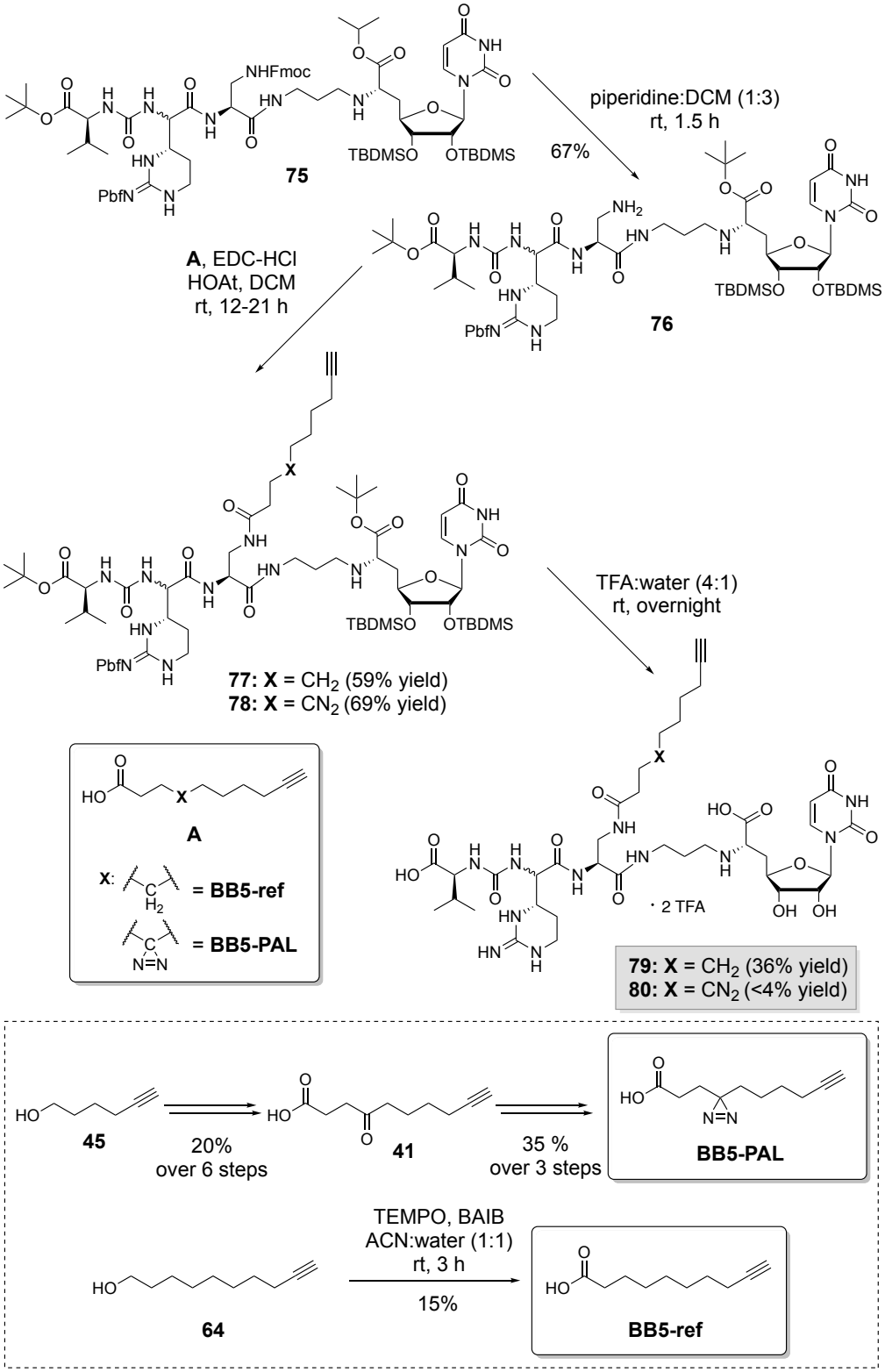
To form the desired Grignard reagents, the primary hydroxyl groups of long chain alcohols were transformed into the respective bromides over three steps with 76% and quantitative yields for the shorter (**58**) and longer (**64**) alkyl chains, respectively. Initial Grignard reactions failed due to old magnesia turnings and a too dilute solution of the bromide reactant. However, even after successful Grignard formation, attempts to couple intermediate **58** with aldehyde **44** led to decomposition and side product formation, while no product could be obtained (not shown). Thorough literature search suggested a different, more elegant synthetic route, using iron (III)-catalyzed conversions of Grignard reagents with acid chlorides, leading to the direct formation of the desired ketone function at the coupling site. Acid chloride **60** was commercially available and reaction with freshly prepared Grignard reagents led to successful formation of the ketone-functionalized fatty acids in moderate yields of 52% and 56% for the short (**62**) and long (**65**) fatty acids, respectively. Final basic deprotection of the TMS-alkyne and the methyl ester yielded key intermediates **63** and **41** in 87% and 57% yield, respectively; meaning an overall yield of 34% for the C_8-based fatty acid **63** and 20% for the C_{10}-based fatty acid **41** (simplified route shown in Scheme 10.4.2; Note: shorter length linker is not included in this scheme).

Generally, diazirines are synthesized from ketones *via* a diaziridine intermediate. In a first step, the ketone was treated with ammonia in MeOH at 0 °C and subsequent addition of HAOSA led to formation of diaziridine **71**. However, it turned out that intermediate **71** was highly unstable and could decompose even during work-up steps. Hence, different work-up techniques were compared, and it was found that filtration, followed by a mere concentration of the reaction mixture under addition of triethylamine as a base and immediate continuation with the subsequent oxidation step, provided optimal conditions to yield the desired diazirine compounds.

In a final step, the diaziridines were oxidized using an iodine solution in MeOH. Including this step, further reactions were carried out in the dark. It was observed that, although the diazirines were fairly stable under common laboratory lighting conditions, column chromatography had to be performed under strict coverage of the column to prevent decomposition of the product in the presence of the silica gel. Octanoic acid **72** and decanoic acid **BB5-PAL** were obtained from their ketone intermediates in overall yields of 40% and 35% over three steps (simplified route shown in Scheme 10.4.2; Note: shorter length linker is not included in this scheme), respectively. Notably, first amide test couplings confirmed stability of the diazirine motif within PyBOP/HOBt and

EDC·HCl/HOAt amide coupling conditions, whereas the latter gave the best yields for the conversion of the employed fatty acids with test coupling partners.

Scheme 10.4.2 | Synthesis of Final Probes 79 and 80



In the last steps of the synthesis of muraymycin-based chemical probes (Scheme 10.4.2), the Fmoc group on the central core structure of muraymycin intermediate **75** had to be cleaved, to allow for the amide coupling with the fatty acids. However, the Fmoc deprotection step turned out to be more troublesome than initially anticipated. A lot of material (**75**) was lost in this step, since separation of the dibenzofulvene-piperidine adduct or piperidine base from product **76** was challenging. Considering the time- and labor-intense synthesis of all building blocks up to that point, a different synthetic procedure here would be desirable. Attempts to use sterically hindered bases, such as DIEA or diethylamine as well as solid phase-supported bases did not lead to any Fmoc-deprotection, leaving piperidine-mediated deprotection conditions as the most efficient. Thus, respective work-up conditions were optimized in an attempt to increase overall yields of this step, resulting in a combination of co-evaporation steps after aqueous work-up and subsequent column chromatography purification as the best option with a yield of 67%.

The resulting free amine **76** was condensed with either diazirine labeled (**BB5-PAL**) or unlabeled (**BB5-ref**) fatty acids, using EDC·HCl/HOAt conditions, in acceptable yields of 69 and 59%, respectively. In a final step, intermediates **77** and **78** were treated with 80% TFA in water for a global cleavage of several protecting groups, including *tert*-butyl esters at the nucleosyl unit and the urea dipeptide valine, the guanidine protecting group Pbf at the epicapreomycidine moiety, as well as the TBDMS groups at the ribose. Subsequent purification by preparative HPLC yielded the final reference compound **79** as a TFA-salt in a yield of 25% from the muraymycin intermediate **76** (Scheme 10.4.2). Unfortunately, NMR analysis revealed at this stage that compound **79** was obtained as a mixture of diastereomers (*d.r.* ~ 1.2:1). Epimerization is a known issue of amide coupling reactions[238] and must had occurred when **BB4** was coupled with **74**, as there was no epimerization observed by NMR analysis for intermediate **73** and following amide coupling reactions could not have led to diastereomer formation at the α-proton of the epicapreomycidine amino acid. Hence, further optimization by for example investigation of alternative amide coupling reagents for this step, introducing the epicapreomycidine moiety, is necessary to obtain stereochemically pure final products.

Unfortunately, the diazirine-labeled probe could not be obtained in sufficient quantity. Hence, optimization especially of the Fmoc deprotection step would be helpful in increasing the overall yield of the probe *via* this synthetic route. While this work was being performed, K. Leyerer in our lab developed solid phase-supported synthetic strategies to assemble the muraymycin scaffold[242]. This approach could be applied to synthesize this probe in higher quantities, as it would specifically help with some steps, such as the Fmoc deprotection. Thus, optimization and simplification of the synthetic route – especially the last steps, introducing the fatty acid moiety – would be essential for the application of this photoactive probe, as high quantities of probe **80** would be needed for further proteomic analysis, while stability issues with diazirine probes during storage also require the probe to be synthesized quickly when needed for an experiment.

Finally, unlabeled final compound **79** exhibited reasonably potent MraY inhibition *in vitro* with an IC_{50} value of 0.99 μM for MraY obtained from *S. aureus* – even though as a mixture of diastereomers –, providing a first *proof of concept* showing that a respective probe would probably mimic muraymycin activity.

10.5 Experimental

10.5.1 General Methods

Synthetic Procedures
Synthetic steps presented within this project were generally carried out under inert gas atmosphere, especially air- and moisture-sensitive reactions. Before use, glass equipment was heated *in vacuo*, using a heat gun, and filled with nitrogen gas upon cooling – repeated three times. Nitrogen gas was passed through a column of phosphorus pentoxide for further removal of moisture as well as silica gel, a good indicator for when to refresh this set-up due to changes in color depending on moisture content of the gel. Reactions carried out below room temperature were cooled in ice water baths or acetone baths cooled with liquid nitrogen, depending on the desired temperature. All reactions, work-ups, or other handling of diazirine-containing compounds were performed using brown glass ware or an appropriate cover to avoid light-induced decomposition.

Reagents/Chemicals
Reagents and starting materials were purchased from *Acros Organics*, *AK Scientific*, *Alfa Aesar*, *Carbolution*, *Fluka Analytical*, *GL Biochem*, *Grüssing*, *Riedel-de Haën*, *Sigma-Aldrich*, *TCI*, and *VWR*, or from a respective facility within Saarland University (Zentrales Chemikalienlager). Purchased chemicals were used without further purification. Notably, aldehyde **BB2** was synthesized by G. Niro and intermediate **21** was synthesized by K. Leyerer within our lab.

Solvents
Most solvents were purchased in technical quality and used without further purification. Deionized water was available within the lab set-up. Certain solvents, such as DCM, Et_2O, EtOAc, and PE, were distilled in our lab before use. For anhydrous conditions, a variety of solvents were also degassed and stored over molecular sieves or dried and purified using a solvent purification system (MB SPS 800, *MBRAUN*). Details are listed in Table 10.5.1.

Chromatographic Purifications
As part of purification processes for intermediate compounds, normal phase column chromatography was method of choice. Columns were packed using Si60 silica gel purchased from *VWR* (pore diameter = 40-63 μm).

Thin Layer Chromatography (TLC)
Reaction controls were usually performed using Si60 F_{254}-coated aluminum plates from *VWR*. TLC plates were analyzed by UV light (254 nm) or staining with appropriate reagents (Table 10.5.2).

Table 10.5.1 | Solvents

Solvent	Purification	Anhydrous
acetone	No	Degassed
acetonitrile	No	SPS
DCM	Distilled	SPS
DMF	No	SPS
DMPU	No	Purchased
Et_2O	Distilled	SPS
EtOAc	Distilled	Pre-dried over K_2CO_3; distilled
n-hexane	No	No
*i*PrOH	No	Pre-dried over K_2CO_3; degassed
MeOH	No	Degassed
PE	Distilled	No
pyridine	No	Pre-dried over CaH_2; distilled
tert-BuOH	No	Degassed
TEA	No	Degassed
THF	No	SPS
toluene	No	Purchased

Table 10.5.2 | TLC Analysis

Stain	Composition[a]
ninhydrin	0.3 g ninhydrin, 3 mL acetic acid, 100 mL 1-butanol
vanillin/sulfuric acid	4 g vanillin, 25 mL $H_2SO_{4\ conc.}$, 80 mL acetic acid, 680 mL MeOH
potassium permanganate	1 g $KMnO_4$, 6 g K_2CO_3, 1.5 mL 5 M NaOH, 100 mL water

[a] conc. = concentrated.

Liquid Chromatography – Mass Spectrometry Analysis (LC-MS)
LC-MS analysis was used as a reaction control and to confirm product mass formation. For this, a *Thermo Scientific* Spectra (TSP) System (*Finnegan*) was used, including an AS3000 auto sampler, a P4000 pump, an SN 4000 controller, an SCM 1000 mixer, a Nucleodur® 100-5 C_{18} column (3 x 125 mm, 5 µm), a UV2000 UV-detector, and an MSQ Plus mass spectrometer (using ESI). UV/Vis detection for 254 nm was recorded for a total runtime of 12 or 25 min at a flowrate of 0.8 mL/min of a water (eluent A)/acetonitrile (eluent B) mixture (0.1% TFA), or a water (0.1% TFA) (eluent A)/MeOH (eluent B) mixture. Used methods are summarized in Table 10.5.3.

Table 10.5.3 | LC-MS Methods

Method	Eluents A/B	Flow rate [mL/min]	Eluent B [%]				
LC-MS analytical 1	water (0.1% TFA)/MeOH	0.8	0 min	6 min	9 min	10.5 min	12 min
			5	100	100	5	5
LC-MS analytical 2	water/acetonitrile (both 0.1% TFA)	0.8	0 min	7 min	8.3 min	9.5 min	12 min
			5	100	100	5	5
LC-MS analytical 3	water (0.1% TFA)/MeOH	0.8	0 min	15 min	20 min	20.1 min	25 min
			5	100	100	5	5
LC-MS analytical 4	water/acetonitrile (both 0.1% TFA)	0.8	0 min	19 min	22 min	23 min	25 min
			0	100	100	5	5

High-performance Liquid Chromatography (HPLC)
Final compounds were purified on an Agilent Technologies 1200 Series, including a LichroCart® Purospher® RP18e column (10 x 250 mm, 5 µm, *VWR*). For this, lyophilized crude reaction products were dissolved in bi-distilled water, acetonitrile, or MeOH (HPLC grade quality) and a few drops of TFA and injected after filtration. UV/Vis detection for 254 nm, 280 nm, and 360 nm was recorded for a total runtime of 35 min at a flow rate of 3 mL/min of a water (eluent A)/acetonitrile (eluent B) mixture (0.1% TFA). Table 10.5.4 gives an overview of the used method. Reported retention times (t_R) [min] of detected product peaks are not corrected for initial column dead times (~2 min).

Table 10.5.4 | HPLC Method

Method	Eluents A/B	Flow rate [mL/min]	Eluent B [%]				
HPLC 1	water/acetonitrile both (0.1% TFA)	3.0	0 min	19 min	25 min	30 min	35 min
			5	100	100	0	0

10.5.2 Analytics

Mass Spectrometry (MS)
LC-MS analysis has already been described above to yield low resolution mass data for compound identification purposes. HRMS data was kindly determined by S. Boettcher in our lab using a *Thermo Scientific* Ultimate 3000 system, including a *Dionex* UltiMate 3000 ultra-high-performance liquid chromatography (UHPLC) system. The latter is a complex of an autosampler, pump, *Thermo* Accucore phenyl-X column (3 x 100 mm, 2.1 µm), diode array detector, and a Q Exactive OrbiTrap (*Thermo Scientific*).

Nuclear Magnetic Resonance (NMR) Analytics
Main tool for the identification of synthesized compounds was ^{1}H and ^{13}C NMR analysis (at 500 MHz and 126 MHz, respectively). For these, *Bruker* NMR spectrometers, including Avance I 500 (B ACS 60 autosampler), Avance DRX 500, and Avance III 500 with a TCI cryo probe head were used. Reported are chemical shifts δ [ppm] (parts per million) in relation to the respective solvents used as an internal standard. Spin multiplicities are indicated as s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), dt (doublet of triplets), tt (triplet of triplets), m (multiplet), and reported with the respective coupling constants *J* [Hz]. For correct peak assignments, $^{1}H,^{1}H$-COSY and $^{1}H,^{13}C$-HSQC/HMBC analysis were used.

Infrared (IR) Spectroscopy
IR spectra were recorded on a *Bruker* Fourier-transform infrared (FTIR) spectrometer ALPHA including an integrated PlatiniumATR® unit, and band wavelengths [cm^{-1}] are reported for supported compound identification purposes.

UV/Vis Spectroscopy
UV/Vis spectra were recorded on an *Agilent Technologies* Cary series 100 UV spectrometer from 200-800 nm and peaks of maximal absorption (λ_{max}) [nm] are reported. Compound solutions in MeOH were analyzed.

10.5.3 General Procedures

Cbz-deprotections
Method A: Protected amine (1.0 eq.) was dissolved in dry isopropanol under nitrogen atmosphere. 1,4-Cyclohexadiene (10.0 eq.), as well as Pd black (a small spatula tip) were added and the mixture was stirred at room temperature for 2-24 hours. After completion of the reaction (TLC control), the reaction mixture was filtered through a syringe filter and washed several times with isopropanol. The solvent was removed *in vacuo* to yield the product in sufficient purity for further reactions.

Method B: Protected amine (1.0 eq.) was dissolved in dry MeOH under nitrogen atmosphere. 1,4-Cyclohexadiene (10.0 eq.) as well as Pd/C (10% loading; a small spatula tip) were added and the mixture was stirred at room temperature for 2-24 hours. After completion of the reaction (TLC control), the reaction mixture was filtered through a syringe filter and washed several times with MeOH. The solvent was removed *in vacuo* to yield the product in sufficient purity for further reactions.

Method C: Protected amine (1.0 eq.) was dissolved in dry MeOH under nitrogen atmosphere. 1,4-Cyclohexadiene (10.0 eq.) as well as Pd black (a small spatula tip) were added and the mixture was stirred at room temperature for 1-24 hours. After completion of the reaction (TLC control), the reaction mixture was filtered through a syringe filter and washed several times with MeOH. The solvent was removed *in vacuo* to yield the product in sufficient purity for further reactions.

Amide Couplings
Method D: The carboxylic acid (1.1 eq.) was dissolved in dry THF under nitrogen atmosphere. HOBt (1.0 eq.), DIEA (1.8 eq.), and PyBOP (1.1 eq.) were added to the solution and stirred at room temperature for about 25 min. The mixture was

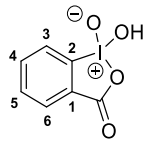
cooled to 0 °C and a solution of amine (1.0 eq.) in dry THF was added. The reaction was further stirred at this temperature for 1.5 hours, warmed to room temperature and stirred for another 3-24 hours.

Method E: The carboxylic acid (1.15 eq.), the amine (1.0 eq.), and HOAt (0.49 eq.) were dissolved in dry DCM under nitrogen atmosphere. EDC·HCl (1.95 eq.) was added and the mixture was stirred at room temperature overnight. After completion of the reaction (TLC control), the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography to yield the pure product.

10.5.4 Synthesis

Synthesis of Reagents

2-Iodoxybenzoic acid (IBX)

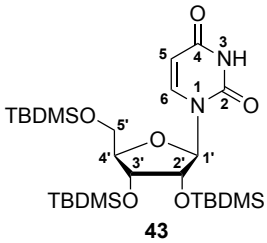


IBX

OXONE® (186 g, 303 mmol, 3.0 eq.) was dissolved in deionized water (1.0 L). 2-Iodobenzoic acid (**42**) (25.0 g, 101 mmol, 1.0 eq.) was added to the solution and the mixture was stirred at 70 °C for 5 hours. Upon completion of the reaction, indicated by the transformation of the suspension to a clear solution, the mixture was cooled to 0 °C and was stirred at this temperature for at least 30 min. The resulting precipitate was filtered off using a medium porosity sintered glass funnel and washed several times with ice cold water and acetone. The product (**IBX**) was dried for several days and could be obtained as a white powder (13.8 g, 49.3 mmol; 49% yield). **^{1}H NMR** (500 MHz, DMSO-d_6) δ 7.84 (dd, *J* = 7.5, 0.8 Hz, 1 H, 5-H), 8.00 (dd, *J* = 7.7, 1.3 Hz, 1 H, 4-H), 8.03 (dd, *J* = 7.4, 1.2 Hz, 1 H, 6-H), 8.15 (d, *J* = 7.8 Hz, 1 H, 3-H); **^{13}C NMR** (126 MHz, DMSO-d_6)) δ 124.97 (C-5), 130.05 (C-4), 131.41 (C-6), 132.92 (C-1), 133.36 (C-2), 146.53 (C-3), 167.45 (C=O).

*Synthesis of the Uridine Building Block (**BB1**)*

2',3',5'-*O*-Tris-(*tert*-butyldimethylsilyl) uridine (43)

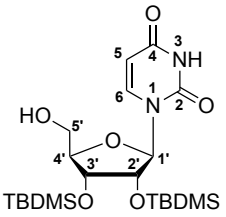


43

After co-evaporation with pyridine (2 x 60 mL), uridine (10.0 g, 41.0 mmol, 1.0 eq.) was dissolved in dry pyridine (120 mL) under nitrogen atmosphere. Imidazole (12.6 g, 185 mmol, 4.5 eq.) and TBDMSCl (27.8 g, 185 mmol, 4.5 eq.) were added and the reaction

mixture was stirred at room temperature for 6 days. Following completion of the reaction (TLC control), the mixture was cooled to 0 °C and water (15 mL) was added. The solvent was removed *in vacuo* and the crude was re-dissolved in EtOAc (1.0 L). The organic layer was washed with water (3 x 480 mL) and brine (150 mL), dried over sodium sulfate (Na_2SO_4), filtered and dried under reduced pressure. The resulting crude product was further purified by flash column chromatography (400 g; PE:EtOAc, 4:1) to give the product (**43**) as a white foam with minor impurities (26.4 g; 100% = 24.1 g). $\mathbf{R}_f$ (TLC) = 0.18 (PE:EtOAc, 4:1); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.06 (s, 3 H, $SiCH_3$), 0.07 (s, 3 H, $SiCH_3$), 0.08 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.11 (s, 3 H, $SiCH_3$), 0.12 (s, 3 H, $SiCH_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 0.90 (s, 9 H, $SiC(CH_3)_3$), 0.94 (s, 9 H, $SiC(CH_3)_3$), 3.75 (dd, J = 11.6, 1.2 Hz, 1 H, 5'-H_a), 3.98 (dd, J = 11.6, 1.9 Hz, 1 H, 5'-H_b), 4.05-4.09 (m, 3 H, 2'-H, 3'-H, 4'-H), 5.68 (dd, J = 8.1, 2.2 Hz, 1 H, 5-H), 5.86 (d, J = 3.6 Hz, 1 H, 1'-H), 8.02 (d, J = 8.2 Hz, 1 H, 6-H), 8.75 (s, 1 H, NH); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -5.45 ($SiCH_3$), -5.25 ($SiCH_3$), -4.72 ($SiCH_3$), -4.66 ($SiCH_3$), -4.44 ($SiCH_3$), -4.09 ($SiCH_3$), 18.11 ($SiC(CH_3)_3$), 18.22 ($SiC(CH_3)_3$), 18.62 ($SiC(CH_3)_3$), 25.79 ($SiC(CH_3)_3$), 25.89 ($SiC(CH_3)_3$), 25.97 ($SiC(CH_3)_3$), 61.96 (C-5'), 70.93 (C-3'), 76.29 (C-2'), 84.69 (C-4'), 89.04 (C-1'), 102.01 (C-5), 140.54 (C-6), 150.30 (C-2), 163.34 (C-4); **IR** (ATR) $\tilde{\nu}$ = 2953, 2929, 2857, 1687, 1462, 1254, 1132, 1070, 834, 778.

2',3'-*O*-Bis-(*tert*-butyldimethylsilyl) uridine (44)

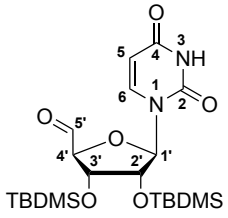


44

43 (6.66 g, 11.1 mmol, 1.0 eq.) was dissolved in THF (125 mL) and cooled to 0 °C. A 1:1-mixture of TFA in water (55 mL) was added dropwise to the solution and stirred for another 6 hours. Upon appearance of first double-deprotection products by TLC control, the reaction was quenched with saturated, aqueous $NaHCO_3$ solution (500 mL) and $NaHCO_3$ powder to yield a pH of 8. EtOAc (1.0 L) was added and the organic layer was washed with saturated, aqueous $NaHCO_3$ solution (500 mL), dried over Na_2SO_4, filtered and dried under reduced pressure. The resulting crude was further purified by flash column chromatography (400 g; DCM:EtOAc, 3:1) to give the product (**44**) as a white foam (3.59 g, 7.59 mmol; 69% yield). $\mathbf{R}_f$ (TLC) = 0.23 (DCM:EtOAc, 7.5:2.5); **MS** (ESI) m/z calc. for $C_{21}H_{40}N_2O_6Si_2$ $[M+H]^+$ 473.25; found 473.07; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.03 (s, 3 H, $SiCH_3$), 0.06 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 0.91 (s, 9 H, $SiC(CH_3)_3$), 3.75 (d, J = 12.1 Hz, 1 H, 5'-H_a), 3.94 (dd, J = 12.3, 2.1 Hz, 1 H, 5'-H_b), 4.09 (dd, J = 5.4, 2.1 Hz, 1 H, 4'-H), 4.17 (dd, J = 4.5, 3.4 Hz, 1 H, 3'-H), 4.56 (dd, J = 5.9, 4.6 Hz, 1 H, 2'-H), 5.45 (d, J = 5.5 Hz, 1 H, 1'-H), 5.72 (dd, J = 8.1, 2.3 Hz, 1 H, 5-H), 7.59 (d, J = 8.1 Hz, 1 H, 6-H), 8.05 (s, 1 H, NH); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.77 ($SiCH_3$), -4.64 ($SiCH_3$), -4.45 ($SiCH_3$), -4.28 ($SiCH_3$), 18.10 ($SiC(CH_3)_3$), 18.22 ($SiC(CH_3)_3$), 25.92 ($SiC(CH_3)_3$), 25.97 ($SiC(CH_3)_3$), 61.94 (C-5'), 71.97 (C-3'), 73.60 (C-2'), 86.47 (C-4'), 94.30 (C-1'), 102.33 (C-5), 143.10 (C-6), 150.08 (C-2),

162.47 (C-4); **UV/Vis** (MeOH) λ_{max} = 262; **IR** (ATR) $\tilde{\nu}$ = 3448, 2952, 2929, 2857, 1693, 1463, 1258, 1159, 1098, 869, 836, 778.

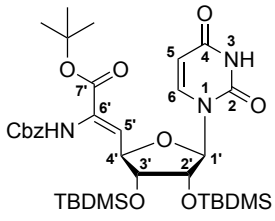
2',3'-*O*-Bis-(*tert*-butyldimethylsilyl) uridine 5'-aldehyde (15)



15

Intermediate **44** (1.70 g, 3.60 mmol, 1.0 eq.) was dissolved in dry acetonitrile (34.4 mL). IBX (2.02 g, 7.20 mmol, 2.0 eq.) was added and the reaction mixture was stirred at 80 °C (reflux) for 1.5 hours. The mixture was cooled to 0 °C for at least 45 min and IBX was filtered off, including several washes with EtOAc. The solvent was removed *in vacuo* (cold!) to give the product (**15**) as a white foam (1.75 g; 100% = 1.69 g). NMR analysis already showed decomposition of the aldehyde, so the product was immediately used for the next reaction without further purification. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.01 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.12 (s, 3 H, $SiCH_3$), 0.12 (s, 3 H, $SiCH_3$), 0.87 (s, 9 H, $SiC(CH_3)_3$), 0.93 (s, 9 H, $SiC(CH_3)_3$), 4.23 (dd, *J* = 3.8, 3.5 Hz, 1 H, 3'-H), 4.32 (dd, *J* = 5.6, 4.1 Hz, 1 H, 2'-H), 4.55 (d, *J* = 3.2 Hz, 1 H, 4'-H), 5.74 (d, *J* = 5.6 Hz, 1 H, 1'-H), 5.80 (dd, *J* = 8.1, 2.3 Hz, 1 H, 5-H), 7.70 (d, *J* = 8.1 Hz, 1 H, 6-H), 8.52 (s, 1 H, NH), 9.82 (s, 1 H, 5'-H).

Z-Didehydro uridinyl amino acid (46)

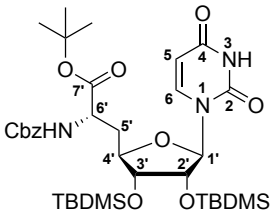


46

Phosphonate **45** (1.08 g, 2.88 mmol, 1.0 eq.) was dissolved in dry THF (14.0 mL) and added to KHMDS (0.5 M in toluene) in dry THF (12.0 mL) at -78 °C under nitrogen atmosphere. The mixture was stirred for about 10 min at this temperature, then aldehyde **15** (1.69 g, 3.60 mmol, 1.25 eq.) in dry THF (18.0 mL) was added dropwise. The reaction was allowed to warm to room temperature overnight and TLC control confirmed sufficient conversion. MeOH (5.0 mL) was added, and the mixture was diluted with EtOAc (100 mL). The organic layer was washed with a 1:1-mixture of brine and water (2 x 100 mL), dried over Na_2SO_4, filtered and dried under reduced pressure. The resulting crude product was further purified by flash column chromatography (200 g; PE:EtOAc, 7:3, then 3:2) to give the product (**46**) as a white foam (1.16 g, 1.62 mmol; 56% yield). **R_f** (TLC) = 0.29 (PE:EtOAc, 3:2); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.07 (s, 3 H, $SiCH_3$), 0.07 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.89 (s, 9 H, $SiC(CH_3)_3$), 0.90 (s, 9 H,

$SiC(CH_3)_3$), 1.48 (s, 9 H, $OC(CH_3)_3$), 3.96 (dd, *J* = 6.3, 3.8 Hz, 1 H, 3'-H), 4.34 (dd, *J* = 3.8, 3.4 Hz, 1 H, 2'-H), 4.87 (dd, *J* = 7.8, 6.0 Hz, 1 H, 4'-H), 5.14 (s, 2 H, Cbz-CH_2), 5.58 (d, *J* = 3.5 Hz, 1 H, 1'-H), 5.73 (dd, *J* = 8.0, 2.5 Hz, 1 H, 5-H), 6.27 (d, *J* = 7.8 Hz, 1 H, 5'-H), 6.72 (s, 1 H, 6'-NH), 7.27 (d, *J* = 8.0 Hz, 1 H, 6-H), 7.34-7.36 (m, 5 H, Cbz-aryl-H), 8.18 (s, 1 H, 3-NH); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.87 ($SiCH_3$), -4.83 ($SiCH_3$), -4.49 ($SiCH_3$), -4.42 ($SiCH_3$), 18.01 ($SiC(CH_3)_3$), 18.09 ($SiC(CH_3)_3$), 25.77 ($SiC(CH_3)_3$), 25.83 ($SiC(CH_3)_3$), 27.84 ($OC(CH_3)_3$), 67.50 (Cbz-CH_2), 74.68 (C-2'), 76.04 (C-3'), 79.14 (C-4'), 82.64 ($OC(CH_3)_3$), 92.84 (C-1'), 102.31 (C-5), 124.61 (C-5'), 128.15, 128.29, 128.50 (Cbz-aryl-C), 131.30 (C-6'), 135.76 (Cbz-C_q), 140.44 (C-6), 149.74 (C-2), 153.52 (Cbz-C=O), 162.69 (C-4), 162.95 (C-7'); **IR** (ATR) $\tilde{\nu}$ = 3249, 2953, 2930, 2887, 2857, 1692, 1460, 1368, 1257, 1223, 1155, 1083, 837, 777.

(6'*S*)-6'*N*-Cbz uridinyl amino acid (BB1)

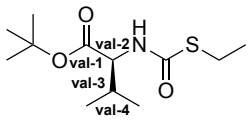


BB1

A solution of **46** (731 mg, 1.02 mmol, 1.0 eq.) in dry MeOH (45.0 mL) was degassed with argon (Ar) (~15 min x 13). (*S*,*S*)-Me-DUPHOS-Rh catalyst (two large spatula tips) was added under inert gas atmosphere. The mixture was flushed with hydrogen gas (grade 6.0) for 5 min, which was repeated every day until completion of the reaction. After seven days, the progress of the reaction was checked by ^{1}H NMR analysis. The reaction was completed after a total of 12 days. The reaction mixture was filtered over a short pad of silica gel and washed with EtOAc. The solvent was removed *in vacuo* and the crude was further purified by flash column chromatography (70 g; DCM:EtOAc, 3:2) to give the product (**BB1**) as a white foam (606 mg, 0.842 mmol; 83% yield). $\mathbf{R}_f$ (TLC) = 0.32 (DCM:EtOAc, 3:2); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.03 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.06 (s, 3 H, $SiCH_3$), 0.09 (s, 3 H, $SiCH_3$), 0.87 (s, 9 H, $SiC(CH_3)_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 1.44 (s, 9 H, $OC(CH_3)_3$), 1.97-2.03 (m, 1 H, 5'-H_a), 2.13-2.17 (m, 1 H, 5'-H_b), 3.64 (dd, *J* = 6.1, 4.3 Hz, 1 H, 3'-H), 4.14-4.20 (m, 2 H, 2'-H, 4'-H), 4.39-4.42 (m, 1 H, 6'-H), 5.08 (s, 2 H, Cbz-CH_2), 5.55 (d, *J* = 6.3 Hz, 1 H, 5-H), 5.60 (d, *J* = 2.8 Hz, 1 H, 1'-H), 5.74 (d, *J* = 7.8 Hz, 1 H, 6'-NH), 7.24-7.34 (m, 6 H, 6-H, Cbz-aryl-H), 8.38 (s, 1 H, 3-NH); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.93 ($SiCH_3$), -4.86 ($SiCH_3$), -4.57 ($SiCH_3$), -4.20 ($SiCH_3$), 17.92 ($SiC(CH_3)_3$), 17.96 ($SiC(CH_3)_3$), 25.74 ($SiC(CH_3)_3$), 25.76 ($SiC(CH_3)_3$), 27.86 ($OC(CH_3)_3$), 36.04 (C-5'), 52.12 (C-6'), 66.86 (Cbz-CH_2), 75.02 (C-2'), 75.20 (C-3'), 80.07 (C-4'), 82.77 ($OC(CH_3)_3$), 91.59 (C-1'), 102.17 (C-5), 124.61 (C-5'), 128.06, 128.15, 128.47 (Cbz-aryl-C), 136.22 (Cbz-C_q), 140.47 (C-6), 150.06 (C-2), 155.52 (Cbz-C=O), 163.35 (C-4), 170.54 (C-7'); **IR** (ATR) $\tilde{\nu}$ = 3232, 2954, 2930, 2895, 2857, 1692, 1461, 1369, 1258, 1158, 1061, 868, 838, 813.

*Synthesis of the Urea Dipeptide Building Block (**BB4**)*

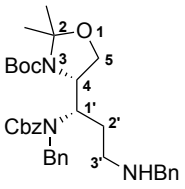
_N_-Thiocarbamate-functionalized valine _tert_-butyl ester (49)



49

Valine·HCl **47** (121 mg, 0.577 mmol, 1.0 eq.) was dissolved in a 1:1-mixture of EtOAc and saturated, aqueous $NaHCO_3$ solution (6.5 mL). Reagent **48** (60 µL, 0.58 mmol, 1.0 eq.) was added and the mixture was stirred at room temperature for 15 min. TLC control confirmed conversion of the starting material, so the reaction was quenched with MeOH (3.0 mL). The organic layer was washed with brine (3 x 20 mL), dried over Na_2SO_4, filtered and the solvent was removed *in vacuo*. The crude product was purified by column chromatography (PE:EtOAc, 9:1) to give the product (**49**) as a white solid (122 mg, 0.467 mmol; 81% yield). **R_f** (TLC) = 0.33 (PE:EtOAc, 9:1); **MS** (ESI) m/z calc. for $C_{12}H_{23}NO_3S$ $[M+Na]^+$ 284.12; found 284.02; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.88 (d, J = 7.0 Hz, 3 H, val-4-H), 0.93 (d, J = 7.0 Hz, 3 H, val-4-H), 1.27 (t, J = 7.0 Hz, 3 H, $CH_2C\underline{H}_3$), 1.45 (s, 9 H, $OC(CH_3)_3$), 2.09-2.16 (m, 1 H, val-3-H), 2.89 (q, J = 7.6, 7.2 Hz, 2 H, $C\underline{H}_2CH_3$), 4.40 (m, 1 H, val-2-H), 5.80 (d, J = 7.8 Hz, 1 H, NH); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 15.56 ($CH_2\underline{C}H_3$), 17.50 (C-val-4), 18.83 (C-val-4), 24.31 ($\underline{C}H_2CH_3$), 27.07 ($OC(\underline{C}H_3)_3$), 31.68 (C-val-3), 58.87 (C-val-2), 82.13 ($O\underline{C}(CH_3)_3$) 167.50 ((C=O)S), 170.62 (C-val-1); **IR** (ATR) $\tilde{\nu}$ = 3331, 2967, 2930, 2874, 1725, 1675, 1515, 1485, 1369, 1348, 1312, 1258, 1198, 1157, 869, 845, 776.

4-Alkyl-functionalized 2,2-dimethyloxazolidin-3-carboxylate (22)

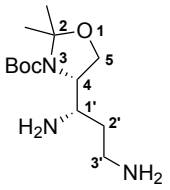


22

Alkene **21** (500 mg, 1.01 mmol, 1.0 eq.) – kindly provided by K. Leyerer in the lab – was dissolved in dry DCM (2.2 mL), dry MeOH (18.8 mL), and pyridine (330 µL, 4.04 mmol, 4.0 eq.) under nitrogen atmosphere. The mixture was cooled to -90 °C and ozone (O_3) was led through the solution for about 15 min. The slightly blue-colored solution was flushed with nitrogen until the blue color disappeared. DMS (750 µL, 10.1 mmol, 10.0 eq.) was added at -86 °C and the reaction was allowed to warm to room temperature overnight. The solvent was removed *in vacuo* and the crude was re-dissolved in dry DCM (14.4 mL), including molecular sieve (pore size = 4 Å), under nitrogen atmosphere. The solution was stirred at room temperature for 30 min, followed by addition of benzylamine (110 µL, 1.01 mmol, 1.0 eq.). The reaction mixture was stirred at room temperature for 23 hours, filtered through a pad of Celite®, and the solvent was removed under reduced pressure. The crude intermediate was re-dissolved in dry MeOH (22 mL) under nitrogen atmosphere and

cooled to 0 °C. At this temperature, $NaBH_4$ (96 mg, 2.5 mmol, 2.5 eq.) was added to the mixture and the reaction was stirred for 1 hour at this temperature and another 2 hours at room temperature. TLC control confirmed product formation, so the reaction was quenched with saturated, aqueous ammonium chloride (NH_4Cl) solution (100 mL). Water (50 mL) was added, and the aqueous layer was extracted with Et_2O (3 x 100 mL). The combined organic layers were dried over Na_2SO_4, filtered and the solvent was removed *in vacuo* to give the product (**22**) as a light-yellow oil (537 mg, 0.914 mmol; 91% yield), which was used without further purification. $\mathbf{R}_f$ (TLC) = 0.45 (DCM:MeOH, 9:1); **MS** (ESI) m/z calc. for $C_{35}H_{45}N_3O_5$ $[M+H]^+$ 588.34; found 588.34; **HRMS** (ESI) m/z calc. for $C_{35}H_{45}N_3O_5$ $[M+H]^+$ 588.3432; found 588.3405; **^{1}H NMR** (500 MHz, DMSO-d_6, 100 °C) δ 1.43 (s, 3 H, $C(CH_3)_2$), 1.45 (s, 9 H, $C(CH_3)_3$), 1.50 (s, 3 H, $C(CH_3)_2$), 1.73-1.78 (m, 1 H, 2'-H_a), 1.84-1.92 (m, 1 H, 2'-H_b), 2.35-2.45 (m, 2 H, 3'-H), 2.99 (s, 1 H, NH), 3.56-3.62 (m, 3 H, 5-H_a, Bn-CH_2), 3.77 (d, J = 9.1 Hz, 1 H, 5-H_b), 3.97-4.04 (m, 1 H, 4-H), 4.15-4.28 (m, 1 H, 1'-H), 4.28 (d, J = 14.4 Hz, 1 H, Bn-C$\underline{H}_a$$H_b$), 4.60 (d, J = 14.4 Hz, 1 H, Bn-$CH_a\underline{H}_b$), 5.14 (s, 2 H, Cbz-CH_2), 7.11-7.33 (m, 15 H, aryl-H); **^{13}C NMR** (126 MHz, DMSO-d_6, 100 °C) δ 24.12 (C-2'), 27.74 ($C(\underline{C}H_3)_2$), 28.57 ($C(\underline{C}H_3)_3$), 46.42 (Bn-CH_2), 53.42 (Bn-CH_2), 58.18 (C-1'), 59.13 (C-5), 65.58 (C-4), 67.18 (Boc-CH_2), 80.04 ($\underline{C}(CH_3)_3$), 94.05 (C-2), 126.81, 127.38, 127.82, 127.98, 128.17, 128.20, 128.40, 128.67, 128.71, 137.23, 139.39, 141.39 (aryl-C), 152.47 (Boc-C=O), 156.98 (Cbz-C=O); **IR** (ATR) $\tilde{\nu}$ = 3063, 3029, 2977, 2933, 2874, 1692, 1495, 1453, 1365, 1240, 1169, 1099, 1079, 847, 766, 734, 698.

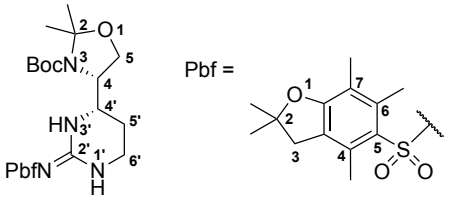
Partially deprotected 4-alkyl-functionalized 2,2-dimethyloxazolidin-3-carboxylate (23)



23

Intermediate **22** (260 mg, 0.442 mmol, 1.0 eq.) was deprotected following Method B. After 4 hours, TLC control confirmed complete deprotection of the starting material, thus, the reaction was filtered and the solvent was removed *in vacuo* to give the crude product (**23**) as a light-yellow oil (207 mg; 100% = 121 mg), which was used immediately in the subsequent reaction without further purification.

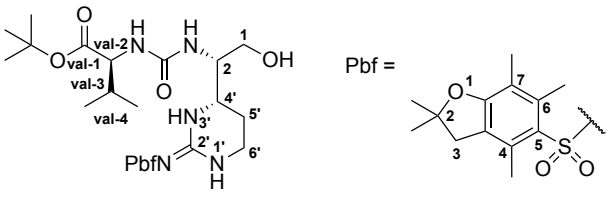
N-Pbf-protected cyclic guanidine (25)



25

Deprotected intermediate **23** (121 mg; crude) was dissolved in a 1:1-mixture of dry THF/dry DMF (5.4 mL) under nitrogen atmosphere. Guanidinylation reagent **24** (182 mg, 0.486 mmol, 1.1 eq.) and TEA (250 µL, 1.81 mmol, 4.1 eq.) were added. After addition of AgOTf (284 mg, 1.11 mmol, 2.5 eq.), the reaction mixture was stirred at room temperature for 19 hours. TLC control confirmed full conversion, and the mixture was filtered. The filtrate was washed with brine (3 x 100 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude was further purified by flash column chromatography (50 g; PE:EtOAc, 2:3) to give the product (**25**) as a clear oil (135 mg, 0.245 mmol; 56% yield over two steps from **22**). $\mathbf{R_f}$ (TLC) = 0.11 (PE:EtOAc, 2:3); **MS** (ESI) m/z calc. for $C_{27}H_{42}N_4O_6S$ $[M+H]^+$ 551.29; found 551.28; **HRMS** (ESI) m/z calc. for $C_{27}H_{42}N_4O_6S$ $[M+H]^+$ 551.2898; found 551.2930; **^{1}H NMR** (500 MHz, DMSO-d_6, 100 °C) δ 1.40 (s, 6 H, Pbf-2-C$(CH_3)_2$), 1.44 (s, 6 H, C$(CH_3)_2$), 1.46 (s, 9 H, Boc-C$(CH_3)_3$), 1.46-1.64 (m, 1 H, 5'-H_a), 1.78-1.85 (m, 1 H, 5'-H_b), 2.04 (s, 3 H, Pbf-7-CH_3), 2.46 (s, 3 H, Pbf-6-CH_3), 2.51 (s, 3 H, Pbf-4-CH_3), 2.97 (s, 2 H, Pbf-3-H), 3.20-3.26 (m, 1 H, 6'-H_a), 3.32-3.37 (m, 1 H, 6'-H_b), 3.61-3.67 (m, 1 H, 4'-H), 3.86-3.88 (m, 1 H, 4-H), 3.91-3.96 (m, 2 H, 5-H), 7.11 (s, 1 H, guanidine-NH), 7.21 (s, 1 H, guanidine-NH); **^{13}C NMR** (126 MHz, DMSO-d_6, 100 °C) δ 12.46 (C-Pbf-7), 17.70 (C-Pbf-6), 19.07 (C-Pbf-4), 23.42 (C($\underline{C}H_3)_2$), 28.50 (Boc-C($\underline{C}H_3)_3$), 28.65 (Pbf-2-C($\underline{C}H_3)_2$), 31.02 (C-5'), 37.98 (C-6'), 43.24 (C-Pbf-3), 51.67 (C-4'), 59.74 (C-4), 63.96 (C-5), 80.55 (Boc-$\underline{C}(CH_3)_3$), 86.59 (Pbf-$\underline{C}(CH_3)_2$), 94.24 (C-2), 116.54, 124.69, 131.98, 135.63, 137.50 (aryl-C), 152.87 (Boc-C=O), 154.11 (guanidine-C), 157.94 (C-Pbf-7_a); **IR** (ATR) $\tilde{\nu}$ = 3325, 3165, 2974, 2932, 1694, 1612, 1553, 1368, 1303, 1257, 1170, 1103, 902, 649.

Urea dipeptide alcohol (27)

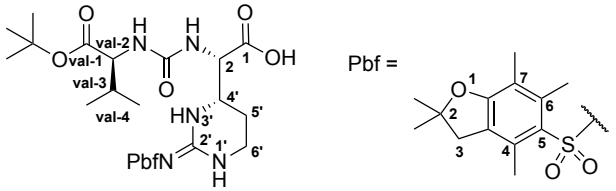


27

Guanidine **25** (100 mg, 0.182 mmol, 1.0 eq.) was dissolved in dry THF (2.0 mL) under nitrogen atmosphere. A 5.0 M HCl solution (0.37 mL) was added to the mixture and the reaction was stirred for 2 hours under reflux (80 °C). The solvent was removed under reduced pressure and the crude was re-dissolved in dry DMF (2.0 mL) under nitrogen atmosphere. Valine **49** (57 mg, 0.22 mmol, 1.2 eq.) was added, followed by dropwise addition of TEA (110 µL, 0.782 mmol, 4.3 eq.). The reaction was stirred for 17 hours at

room temperature. After complete conversion, the mixture was filtered, and the filtrate was washed with saturated, aqueous $NaHCO_3$ solution (2 x 50 mL) and brine (40 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude was purified by flash column chromatography (11 g; DCM:MeOH, 97:3) to give the product (**27**) as a clear oil (79 mg, 0.13 mmol; 71% yield). $\mathbf{R_f}$ (TLC) = 0.49 (DCM:MeOH, 99:1); **MS** (ESI) m/z calc. for $C_{29}H_{47}N_5O_7S$ $[M+H]^+$ 610.32; found 610.30; **HRMS** (ESI) m/z calc. for $C_{29}H_{47}N_5O_7S$ $[M+H]^+$ 610.3269; found 610.3259; **^{1}H NMR** (500 MHz, DMSO-d_6, 100 °C) δ 0.90 (d, *J* = 6.6 Hz, 3 H, val-4-H), 0.91 (d, *J* = 6.6 Hz, 3 H, val-4-H), 1.43 (s, 9 H, $OC(CH_3)_3$), 1.45 (s, 6 H, Pbf-2-$C(CH_3)_2$), 1.71-1.83 (m, 2 H, 5'-H), 1.96-2.03 (m, 1 H, val-3-H), 2.05 (s, 3 H, Pbf-7-CH_3), 2.46 (s, 3 H, Pbf-6-CH_3), 2.50-2.52 (m, 3 H, Pbf-4-CH_3), 2.98 (s, 2 H, Pbf-3-H), 3.13-3.18 (m, 1 H, 6'-H_a), 3.29-3.34 (m, 1 H, 6'-H_b), 3.43-3.48 (m, 2 H, 4'-H, 1-H_a), 3.56-3.61 (m, 2 H, 2-H, 1-H_b), 3.94-3.97 (m, 1 H, val-2-H), 6.08 (d, *J* = 11.7 Hz, 1 H, urea-NH), 6.15 (d, *J* = 11.7 Hz, 1 H, urea-NH), 7.14 (s, 2 H, guanidine-NH); **^{13}C NMR** (126 MHz, DMSO-d_6, 100 °C) δ 12.48 (C-Pbf-7), 17.73 (C-Pbf-6), 18.35 (C-val-4), 19.09 (C-Pbf-4), 19.32 (C-val-4), 22.32 ($OC(\underline{C}H_3)_3$), 28.30 (C-val-3), 28.66 (Pbf-2-$C(\underline{C}H_3)_2$), 30.94 (C-5'), 37.30 (C-6'), 43.24 (C-Pbf-3), 50.87 (C-4'), 53.66 (C-val-2), 59.28 (C-2), 61.67 (C-1), 80.76 ($O\underline{C}(CH_3)_3$), 86.58 (Pbf-2-$\underline{C}(CH_3)_2$), 116.55, 124.70, 132.01, 135.57, 137.55 (aryl-C), 153.73 (guanidine-C), 157.93 (C-Pbf-7_a), 158.42 (urea-C=O), 172.01(C-val-1); **IR** (ATR) $\tilde{\nu}$ = 3345, 2970, 2930, 1730, 1651, 1610, 1550, 1457, 1369, 1349, 1305, 1249, 1156, 1093, 848, 657.

Urea dipeptide (BB4)



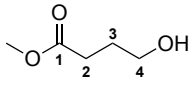
BB4

Alcohol **27** (82 mg, 0.14 mmol, 1.0 eq.), and TEMPO (6 mg, 0.04 mmol, 0.3 eq.) were dissolved in acetonitrile (4.8 mL) and phosphate buffer (0.67 M; pH = 6.7; 2.9 mL). $NaClO_2$(39 mg, 0.43 mmol, 3.2 eq.) was added and the mixture was heated to 35 °C. After dropwise addition of an aqueous 5% NaOCl solution (40 µL, 0.041 mmol, 0.3 eq.), the reaction was stirred for 24 hours. For complete conversion, another batch of TEMPO (6 mg, 0.04 mmol, 0.3 eq), $NaClO_2$ (39 mg, 0.43 mmol, 3.2 eq.), and aqueous 5% NaOCl solution (40 µL, 0.041 mmol, 0.3 eq.) was added and the mixture was stirred for another 48 hours. TLC control confirmed product formation and the mixture was diluted with EtOAc (50 mL). The organic layer was washed with saturated, aqueous sodium thiosulfate ($Na_2S_2O_3$) solution (2 x 50 mL) and brine (2 x 40 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography (10 g; DCM:MeOH, 95:5, then 8:1) to give the product (**BB4**) as a white solid (75 mg, 0.12 mmol; 89% yield). $\mathbf{R_f}$ (TLC) = 0.32 (DCM:MeOH, 9:1); **MS** (ESI) m/z calc. for $C_{29}H_{45}N_5O_8S$ $[M+H]^+$ 624.31; found 624.29; **HRMS** (ESI) m/z calc. for $C_{29}H_{45}N_5O_8S$ $[M+H]^+$ 624.3062; found 624.3076; **^{1}H NMR** (500 MHz, DMSO-d_6, 100 °C) δ 0.90 (d, *J* = 1.4 Hz, 3 H, val-4-H), 0.91 (d, *J* = 1.4 Hz, 3 H, val-4-H), 1.42 (s, 9 H, $OC(CH_3)_3$), 1.44 (s, 6 H, Pbf-2-$C(CH_3)_2$), 1.58-1.66 (m, 1 H, 5'-H_a), 1.80-1.88 (m, 1 H, 5'-

H_b), 1.95-2.04 (m, 1 H, val-3-H), 2.04 (s, 3 H, Pbf-7-CH_3), 2.46 (s, 3 H, Pbf-6-CH_3), 2.51 (s, 3 H, Pbf-4-CH_3), 2.97 (s, 2 H, Pbf-3-H), 3.09-3.14 (m, 1 H, 6'-H_a), 3.32-3.38 (m, 2 H, 6'-H_b, 4'-H), 3.81-3.84 (t, *J* = 7.1 Hz, 1 H, 2-H), 3.95 (dd, *J* = 5.4, 3.1 Hz, 1 H, val-2-H), 6.32 (d, *J* = 6.1 Hz, 1 H, urea-NH), 6.41 (d, *J* = 8.5 Hz, 1 H, urea-NH), 7.09 (s, 1 H, guanidine-NH), 7.52 (s, 1 H, guanidine-NH); **^{13}C NMR** (126 MHz, DMSO-d_6, 100 °C) δ 12.49 (C-Pbf-7), 17.73 (C-Pbf-6), 18.53 (C-val-4), 19.11 (C-Pbf-4), 19.35 (C-val-4), 23.95 (OC($\underline{C}H_3$)$_3$), 28.33 (C-val-3), 28.69 (Pbf-2-C($\underline{C}H_3$)$_2$), 30.92 (C-5'), 38.73 (C-6'), 43.28 (C-Pbf-3), 53.80 (C-4'), 57.52 (C-val-1), 59.49 (C-2), 80.56 ($\underline{C}(CH_3)_3$), 86.51 (Pbf-2-$\underline{C}(CH_3)_2$), 116.45, 124.59, 132.03, 135.81, 137.53 (aryl-C), 153.63 (guanidine-C), 157.86 (C-Pbf-7$_a$), 158.88 (urea-C=O), 172.03 (C-val-1), 173.61 (C-1); **UV/Vis** (MeOH) λ_{max} = 253; **IR** (ATR) $\tilde{\nu}$ = 3347, 2969, 2930, 1730, 1602, 1553, 1456, 1407, 1370, 1305, 1259, 1155, 1105, 1007, 845, 622.

Synthesis of Fatty Acid Derivatives

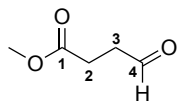
Methyl 4-hydroxybutyrate (50)



50

γ-Butyrolactone (**40**) (2.0 mL, 26 mmol, 1.0 eq.) was dissolved in MeOH (131 mL). TEA (18.5mL, 132 mmol, 5.1 eq.) was added and the mixture was stirred under reflux (60 °C) for 15 days. The mixture was diluted with toluene (131 mL) and the solvent was removed *in vacuo*. The crude was co-evaporated with toluene a few more times (2 x 50 mL) and the crude product (**50**) was yielded as a yellow oil (2.54 g; 100% = 3.10 g), which was used without further purification. Note: NMR analysis showed a ratio of ~1:4 of lactone to ring-opened product. $\mathbf{R}_f$ (TLC) = 0.50 (DCM:MeOH, 9:1); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.83-1.90 (m, 2 H, 3-H), 2.43 (t, *J* = 6.7 Hz, 2 H, 2-H), 3.65-3.68 (m, 5 H, 4-H, OCH_3); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 27.62 (C-3), 30.76 (C-2), 51.68 (OCH_3), 62.02 (C-4), 174.39 (C-1); **IR** (ATR) $\tilde{\nu}$ = 3376, 2959, 2853, 1715, 1644, 1441, 1367, 1325, 1259, 1234, 1206, 1058, 799.

Methyl 4-oxobutyrate (38)

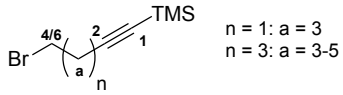


38

Alcohol **50** (503 mg, 3.60 mmol (calc.; assuming 1:4 ratio of **40**:**50**), 1.0 eq.) was dissolved in dry acetonitrile (30 mL) under nitrogen atmosphere. IBX (2.84 g, 10.2 mmol, 2.3 eq.) was added and the mixture was stirred at 80 °C for 5 hours. TLC control showed product formation and the mixture was cooled to 0 °C for at least 30 min. The precipitate was filtered off and washed several times with ice cold EtOAc (200 mL). The filtrate was dried *in vacuo* to give the product (**38**) as a yellow solid (513 mg; 100% = 418 mg; 70% calc. yield over 2 steps from **40**). Note: NMR analysis showed a ratio of 1:4 of lactone (**40**) to

aldehyde (**38**). $\mathbf{R}_f$ (TLC) = 0.58 (DCM:MeOH, 9:1); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 2.62 (t, *J* = 6.7 Hz, 2 H, 2-H), 2.79 (t, *J* = 6.7 Hz, 2 H, 3-H), 3.67 (s, 3 H, OCH_3), 9.80 (s, 1 H, 4-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 26.28 (C-2), 38.51 (C-3), 51.95 (OCH_3), 172.90 (C-1), 199.97 (C-4).

TMS-protected 4/6-bromoalkynes (52/58)



52: n = 1
58: n = 3

n = 1:

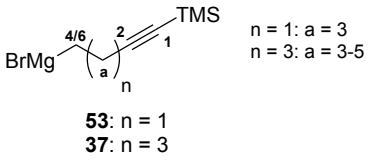
Alcohol **51** (1.85 g, 26.4 mmol, 1.0 eq.) was dissolved in dry THF (40 mL) under nitrogen atmosphere. The solution was cooled to -78 °C and *n*-BuLi (2.5 M in *n*-hexanes; 22.2 mL, 55.5 mmol, 2.1 eq.) was added dropwise at this temperature. The mixture was stirred for ~10 min, followed by the addition of TMSCl (7.39 mL, 58.1 mmol, 2.2 eq.). The reaction was allowed to warm to room temperature and TLC showed formation of TMS-protected products. The double-protected intermediate was converted to the free alcohol by treatment with a 3.0 M HCl solution (20 mL) and was stirred overnight at room temperature. TLC control confirmed cleavage of the alcoholic TMS group. The reaction mixture was then extracted with Et_2O (70 mL + 2 x 40 mL). The combined organics were washed with saturated, aqueous $NaHCO_3$ (2 x 50 mL) and brine (50 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude was re-dissolved in a 1:1-mixture of dry Et_2O/acetonitrile (3.4 mL) under nitrogen atmosphere. PPh_3 (10.4 g, 39.6 mmol, 1.5 eq.) was added to the solution. After cooling to 0 °C, NBS (7.04 g, 39.6 mmol, 1.5 eq.) was added over the course of ~10 min. The reaction mixture was allowed to warm to room temperature and stirred for 3 hours. TLC confirmed conversion, so the mixture was diluted with PE (40 mL) and filtered through a pad of Celite®, which was washed with PE several times. The filtrate was dried under reduced pressure and the crude product was purified through a pad of silica gel (PE:EtOAc, 9:1) to give the product (**64**) as a yellow oil (4.12 g, 20.1 mmol; 76% yield). $\mathbf{R}_f$ (TLC) = 0.61 (PE:EtOAc, 9:1); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.16 (s, 9 H, $Si(CH_3)_3$), 2.77 (t, *J* = 7.5 Hz, 2 H, 3-H), 3.43 (t, *J* = 7.5 Hz, 2 H, 4-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 0.08 ($Si(CH_3)_3$), 24.42 (C-3), 29.34 (C-4), 87.16 (C-2), 103.33 (C-1); **UV/Vis** (MeOH) λ_{max} = 205; **IR** (ATR) $\tilde{\nu}$ = 2962, 1260, 1092, 1018, 842, 799.

n = 3:

Alcohol **39** (3.0 mL, 27 mmol, 1.0 eq.) was dissolved in dry THF (60 mL) under nitrogen atmosphere. The solution was cooled to -78 °C and *n*-BuLi (2.5 M in *n*-hexanes; 22.6 mL, 56.5 mmol, 2.1 eq.) was added dropwise at this temperature. The mixture was stirred for ~10 min, followed by the addition of TMSCl (7.53 mL, 59.2 mmol, 2.2 eq.). The reaction was allowed to warm to room temperature and TLC showed formation of TMS-protected products. The double-protected intermediate was converted to the free alcohol by treatment with a 3.0 M HCl solution (30 mL) and was stirred overnight at room temperature. TLC control confirmed cleavage of the alcoholic TMS group. The reaction mixture was extracted with Et_2O (90 mL + 2 x 60 mL). The combined organic layers were washed with

saturated, aqueous $NaHCO_3$ (2 x 60 mL) and brine (60 mL), dried over Na_2SO_4, filtered and the solvent was removed *in vacuo*. The crude was re-dissolved in a 1:1-mixture of dry Et_2O/acetonitrile (3.4 mL) under nitrogen atmosphere. PPh_3 (10.60 g, 40.4 mmol, 1.5 eq.) was added to the solution. After cooling to 0 °C, NBS (7.19 g, 40.4 mmol, 1.5 eq.) was added over the course of ~10 min. The reaction mixture was allowed to warm to room temperature and stirred for 3 hours. TLC confirmed conversion, so the mixture was diluted with PE (150 mL) and filtered through a pad of Celite®, which was washed with PE several times. The filtrate was dried under reduced pressure and the crude was purified through a pad of silica gel (PE:EtOAc, 9:1) to give the product (**58**) as a light-yellow oil (6.27 g, 26.9 mmol; 100% yield). $\mathbf{R_f}$ (TLC) = 0.70 (PE:EtOAc, 9:1); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.15 (s, 9 H, $Si(CH_3)_3$), 1.67 (tt, *J* = 7.1, 7.1 Hz, 2 H, 4-H), 1.95-2.00 (m, 2 H, 5-H), 2.27 (t, *J* = 7.0 Hz, 2 H, 3-H), 3.44 (t, *J* = 6.7 Hz, 2 H, 6-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 0.27 ($Si(CH_3)_3$), 19.17 (C-3), 27.09 (C-4), 31.81 (C-5), 33.38 (C-6), 85.37 (C-2), 106.56 (C-1); **UV/Vis** (MeOH) λ_{max} = 203; **IR** (ATR) $\tilde{\nu}$ = 2958, 2174, 1249, 841, 759.

TMS-protected Grignard reagents (53/37)



53: n = 1
37: n = 3

<u>n = 1:</u>

First Attempt: Magnesia turnings (199 mg, 8.19 mmol, 3.4 eq.) were stirred without solvent under nitrogen atmosphere for about 10 min. Dry THF (3.0 mL) was added and the mixture was stirred for another 30 min. Bromide **52** (487 mg, 2.38 mmol, 1.0 eq.) was dissolved in dry THF (9 mL) and was added dropwise to the magnesia turnings. To initiate the conversion, a few crumbs of iodine were added. After about 2 hours of stirring at room temperature, no conversion had happened.

Second Attempt: Magnesia turnings (70 mg, 2.9 mmol, 1.5 eq.) were stirred without solvent under nitrogen atmosphere for about 10 min. Dry THF (2.0 mL) was added and the mixture was stirred for another 20 min. Bromide **52** (195 mg, 0.952 mmol, 1.0 eq.) was dissolved in dry THF (4.0 mL) and was added dropwise to the magnesia turnings. To help start the conversion, a few crumbs of iodine were added, and the mixture was gently heated using a heat gun. The mixture was heated until reflux without conversion.

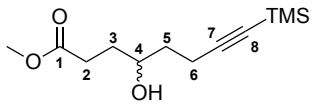
Third Attempt: Magnesia turnings (40 mg, 1.7 mmol, 1.7 eq.) were polished and stirred in dry THF (2.0 mL) under nitrogen atmosphere for about 10 min. A few crumbs of iodine were added to help activate the magnesia turnings. Bromide **52** (195 mg, 0.952 mmol, 1.0 eq.) was dissolved in dry THF (6 mL) and added dropwise to the magnesia turnings. The mixture was gently heated with a heat gun, but no conversion was observed.

Fourth Attempt: A new batch of magnesia turnings (183 mg, 7.53 mmol, 1.5 eq.) was stirred with dry THF (1.0 mL) under nitrogen atmosphere for about 3 hours. A few crumbs of iodine were added to help activate the magnesia turnings (stirred for about 5 min). Bromide **52** (1.03 g, 5.02 mmol, 1.0 eq.) was dissolved in dry THF (5.8 mL) and 0.5 mL were added immediately to the magnesia turnings. The mixture was gently heated with a heat gun and conversion was observed. The rest of the bromide solution was added in small portions, so that the solvent was steadily bubbling. After the reaction was cooling down, it was heated for another 30 min at 60 °C. The THF solution was immediately used for the following reaction.

n = 3:

Magnesia turnings (532 mg, 21.9 mmol, 1.5 eq.) were stirred with dry THF (3.3 mL) under nitrogen atmosphere for about 30 min. A few crumbs of iodine were added to help activate the magnesia turnings (stirred for about 5 min). Bromide **58** (3.00 g, 14.6 mmol, 1.0 eq.) was dissolved in dry THF (9.5 mL) and 0.5 mL were added immediately to the magnesia turnings and conversion was already observed. The rest of the bromide solution was added in small portions, so that the solvent was steadily bubbling. After the reaction was cooling down, it was heated for another 30 min at 60 °C. The THF solution was immediately used for the following reaction.

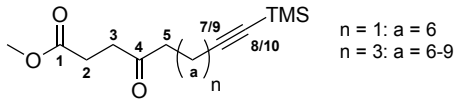
TMS-protected 4-hydroxy fatty acid ester (55)



55

For the attempted synthesis of **55**, freshly prepared Grignard reagent **53** (1.15 g, 5.02 mmol, 1.2 eq.) in dry THF (11.0 mL) was added dropwise to a solution of aldehyde **38** (485 mg, 4.18 mmol, 1.0 eq.) in dry THF (4.5 mL) under nitrogen atmosphere at -78 °C. The mixture was stirred at this temperature for 1.5 hours. TLC control showed side product formation and no product could be identified.

TMS-protected 4-oxo fatty acid esters (56/59)



56: n = 1
59: n = 3

n = 1:

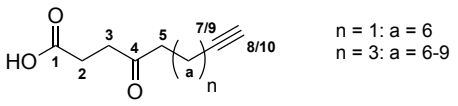
Acid chloride **54** (1.98 mL, 16.1 mmol, 1.1 eq.) and $Fe(acac)_3$ (516 mg, 1.46 mmol, 0.1 eq.) were dissolved in dry THF (30 mL) under nitrogen atmosphere. The mixture was cooled to -78 °C and freshly prepared Grignard reagent **53** (3.35 g, 14.6 mmol, 1.0 eq.) in dry THF (33 mL) was added dropwise to the reaction mixture. TLC control confirmed product formation and the reaction was quenched with saturated, aqueous NH_4Cl solution (100 mL). The aqueous layer was extracted with EtOAc (3 x 100 mL) and the combined

organics were dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography (120 g, PE:EtOAc, 99:1) to give the product (**56**) as a colorless oil (1.84 g, 7.65 mmol; 52% yield). $\mathbf{R}_f$ (TLC) = 0.34 (PE:EtOAc, 9:1); **MS** (ESI) m/z calc. for $C_{12}H_{20}O_3Si$ $[M+Na]^+$ 263.11; found 263.01; **HRMS** (ESI) m/z calc. for $C_{12}H_{20}O_3Si$ $[M+H]^+$ 241.1245; found 241.1235; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.11 (s, 9 H, $Si(CH_3)_3$), 2.47 (dd, *J* = 6.4, 2.1 Hz, 2 H, 6-H), 2.57 (t, *J* = 6.1 Hz, 2 H, 2-H), 2.68 (dd, *J* = 6.4, 2.1 Hz, 2 H, 5-H), 2.72 (t, *J* = 7.5 Hz, 2 H, 3-H), 3.64 (s, 3 H, OCH_3); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 0.07 ($Si(CH_3)_3$), 14.37 (C-6), 27.64 (C-2), 37.11 (C-3), 41.55 (C-5), 51.78 (OCH_3), 85.14 (C-8), 105.47 (C-7), 173.08 (C-1), 206.68 (C-4); **UV/Vis** (MeOH) λ_{max} = 206, 276; **IR** (ATR) $\tilde{\nu}$ = 2956, 2175, 1736, 1718, 1437, 1360, 1249, 1197, 1175, 1096, 838, 759, 698, 638.

n = 3:

Acid chloride **54** (290 µL, 2.35 mmol, 1.1 eq.) and $Fe(acac)_3$ (74 mg, 0.21 mmol, 0.1 eq.) were dissolved in dry THF (4.2 mL) under nitrogen atmosphere. The mixture was cooled to -78 °C and freshly prepared Grignard reagent **37** (551 mg, 2.14 mmol, 1.0 eq.) in dry THF (4.9 mL) was added dropwise to the reaction mixture. TLC control confirmed product formation and the reaction was quenched with saturated, aqueous NH_4Cl solution (12 mL). The aqueous layer was extracted with EtOAc (3 x 20 mL) and the combined organics were dried over Na_2SO_4, filtered and the solvent was removed *in vacuo*. The crude was further purified by flash column chromatography (60 g, PE:EtOAc, 99:1, then 9:1) to give the product (**71**) as a yellow oil (319 mg, 1.19 mmol; 56% yield). $\mathbf{R}_f$ (TLC) = 0.16 (PE:EtOAc, 9:1); **MS** (ESI) m/z calc. for $C_{14}H_{24}O_3Si$ $[M+Na]^+$ 291.14; found 290.96; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.10 (s, 9 H, $Si(CH_3)_3$), 1.45-1.51 (m, 2 H, 7-H), 1.63-1.69 (m, 2 H, 6-H), 2.19 (t, *J* = 7.1 Hz, 2 H, 8-H), 2.45 (t, *J* = 7.1 Hz, 2 H, 5-H), 2.56 (t, *J* = 6.4 Hz, 2 H, 3-H), 2.67 (t, *J* = 6.4 Hz, 2 H, 2-H), 3.64 (s, 3 H, OCH_3); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 0.34 ($Si(CH_3)_3$), 19.85 (C-8), 23.09 (C-6), 27.90 (C-2), 28.21 (C-7), 37.18 (C-3), 42.36 (C-5), 51.98 (OCH_3), 84.97 (C-9), 107.09 (C-10), 173.46 (C-1), 208.80 (C-4); **UV/Vis** (MeOH) λ_{max} = 206, 276; **IR** (ATR) $\tilde{\nu}$ = 2954, 2172, 1736, 1716, 1437, 1410, 1360, 1248, 1203, 1160, 1101, 839, 759, 698, 639.

4-Oxo fatty acids (57/35)



57: n = 1
35: n = 3

n = 1:

Protected fatty acid **56** (967 mg, 4.02 mmol, 1.0 eq.) was dissolved in THF (29 mL) and 4.4 M NaOH solution (9.1 mL) was added. The mixture was stirred at room temperature for 18 hours and TLC control confirmed complete conversion. The reaction was quenched with a 0.1 M HCl solution (420 mL) and the aqueous layer was extracted with DCM (3 x 420 mL). The combined organics were dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography (60 g, DCM:MeOH, 9:1) to give the product (**57**) as a light-yellow oil (537 mg, 3.48 mmol; 87%

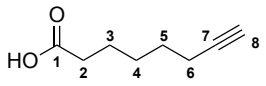
yield). $\mathbf{R_f}$ (TLC) = 0.23 (DCM:MeOH, 9:1, 0.1% TFA); **HRMS** (ESI) m/z calc. for $C_8H_{10}O_3$ $[M-H]^-$ 181.0870; found 181.0849; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.93 (t, *J* = 2.7 Hz, 1 H, 8-H), 2.44 (dt, *J* = 7.3, 2.7 Hz, 2 H, 6-H), 2.63 (t, *J* = 5.9 Hz, 2 H, 2-H), 2.67-2.73 (m, 4 H, 5-H, 3-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 12.90 (C-6), 27.63 (C-2), 36.78 (C-3), 41.24 (C-5), 68.81 (C-8), 82.81 (C-7), 178.44 (C-1), 206.32 (C-4); **UV/Vis** (MeOH) λ_{max} = 206, 276; **IR** (ATR) $\tilde{\nu}$ = 3289, 2936, 2869, 1708, 1405, 1375, 1211, 1180, 1103, 643.

n = 3:

Protected fatty acid **59** (300 mg, 1.12 mmol, 1.0 eq.) was dissolved in THF (8.0 mL) and 4.4 M NaOH solution (2.6 mL) was added. The mixture was stirred at room temperature for 19 hours and TLC control confirmed complete conversion. The reaction was quenched with a 0.1 M HCl solution (120 mL), and the aqueous layer was extracted with DCM (3 x 120 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography (45 g, DCM:MeOH, 3:1) to give the product (**35**) as a light-yellow solid (73 mg, 0.40 mmol; 36% yield). $\mathbf{R_f}$ (TLC) = 0.23 (DCM:MeOH, 3:1); **MS** (ESI) m/z calc. for $C_{10}H_{14}O_3$ $[M+H]^+$ 155.07; found 154.98; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.48-1.54 (m, 2 H, 7-H), 1.67-1.73 (m, 2 H, 6-H), 1.93 (t, *J* = 2.7 Hz, 1 H, 10-H), 2.18 (dt, *J* = 6.9, 2.7 Hz, 2 H, 8-H), 2.45 (t, *J* = 7.1 Hz, 2 H, 5-H), 2.62 (t, *J* = 6.4 Hz, 2 H, 3-H), 2.70 (t, *J* = 6.4 Hz, 2 H, 2-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 18.19 (C-8), 22.73 (C-6), 27.68 (C-2), 27.78 (C-7), 36.72 (C-3), 42.02 (C-5), 68.57 (C-10), 83.97 (C-9), 178.30 (C-1), 208.39 (C-4); **IR** (ATR) $\tilde{\nu}$ = 3288, 2923, 1709, 1405, 1374, 2151, 1190, 1098, 843, 646.

*Synthesis of the Reference Fatty Acid (**BB5-ref**)*

C_8 reference fatty acid (63)



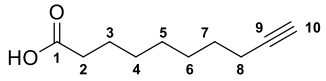
63

First Attempt: A 90% complex of lithium acetylide and ethylenediamine (917 mg, 8.96 mmol, 2.0 eq.) was dissolved in dry THF (11.3 mL) under nitrogen atmosphere. Acid bromide **60** (1.00 g, 4.48 mmol, 1.0 eq.) was dissolved in DMPU (5.8 mL) and slowly added to the mixture at 0 °C. The reaction was allowed to warm to room temperature overnight. However, no conversion was detected by TLC control and ^{1}H NMR analysis.

Second Attempt: Alkyne alcohol **60** (1.0 mL, 7.0 mmol, 1.0 eq.), TEMPO (330 mg, 2.11 mmol, 0.3 eq.), and BAIB (5.0 g, 15.5 mmol, 2.2 eq.) were dissolved in a 1:1-mixture of acetonitrile and water (40 mL). The reaction was stirred at room temperature for 3 hours, and TLC control verified conversion of the starting material. The organic layer was washed with 1.0 M HCl solution (30 mL). The aqueous layer was again extracted with EtOAc (3 x 30 mL). The combined organics were dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo*. The crude was purified by flash column chromatography (80 g, DCM:MeOH, 98:2; 0.1%

formic acid) to give the product (**75**) as a colorless oil (245 mg, 1.74 mmol; 25% yield). $\mathbf{R}_f$ (TLC) = 0.21 (DCM:MeOH, 98:2); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.42-1.47 (m, 2 H, 4-H), 1.51-1.56 (m, 2 H, 5-H), 1.61-1.66 (m, 2 H, 3-H), 1.92 (t, *J* = 2.7 Hz, 1 H, 8-H), 2.18 (dt, *J* = 7.3, 2.7 Hz, 2 H, 6-H), 2.34 (t, *J* = 7.4 Hz, 2 H, 2 H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 18.21 (C-6), 24.21 (C-3), 28.04 (C-5), 28.10 (C-4), 33.99 (C-2), 68.35 (C-8), 84.27 (C-7), 179.69 (C-1); **IR** (ATR) $\tilde{\nu}$ = 3296, 2939, 2863, 1708, 1460, 1412, 1275, 1225, 942, 636.

C_{10} reference fatty acid (BB5-ref)

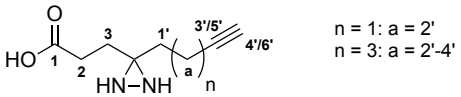


BB5-ref

Alcohol **76** (1.0 mL, 5.6 mmol, 1.0 eq.), TEMPO (264 mg, 1.69 mmol, 0.3 eq.), and BAIB (3.99 g, 12.4 mmol, 2.2 eq.) were dissolved in a 1:1-mixture of acetonitrile and water (32 mL). The reaction was stirred at room temperature for 18 hours, and TLC control verified conversion of the starting material. The organic layer was washed with 1.0 M HCl solution (20 mL). The aqueous layer was again extracted with EtOAc (3 x 20 mL). The combined organic layers were dried over Na_2SO_4, filtered and the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography (150 g, DCM:MeOH, 98:2; 0.1% formic acid) to give the product (**BB5-ref**) as a colorless oil (142 mg, 0.844 mmol; 15% yield). $\mathbf{R}_f$ (TLC) = 0.21 (DCM:MeOH, 98:2); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.27-1.35 (m, 4 H, 5-H, 6-H), 1.35-1.41 (m, 2 H, 4-H), 1.47-1.53 (m, 2 H, 7-H), 1.59-1.65 (m, 2 H, 3-H), 1.92 (t, *J* = 2.7 Hz, 1 H, 10-H), 2.16 (dt, *J* = 7.3, 2.6 Hz, 2 H, 8-H), 2.33 (t, *J* = 7.4 Hz, 2 H, 2-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 18.31 (C-8), 24.59 (C-3), 28.32, 28.45 (C-5, C-6), 28.66 (C-7), 28.85 (C-4), 34.01 (C-2), 68.11 (C-10), 84.59 (C-9), 179.90 (C-1); **IR** (ATR) $\tilde{\nu}$ = 3298, 2932, 2857, 1704, 1459, 1411, 1284, 1246, 934, 627.

*Synthesis of the Diazirine Function (**BB5-PAL**)*

3-(3-(but-3-yn-1-yl)diaziridin-3-yl)propanoic acid (65)/ 3-(3-(hex-5-yn-1-yl)diaziridin-3-yl)propanoic acid (34)



65: n = 1
34: n = 3

<u>n = 1:</u>

Fatty acid **57** (25 mg, 0.16 mmol, 1.0 eq.) was dissolved in dry MeOH (100 µL) and cooled to 0 °C. A 7.0 N ammonia in MeOH solution (2.4 mL) was slowly added and the mixture was stirred for 4 hours. After the mixture was cooled to -12 °C, HAOSA (20 mg, 0.178 mmol, 1.1 eq.) in dry MeOH (340 µL) was added and the reaction was allowed to warm to room temperature for 22 hours.

Work-up 1: The suspension was filtered, and the solvent was removed *in vacuo*. The resulting intermediate turned out to be very instable and would decompose spontaneously.

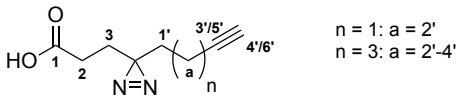
Work-up-2: The reaction mixture was dried under a stream of nitrogen. The crude was re-dissolved in MeOH, filtered and evaporated *in vacuo*. The concentrated crude was diluted with more MeOH and it was continued with the following step without further work-up. The intermediate seemed to be stable.

Work-up-3: The suspension was filtered, and TEA (1.6 eq.) was added. The basic filtrate was evaporated to ~0.5 mL and more TEA (1.6 eq.) was added. It was continued with the next reaction without any further work-up, and the intermediate was stable. Decomposition was observed in the absence of TEA, thus, potential decomposition during work-up 2 could occur.

n = 3:

Fatty acid **35** (25 mg, 0.13 mmol, 1.0 eq.) was dissolved in dry MeOH (100 µL) and cooled to 0 °C. A 7.0 N ammonia in MeOH solution (2.0 mL) was slowly added and the mixture was stirred for 3 hours. After the mixture was cooled to -12 °C, HAOSA (30 mg, 0.27 mmol, 2.0 eq.) in dry MeOH (510 µL) was added and the reaction was allowed to warm to room temperature for 18.5 hours. The suspension was filtered, and TEA (20 µL, 0.14 mmol, 1.04 eq.) was added. The basic filtrate was evaporated to ~0.5 mL and more TEA (20 µL, 0.14 mmol, 1.04 eq.) was added. It was continued with the next reaction without any further work-up, and the intermediate was stable as confirmed by TLC control.

3-(3-(but-3-yn-1-yl)-3*H*-diazirin-3-yl)propanoic acid (66)/
3-(3-(hex-5-yn-1-yl)-3*H*-diazirin-3-yl)propanoic acid (BB5-PAL)



66: n = 1
BB5-PAL: n = 3

n = 1:

Crude diaziridine **65** (109 mg, 0.649 mmol, 1.0 eq.) in MeOH (0.7 mL) and TEA (145 µL, 1.04 mmol, 1.6 eq.) were stirred at 0 °C using brown-glass ware. In the dark, a 0.1 M solution of iodine (275 mg, 1.04 mmol, 1.6 eq.) in MeOH (10.4 mL) was added dropwise to the mixture until the yellow-brown color of the iodine would stay for longer than ~1 min. The solvent was removed under reduced pressure in the cold. The crude was re-dissolved in water (40 mL) and the aqueous phase was extracted with EtOAc (2 x 40 mL). The combined organics were washed with brine (40 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo* (cold). The crude product was purified by flash column chromatography (10 g, DCM:MeOH, 9:1) under strict exclusion of light to give the product (**66**) as a yellow oil (44 mg, 0.27 mmol; 40% yield over 2 steps from **57**). $\mathbf{R_f}$ (TLC) = 0.69 (DCM:MeOH, 3:2); **MS** (ESI) m/z calc. for $C_8H_{10}N_2O_2$ $[M+H]^+$ 167.08; found 166.98; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.64 (t, *J* = 7.1 Hz, 1 H, 1'-H), 1.80 (t, *J* = 7.8 Hz, 2 H, 3-H), 1.97-2.02 (m, 3 H, 2'-H, 4'-H), 2.16 (t, *J* = 7.4 Hz, 2 H, 2-H); **^{13}C NMR** (126 MHz, $CDCl_3$)

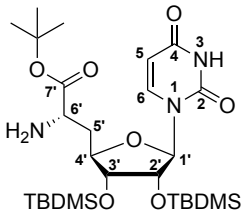
δ 13.22 (C-2'), 27.44 (C-2), 27.75 (C-1'), 28.05 (C-3), 32.18 (CN_2), 69.29 (C-4'), 82.52 (C-3'), 177.56 (C-1); **UV/Vis** (MeOH) λ_{max} = 226, 345; **IR** (ATR) $\tilde{\nu}$ = 3295, 2924, 2856, 1709, 1586, 1433, 1260, 1212, 932, 803, 638.

n = 3:

Crude diaziridine **34** (108 mg, 0.549 mmol, 1.0 eq.) in MeOH (0.41 mL) and TEA (122 µL, 0.878 mmol, 1.6 eq.) were stirred at 0 °C using brown-glass ware. In the dark, a 0.1 M solution of iodine (223 mg, 0.878 mmol, 1.6 eq.) in MeOH (8.80 mL) was added dropwise to the mixture until the yellow-brown color of the iodine would stay for longer than ~1 min. The solvent was removed under reduced pressure in the cold. The crude was re-dissolved in water (40 mL) and the aqueous phase was extracted with EtOAc (2 x 40 mL). The combined organics were washed with brine (40 mL), dried over Na_2SO_4, filtered, and the solvent was removed *in vacuo* (cold). The crude product was purified by flash column chromatography (10 g, DCM:MeOH, 9:1) under strict exclusion of light to give the product (**BB5-PAL**) as a brown oil (28 mg, 0.14 mmol; 26% yield over 2 steps from **35**). $\mathbf{R_f}$ (TLC) = 0.24 (DCM:MeOH, 9:1); **HRMS** (ESI) m/z calc. for $C_{10}H_{14}N_2O_2$ $[M-H]^-$ 193.0938; found 193.0956; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 1.17-1.24 (m, 2 H, 1'-H), 1.39-1.48 (m, 4 H, 2'-H, 3'-H), 1.72 (t, *J* = 7.6 Hz, 2 H, 3-H), 1.92 (t, *J* = 2.5 Hz, 1 H, 6'-H), 2.11-2.15 (m, 4 H, 4'-H, 2-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ 18.39 (C-4'), 23.12 (C-3'), 28.01 (C-2, C-2'), 28.14 (C-1'), 28.57 (C-3), 32.45 (CN_2), 68.89 (C-6'), 84.09 (C-5'), 178.41 (C-1); **IR** (ATR) $\tilde{\nu}$ = 3299, 2935, 2862, 1711, 1584, 1433, 1289, 1219, 917, 838, 639.

*Synthesis of Final Compound **79***

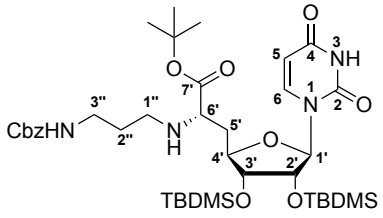
(6'*S*)-Uridinyl amino acid (70)



70

The Cbz group in intermediate **BB1** (480 mg, 0.667 mmol, 1.0 eq.) was cleaved following Method A. After 22 hours, the reaction was filtered through a syringe filter, which was washed several times with isopropanol. The solvent of the filtrate was removed *in vacuo* to give the product (**70**) as a white foam (400 mg; 100% = 391 mg). **HRMS** (ESI) m/z calc. for $C_{27}H_{51}N_3O_7Si_2$ $[M+H]^+$ 586.3338; found 586.3329; **UV/Vis** (MeOH) λ_{max} = 262; **IR** (ATR) $\tilde{\nu}$ = 3206, 2954, 2929, 2895, 2857, 1693, 1461, 1369, 1253, 1157, 1115, 1090, 1059, 868, 837, 813, 777.

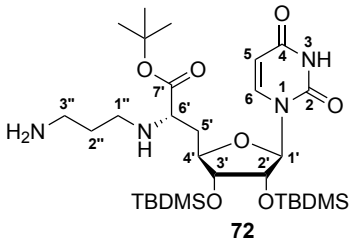
Cbz-protected (6'*S*)-uridinyl amino acid (71)



71

Amine **70** (360 mg, 0.614 mmol, 1.0 eq.) was dissolved in dry THF (16 mL), including molecular sieve (pore size = 4 Å), under nitrogen atmosphere. The mixture was stirred for about 10 min at room temperature. Aldehyde **BB2** (146 mg, 0.706 mmol, 1.15 eq.) – kindly provided by G. Niro in the lab – was added to the mixture. After stirring the reaction overnight at room temperature, $NaBH(AcO)_3$ (260 mg, 1.23 mmol, 2.0 eq.) and Amberlyst® 15 (42 mg, 0.14 mmol, 0.22 eq.) were added and the mixture was stirred for another 21 hours at room temperature. TLC control confirmed conversion and the reaction was filtered. The filter was washed with EtOAc (3 x 50 mL), and the organic layer was washed with saturated, aqueous Na_2CO_3 solution (200 mL). The aqueous layer was again extracted with EtOAc (150 mL), and the combined organics were dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude was purified by flash column chromatography (75 g; DCM:MeOH, 98:2) to give the product (**71**) as a white foam (314 mg, 0.404 mmol; 66% yield). $\mathbf{R_f}$ (TLC) = 0.34 (DCM:MeOH, 96:4); **MS** (ESI) m/z calc. for $C_{38}H_{64}N_4O_9Si_2$ $[M+H]^+$ 777.42; found 777.59; **HRMS** (ESI) m/z calc. for $C_{38}H_{64}N_4O_9Si_2$ $[M+H]^+$ 777.4285; found 777.4271; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.04 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.07 (s, 3 H, $SiCH_3$), 0.87 (s, 9 H, $SiC(CH_3)_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 1.44 (s, 9 H, $OC(CH_3)_3$), 1.59-1.70 (m, 2 H, 2''-H), 1.77-1.84 (m, 1 H, 5'-H_a), 1.92-1.97 (m, 1 H, 5'-H_b), 2.49-2.54 (m, 1 H, 1''-H_a), 2.69-2.74 (m, 1 H, 1''-H_b), 3.19-3.38 (m, 3 H, 3''-H, 6'-H), 3.63-3.66 (m, 1 H, 3'-H), 4.06-4.11 (m, 1 H, 4'-H), 4.24-4.26 (m, 1 H, 2'-H), 5.04-5.17 (m, 2 H, Cbz-CH_2), 5.53 (d, J = 3.8 Hz, 1 H, 1'-H), 5.69 (d, J = 7.5 Hz, 1 H, 5-H), 7.27-7.35 (m, 6 H, 6-H, Cbz-aryl-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.55 ($SiCH_3$), -4.55 ($SiCH_3$), -4.35 ($SiCH_3$), -3.93 ($SiCH_3$), 18.20 ($SiC(CH_3)_3$), 18.30 ($SiC(CH_3)_3$), 26.00 ($SiC(CH_3)_3$), 26.06 ($SiC(CH_3)_3$), 28.32 ($OC(CH_3)_3$), 29.94 (C-2''), 37.41 (C-5'), 39.63 (C-3''), 45.99 (C-1''), 60.07 (C-6'), 66.73 (Cbz-CH_2), 75.74 (C-2'), 75.52 (C-3'), 81.35 (C-4'), 82.77 ($OC(CH_3)_3$), 92.67 (C-1'), 102.34 (C-5), 128.27, 128.28, 128.72 (Cbz-aryl-C), 136.94 (Cbz-C_q), 140.78 (C-6), 150.30 (C-2), 156.80 (Cbz-C=O), 163.00 (C-4), 175.40 (C-7'); **UV/Vis** (MeOH) λ_{max} = 262.

Cbz-deprotected (6'*S*)-uridinyl amino acid (72)



72

First Attempt: The Cbz group in intermediate **71** (10 mg, 0.013 mmol, 1.0 eq.) was attempted to be cleaved following Method A. However, no conversion was observed after 24 hours, several work-up and new set-up cycles, and careful warming of the reaction to 27 °C.

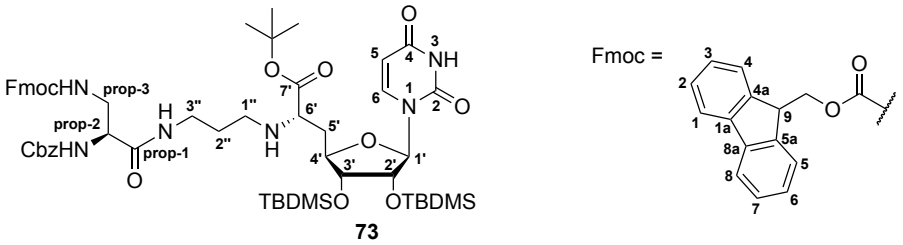
Second Attempt: The Cbz group in intermediate **71** (33 mg, 0.043 mmol, 1.0 eq.) was attempted to be cleaved following Method C. After overnight reaction, partial conversion was observed by TLC control. The reaction was gently warmed to 25 °C and stirred for another 2.5 hours and no more conversion was observed. However, side product formation was identified. The crude was filtered through a syringe filter and the solvent was removed *in vacuo*. The side products were separated from the product by column chromatography using a small pad of silica gel (DCM:MeOH, 3:2). Only 2.4 mg (9% yield) product (**72**) could be isolated. Although the side product could not be identified, it would pose the majority of the conversion.

Third Attempt: The Cbz group in intermediate **71** (9.0 mg, 0.012 mmol, 1.0 eq.) was cleaved following Method B. After overnight conversion, almost all starting material was deprotected as observed by TLC control. For complete conversion, more Pd/C (10% loading) was added and stirred for another 24 hours. The reaction was filtered through a syringe filter, which was washed several times with MeOH. The solvent of the filtrate was removed *in vacuo* to give the product (**72**) as a white foam (6.7 mg; 100% = 7.5 mg).

Scale-Up: The Cbz group in intermediate **71** (30 mg, 0.039 mmol, 1.0 eq.) was cleaved following Method B. After 2 hours, the reaction was filtered through a syringe filter, which was washed several times with MeOH. The solvent was removed *in vacuo* to give the product (**72**) as a white foam (25 mg; 100% = 25 mg). **MS** (ESI) m/z calc. for $C_{30}H_{58}N_4O_7Si_2$ $[M+H]^+$ 643.39; found 643.35; **HRMS** (ESI) m/z calc. for $C_{30}H_{58}N_4O_7Si_2$ $[M+H]^+$ 643.3917; found 643.3907; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.04 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.06 (s, 3 H, $SiCH_3$), 0.86 (s, 9 H, $SiC(CH_3)_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 1.45 (s, 9 H, $OC(CH_3)_3$), 1.59-1.65 (m, 2 H, 2"-H), 1.85-1.93 (m, 2 H, 5'-H), 2.52-2.57 (m, 1 H, 1"-H_a), 2.69-2.74 (m, 1 H, 1"-H_b), 2.78-2.86 (m, 2 H, 3"-H), 3.31 (dd, *J* = 7.0, 7.0 Hz, 1 H, 6'-H), 3.65-3.70 (m, 1 H, 3'-H), 4.07-4.11 (m, 1 H, 4'-H), 4.29-4.32 (m, 1 H, 2'-H), 5.53 (d, *J* = 3.7 Hz, 1 H, 1'-H), 5.72 (d, *J* = 8.4 Hz, 1 H, 5-H), 7.32 (d, *J* = 7.9 Hz, 1 H, 6-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.82 ($SiCH_3$), -4.80 ($SiCH_3$), -4.64 ($SiCH_3$), -4.23 ($SiCH_3$), 17.94 ($SiC(CH_3)_3$), 18.04 ($SiC(CH_3)_3$), 25.75

(SiC($\underline{C}H_3)_3$), 25.80 (SiC($\underline{C}H_3)_3$), 28.08 (OC($\underline{C}H_3)_3$), 29.67 (C-2''), 37.29 (C-5'), 40.43 (C-3''), 46.20 (C-1''), 59.51 (C-6'), 74.30 (C-2'), 75.38 (C-3'), 81.21 (C-4'), 81.61 (O$\underline{C}(CH_3)_3$), 92.68 (C-1'), 102.16 (C-5), 140.86 (C-6), 154.04 (C-2), 163.35 (C-4), 173.85 (C-7'); **UV/Vis** (MeOH) λ_{max} = 262; **IR** (ATR) $\tilde{\nu}$ = 3279, 2953, 2928, 2856, 1694, 1462, 1389, 1368, 1254, 1158, 116, 1086, 1056, 869, 838, 778.

N_α-Cbz-N_β-Fmoc-protected 2,3-diaminopropionamide-extended (6'*S*)-uridinyl amino acid (73)



73

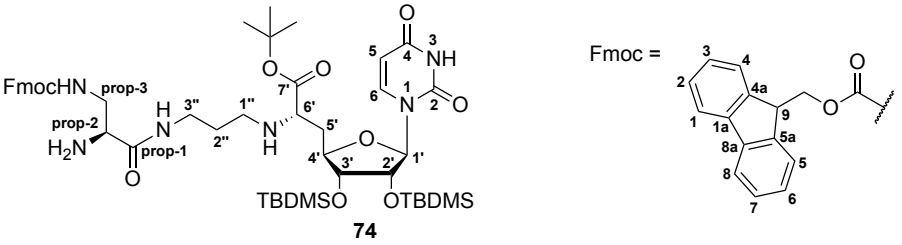
First Attempt: Amine **72** (4.2 mg, 0.0065 mmol, 1.0 eq.) was coupled with carboxylic acid **BB3** (3 mg, 0.007 mmol, 1.05 eq.) following Method D. However, no conversion was observed after 24 hours.

Second Attempt: Amine **72** (7.5 mg, 0.0116 mmol, 1.0 eq.) was coupled with carboxylic acid **BB3** (6 mg, 0.013 mmol, 1.15 eq.) following Method E. After 24 hours, the solvent was removed under reduced pressure and the crude was purified by flash column chromatography (6 g; DCM:MeOH, 98:2) to give the product (**73**) as a white foam (3.8 mg, 0.0035 mmol; 30% yield).

Scale-Up: Amine **72** (71 mg, 0.11 mmol, 1.0 eq.) was coupled with carboxylic acid **BB3** (59 mg, 0.13 mmol, 1.15 eq.) following Method E. After 5.5 hours, the solvent was removed *in vacuo* and the crude was purified by flash column chromatography (20 g; DCM:MeOH, 98:2) to give the product (**73**) as a white foam (89 mg, 0.082 mmol; 74% yield). $\mathbf{R}_f$ (TLC) = 0.11 (DCM:MeOH, 98:2); **MS** (ESI) m/z calc. for $C_{56}H_{80}N_6O_{12}Si_2$ $[M+H]^+$ 1085.54; found 1085.25; **HRMS** (ESI) m/z calc. for $C_{56}H_{80}N_6O_{12}Si_2$ $[M+H]^+$ 1085.5446; found 1085.5432; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.03 (s, 3 H, $SiCH_3$), 0.04 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.82 (s, 9 H, SiC$(CH_3)_3$), 0.87 (s, 9 H, SiC$(CH_3)_3$), 1.43 (s, 9 H, OC$(CH_3)_3$), 1.52-1.66 (m, 2 H, 2''-H), 1.70-1.75 (m, 1 H, 5'-H_a), 1.89-2.09 (m, 1 H, 5'-H_b), 2.47-2.57 (m, 1 H, 1''-H_a), 2.67-2.73 (m, 1 H, 1''-H_b), 3.25-3.42 (m, 3 H, 3''-H, 6'-H), 3.56-3.66 (m, 3 H, 3'-H, prop-3-H), 3.94-4.00 (m, 1 H, 4'-H), 4.32-4.40 (m, 3 H, prop-2-H, Fmoc-CH_2), 4.52-4.59 (m, 1 H, 2'-H), 5.00 (d, *J* = 12.0 Hz, 1 H, Cbz-CH_{2a}), 5.17 (d, *J* = 12.0 Hz, 1 H, Cbz-CH_{2b}), 5.35 (s, 1 H, 1'-H), 5.67 (d, *J* = 10.5 Hz, 2 H, 5-H), 6.27 (d, *J* = 7.4 Hz, 1 H, Fmoc-9-H), 7.23-7.30 (m, 8 H, 6-H, Cbz-aryl-H, Fmoc-3-H, Fmoc-6-H), 7.36 (ddd, *J* = 7.3, 7.3, 2.4 Hz, 2 H, Fmoc-2-H, Fmoc-7-H), 7.55 (m, 2 H, Fmoc-4-H, Fmoc-5-H), 7.73 (d, *J* = 7.3 Hz, 2 H, Fmoc-1-H, Fmoc-8-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.85 ($SiCH_3$), -4.75 ($SiCH_3$), -4.71 ($SiCH_3$), -4.33 ($SiCH_3$), 17.86 (Si$\underline{C}(CH_3)_3$), 18.00 (Si$\underline{C}(CH_3)_3$), 25.71 (SiC($\underline{C}H_3)_3$), 25.77 (SiC($\underline{C}H_3)_3$), 28.09 (OC($\underline{C}H_3)_3$), 28.36 (C-2''), 37.00 (C-5'), 39.03 (C-3''), 43.27 (C-prop-3), 46.27 (C-1''), 47.07 (C-Fmoc-9), 55.74 (C-prop-2), 59.83 (C-6'), 67.17 (Fmoc-CH_2, Cbz-CH_2), 72.81

(C-2'), 75.60 (C-3'), 81.73 (C-4'), 82.40 (O<u>C</u>(CH$_3$)$_3$), 94.37 (C-1'), 102.27 (C-5), 119.92 (Fmoc-C-1, Fmoc-C-8), 125.09, 125.16 (Fmoc-C-4, Fmoc-C-5), 127.06, 127.66, 127.69, 128.24, 128.49, 136.14 (Cbz-aryl, Fmoc-C-2, Fmoc-C-3, Fmoc-C-6, Fmoc-C7), 141.22, 141.24 (C-Fmoc-4$_a$, C-Fmoc-5$_a$), 142.39 (C-6), 143.72, 143.78 (C-Fmoc-1$_a$, C-Fmoc-8$_a$), 150.07 (C-2), 156.78 (Cbz-C=O), 157.45 (Fmoc-C=O), 163.12 (C-4), 169.64 (C-prop-1), 173.69 (C-7'); **IR** (ATR) $\tilde{\nu}$ = 3296, 2952, 2928, 2856, 1696, 1529, 1461, 1389, 1368, 1253, 1155, 114, 1086, 1068, 868, 839, 778, 740, 697, 674.

N$_\beta$-Fmoc-protected 2,3-diaminopropionamide-extended (6'*S*)-uridinyl amino acid (74)



74

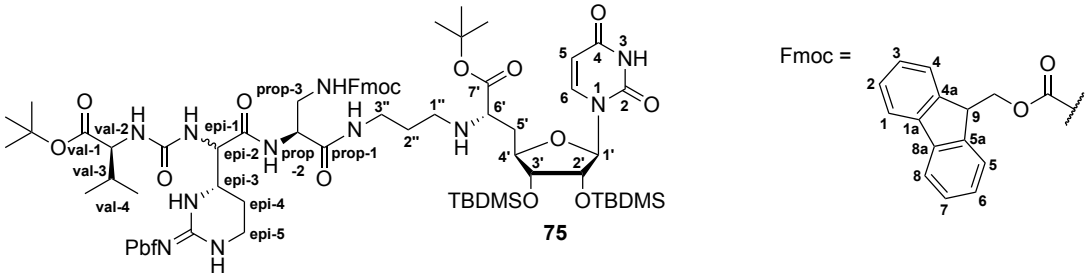
First Attempt: The Cbz group in intermediate **73** (3.8 mg, 0.0035 mmol, 1.0 eq.) was cleaved following Method B. However, no conversion was observed, even after adding more Pd/C (10% loading) or warming the reaction to 25 °C. After addition of Pd black (one small spatula tip), full conversion was observed. The mixture was filtered through a syringe filter and the solvent was removed *in vacuo* to yield the product (**74**) (2.2 mg; 100% = 3.3 mg).

Second Attempt: The Cbz group in intermediate **73** (26 mg, 0.024 mmol, 1.0 eq.) was cleaved following Method C. However, the mixture showed only partial conversion, but could be fully converted after setting up the reaction with the crude again. The mixture was filtered through a syringe filter and the solvent was removed *in vacuo* to yield the product (**74**) (11 mg; 100% = 22 mg).

Scale-Up: The Cbz group in intermediate **73** (164 mg, 0.151 mmol, 1.0 eq.) was cleaved following Method C. After 1 hour, the reaction was filtered through a syringe filter, which was washed several times with MeOH. The solvent of the filtrate was removed *in vacuo* to give the product (**74**) as a white foam (149 mg; 100% = 144 mg). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.03 (s, 3 H, $SiCH_3$), 0.03 (s, 3 H, $SiCH_3$), 0.04 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.85 (s, 9 H, $SiC(CH_3)_3$), 0.88 (s, 9 H, $SiC(CH_3)_3$), 1.44 (s, 9 H, $OC(CH_3)_3$), 1.59-1.67 (m, 2 H, 2''-H), 1.84-1.89 (m, 1 H, 5'-H$_a$), 1.90-1.97 (m, 1 H, 5'-H$_b$), 2.50-2.56 (m, 1 H, 1''-H$_a$), 2.66-2.71 (m, 1 H, 1''-H$_b$), 3.28-3.31 (m, 2 H, 6'-H), 3.35 (d, *J* = 6.6 Hz, 1 H, 3''-H), 3.46-3.52 (m, 2 H, prop-3-H), 3.71 (dd, *J* = 4.6, 4.6 Hz, 1 H, 3'-H), 4.04-4.08 (m, 1 H, 4'-H), 4.19 (d, *J* = 7.2, 7.2 Hz, 1 H, prop-2-H), 4.31-4.43 (m, 3 H, 2'-H, Fmoc-CH_2), 5.43 (d, *J* = 3.7 Hz, 1 H, 1'-H), 5.63-5.67 (m, 1 H, Fmoc-9-H), 5.69 (d, *J* = 8.0 Hz, 1 H, 5-H), 7.26-7.32 (m, 3 H, 6-H, Fmoc-3-H, Fmoc-6-H), 7.33-7.39 (m, 2 H, Fmoc-2-H, Fmoc-7-H), 7.55-7.59 (m, 2 H, Fmoc-4-H, Fmoc-5-H), 7.73 (d, *J* = 7.3 Hz, 2 H, Fmoc-1-H, Fmoc-8-H); **^{13}C NMR** (126 MHz, $CDCl_3$) δ -4.77 ($SiCH_3$), -4.74 ($SiCH_3$), -4.72 ($SiCH_3$), -4.28 ($SiCH_3$), 17.92 (Si<u>C</u>$(CH_3)_3$), 18.03 (Si<u>C</u>$(CH_3)_3$), 25.74 (SiC(<u>C</u>$H_3)_3$), 25.79

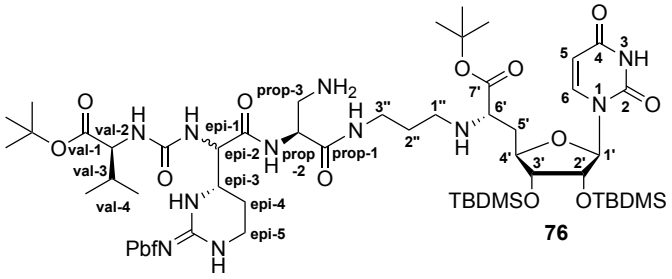
(SiC($\underline{C}H_3)_3$), 28.08 (OC($\underline{C}H_3)_3$), 29.35 (C-2''), 37.23 (C-5'), 37.59 (C-3''), 45.25 (C-prop-3), 45.59 (C-1''), 47.20 (C-Fmoc-9), 59.43 (C-6'), 66.82 (Fmoc-CH_2), 73.61 (C-2'), 75.50 (C-3'), 81.61 (C-4'), 81.78 (O$\underline{C}(CH_3)_3$), 95.00 (C-1'), 102.20 (C-5), 119.93 (Fmoc-C-1, Fmoc-C-8), 123.97, 125.07 (Fmoc-C-4, Fmoc-C-5), 126.87, 127.02, 127.67 (Fmoc-C-2, Fmoc-C-3, Fmoc-C-6, Fmoc-C-7), 141.27 (C-Fmoc-4_a, C-Fmoc-5_a), 143.87 (C-Fmoc-1_a, C-Fmoc-8_a), 143.87 (C-6), 150.05 (C-2), 157.55 (Fmoc-C=O), 163.12 (C-4), 169.70 (C-prop-1), 173.87 (C-7'); **IR** (ATR) $\tilde{\nu}$ = 3296, 2952, 2928, 2856, 1696, 1529, 1461, 1389, 1368, 1253, 1155, 114, 1086, 1068, 868, 839, 778, 740, 697, 674.

Fully protected muraymycin core structure (75)



Amine **74** (40 mg, 0.038 mmol, 1.0 eq.) was coupled with carboxylic acid **BB4** (31 mg, 0.049 mmol, 1.15 eq.) following Method E. After 19 hours, the solvent was removed *in vacuo* and the crude was purified by flash column chromatography (12 g; DCM:MeOH, 98:2, then 9:1) to give the product (**75**) as a white foam (40 mg, 0.026 mmol; 68% yield). NMR analytics in later steps confirmed that **75** was obtained as an epimeric mixture (C-epi-2). $\mathbf{R}_f$ (TLC) = 0.29 (DCM:MeOH, 98:2); **IR** (ATR) $\tilde{\nu}$ = 3342, 2966, 2930, 2857, 1717, 1651, 1607, 1566, 1460, 1370, 1260, 1155, 1092, 1033, 841, 619.

Fmoc-deprotected muraymycin core structure (76)



First Attempt: Fmoc deprotection of muraymycin core structure **75** (34 mg, 0.022 mmol, 1.0 eq.) was attempted by dissolving the intermediate in a 25% piperidine in DCM solution (1.0 mL) and stirring at room temperature for 1.5 hours. TLC control showed conversion of the starting material. Chloroform (10 mL) was added and the organic layer was washed with saturated, aqueous Na_2CO_3 (2 x 5 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude was purified by flash column chromatography (3 g; DCM:MeOH, 99:1, then 95:5, then 94:6) to give the product (**76**) as a white foam (11 mg, 0.0082 mmol; 38% yield).

Second Attempt: Fmoc deprotection of muraymycin core structure **75** (11 mg, 0.0071 mmol, 1.0 eq.) was attempted by dissolving the intermediate in a 10% piperidine in DCM solution (1.0 mL) and stirring at room temperature for 1.5 hours. TLC control showed conversion of the starting material, but slower as compared to the first method. The solvent was removed under reduced pressure and the crude product was purified by flash column chromatography (3 g; DCM:MeOH, 99:1, then 9:1) to give the product (**76**) as an impure mixture. The collected fractions were lyophilized due to residual base, and the crude was again purified by column chromatography (3 g; DCM:MeOH, 99:1) to give the product (**76**) (2.1 mg, 0.0016 mmol; 23%yield).

Third Attempt: Fmoc deprotection of muraymycin core structure **75** (11 mg, 0.0071 mmol, 1.0 eq.) was attempted by dissolving the intermediate in a 1:1-mixture of DIEA in DCM solution (1.0 mL) and stirring at room temperature for up to 3 days. However, no conversion was observed by TLC control.

Fourth Attempt: Fmoc deprotection of muraymycin core structure **75** (11 mg, 0.0071 mmol, 1.0 eq.) was attempted by dissolving the intermediate in a 1:1-mixture of DEA in DCM solution (1.0 mL) and stirring at room temperature for 3 hours. However, side product formation was observed. The solvent was removed under reduced pressure and the crude was purified by column chromatography (3 g, DCM:MeOH, 9:1), but only traces of product (**76)** could be isolated (<1 mg).

Fifth Attempt: Fmoc deprotection of muraymycin core structure **75** (7.0 mg, 0.0045 mmol, 1.0 eq.) was attempted by dissolving the intermediate in DCM (1.0 mL), followed by the addition of polymer-bound piperazine (1.0-2.0 mmol/g; 27 g). However, no conversion was observed by TLC control after 24 hours.

Given the above, the method described in the first deprotection attempt showed the best conversion and yield. Nevertheless, different work-up conditions were investigated to potentially increase the yields.

Work-up 1: The solvent was removed *in vacuo* and the crude was co-evaporated using toluene, acetonitrile, and DCM. Following column chromatography (3 g, DCM:MeOH, 9:1), only mixed fractions with side product (**76**) could be isolated.

Work-up 2: Chloroform (10 mL) was added and the organic layer was washed with saturated, aqueous Na_2CO_3 (2 x 5 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude was co-evaporated with toluene, acetonitrile, and DCM until completely dried (white foam). It was continued without further purification. However, the follow-up amide coupling failed due to piperidine residues in the crude.

Work-up 3: Chloroform (10 mL) was added and the organic layer was washed with saturated, aqueous Na_2CO_3 (2 x 5 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude was co-

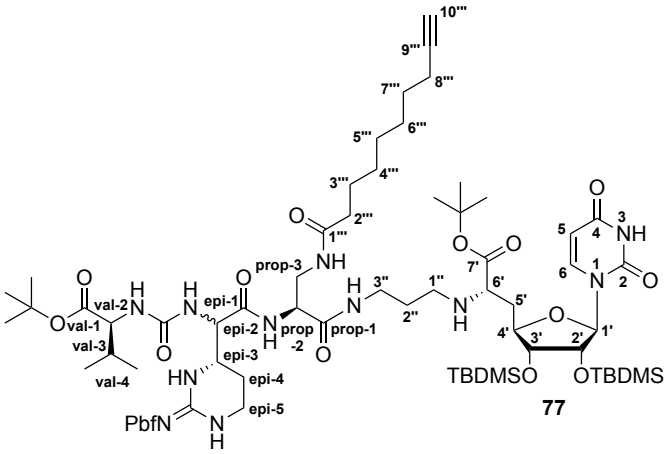
evaporated with DCM and lyophilized twice. However, piperidine traces could still be found in TLC control of the crude. After column purification (3 g, DCM, 100%, then DCM:MeOH, 4:1), pure product (**76**) (49% yield) could be obtained.

Work-up 4: The solvent was removed under reduced pressure and the crude was co-evaporated with toluene and acetonitrile. The crude was re-dissolved in chloroform (25 mL) and the organic layer was washed with saturated, aqueous Na_2CO_3 (2 x 10 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. After column purification (3 g, DCM, 100%, then DCM:MeOH, 4:1), pure product (**76**) (67% yield) could be obtained.

Given the above, work-up 4 seemed to furnish the best results.

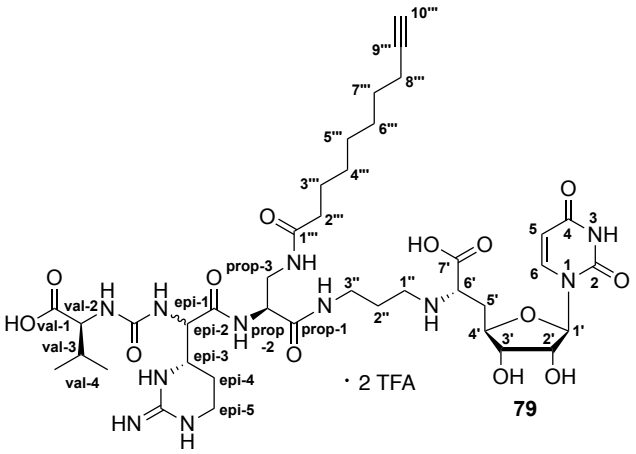
Combined: Fmoc deprotection of muraymycin core **75** (35 mg, 0.023 mmol, 1.0 eq.) was attempted by dissolving the intermediate in a 25% piperidine in DCM solution (1.0 mL) and stirring at room temperature for 1 hour. TLC control showed conversion of the starting material. The solvent was removed under reduced pressure and the crude was co-evaporated with toluene and acetonitrile. The crude was re-dissolved in chloroform (25 mL) and the organic layer was washed with saturated, aqueous Na_2CO_3 (2 x 10 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography (3 g; DCM:MeOH, 9:1) to give the product (**76**) as a white foam (22 mg, 0.016 mmol; 73% yield). $\mathbf{R_f}$ (TLC) = 0.18 (DCM:MeOH, 9:1); **MS** (ESI) m/z calc. for $C_{62}H_{107}N_{11}O_{15}SSi_2$ $[M+H]^+$ 1334.73; found 1334.99; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 0.02 (s, 3 H, $SiCH_3$), 0.04 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.05 (s, 3 H, $SiCH_3$), 0.81-0.89 (m, 24 H, $SiC(CH_3)_3$, val-4-H), 1.40-1.45 (m, 24 H, 7'-$OC(CH_3)_3$, val-$OC(CH_3)_3$, Pbf-2-$C(CH_3)_2$), 1.53-1.70 (m, 2 H, 2''-H), 1.79-1.84 (m, 2 H, epi-4-H), 1.87-1.96 (m, 1 H, 5'-H_a), 1.98-2.13 (m, 5 H, Pbf-7-CH_3, val-3-H, 5'-H_b), 2.44 (s, 3 H, Pbf-6-CH_3), 2.48-2.58 (m, 4 H, Pbf-4-CH_3, 1''-H_a), 2.64-2.74 (m, 1 H, 1''-H_b), 2.91 (s, 2 H, Pbf-3-H), 3.12-3.24 (m, 2 H, epi-5-H), 3.28-3.47 (m, 6 H, prop-3-H, 3''-H, epi-3-H, 6'-H), 3.64-3.71 (m, 1 H, epi-2-H), 3.78-3.82 (m, 1 H, 3'-H), 3.99-4.06 (m, 1 H, 4'-H), 4.13-4.19 (m, 1 H, val-2-H), 4.34-4.43 (m, 1 H, prop-2-H), 4.63-4.73 (m, 1 H, 2'-H), 5.29-5.40 (m, 1 H, 1'-H), 5.67-5.74 (m, 1 H, 5-H), 7.06-7.30 (m, 1 H, 6-H). (Note: The spectrum resolution in this step did not allow for identification of an epimeric mixture of **76**, which is why the analytical data for both epimers is reported together.)

Fully protected reference probe (77)



Amine **76** (15 mg, 0.011 mmol, 1.0 eq.) was coupled with fatty acid **BB5-ref** (2.2 mg, 0.013 mmol, 1.15 eq.) following Method E. After 21 hours, the solvent was removed *in vacuo* and the crude was purified by flash column chromatography (3 g; DCM:MeOH, 96:4, then 99:1) to give the product (**77**) as a white foam (10 mg, 0.0066 mmol; 59% yield). $\mathbf{R}_f$ (TLC) = 0.43 (DCM:MeOH, 96:4); **MS** (ESI) m/z calc. for $C_{72}H_{121}N_{11}O_{16}SSi_2$ $[M+H]^+$ 1484.83; found 1485.00.

Reference probe (79)



Fully protected muraymycin intermediate **77** (19 mg, 0.013 mmol, 1.0 eq.) was dissolved in an 80% TFA in water solution (5.0 mL). The mixture was stirred at room temperature for 12 hours. The solvent was removed *in vacuo* and the crude was purified by preparative HPLC to give an apparent diastereomeric mixture of the TFA-salt of the product (**79**) as a white solid (4.1 mg, 0.0036 mmol; 36% yield). Note: NMR analysis revealed a mixture of diastereomers (*d.r.* ~ 1.2:1).

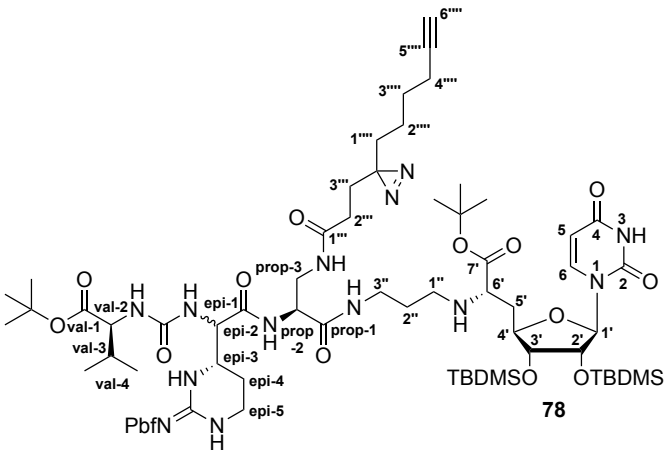
Analytic data of excess epimer:
MS (ESI) m/z calc. for $C_{39}H_{61}N_{11}O_{13}$ $[M+H]^+$ 892.45; found 892.41; **^{1}H NMR** (500 MHz, D_2O) δ 0.83-0.87 (m, 3 H, val-4-H), 0.87-0.90 (m, 3 H, val-4-H), 1.15-1.23 (m, 4 H, 5'''-H,

6'''-H), 1.27-1.30 (m, 2 H, 4'''-H), 1.36-1.42 (m, 2 H, 3'''-H), 1.42-1.48 (m, 2 H, 7'''-H), 1.75-1.88 (m, 4 H, epi-4-H, 2''-H), 2.06-2.14 (m, 4 H, val-3-H, 8'''-H, 2'''-H), 2.14-2.22 (m, 1 H, 5'-H_a), 2.24 (t, *J* = 2.6 Hz, 1 H, 10'''-H), 2.33-2.38 (m, 1 H, 5'-H_b), 2.97-3.04 (m, 2 H, 1''-H), 3.18-3.32 (m, 4 H, epi-5-H, 3''-H), 3.34-3.38 (m, 0.5 H, prop-3-H_a), 3.53-3.57 (m, 0.5 H, prop-3-H_b), 3.77-3.85 (m, 0.5 H, epi-3-H), 3.83-3.87 (m, 1 H, 6'-H), 3.97-4.01 (m, 2 H, val-2-H, 3'-H), 4.04-4.09 (m, 1 H, 4'-H), 4.27 (dd, *J* = 7.6, 2.3 Hz, 0.5 H, epi-2-H), 4.31-4.36 (m, 2 H, prop-2-H, 2'-H), 5.66 (d, *J* = 3.8 Hz, 1 H, 1'-H), 5.79 (d, *J* = 7.3 Hz, 1 H, 5-H), 7.56 (d, *J* = 8.2 Hz, 1 H, 6-H); **^{13}C NMR** (126 MHz, D_2O) δ 16.95, 17.04 (C-val-4), 17.43 (C-8'''), 18.45 (C-val-4), 20.58 (C-epi-4), 25.23 (C-7'''), 25.49 (C-2''), 27.55, 27.58 (C-3'''), 27.70, 27.74, 27.95, 28.05, (C-4'''-6'''), 29.78 (C-val-3'), 32.66 (C-5'), 35.55 (C-2'''), 35.93 (C-epi-5), 36.13 (C-3''), 40.09 (C-prop-3), 44.16, 44.23 (C-1''), 49.63 (C-epi-3), 54.31 (C-prop-2), 55.93 (C-epi-2), 59.00 (C-val-2), 59.41 (C-6'), 68.98 (C-10'''), 72.57, 72.57 (C-2'), 72.87 (C-3'), 79.74 (C-4'), 86.52 (C-9'''), 91.65, 91.68 (C-1'), 102.16 (C-5), 116.28 (q, $^1J_{CF}$ = 291.9 Hz, TFA-$\underline{C}F_3$), 142.73 (C-6), 151.35 (C-2), 153.80, 153.83 (C=N), 159.07 (urea-C=O), 159.36 (C-4), 162.97 (q, $^1J_{CF}$ = 35.6 Hz, TFA-C=O), 166.13 (C-1'''), 171.93 (C-prop-1), 176.54 (C-epi-1), 177.37 (C-7'), 178.16 (C-val-1).

Analytic data of minor epimer:
MS (ESI) m/z calc. for $C_{39}H_{61}N_{11}O_{13}$ $[M+H]^+$ 892.45; found 892.41; **^{1}H NMR** (500 MHz, D_2O) δ 0.83-0.87 (m, 3 H, val-4-H), 0.87-0.90 (m, 3 H, val-4-H), 1.15-1.23 (m, 4 H, 5'''-H, 6'''-H), 1.27-1.30 (m, 2 H, 4'''-H), 1.36-1.42 (m, 2 H, 3'''-H), 1.42-1.48 (m, 2 H, 7'''-H), 1.75-1.88 (m, 4 H, epi-4-H, 2''-H), 2.06-2.14 (m, 4 H, val-3-H, 8'''-H, 2'''-H), 2.14-2.22 (m, 1 H, 5'-H_a), 2.24 (t, *J* = 2.6 Hz, 1 H, 10'''-H), 2.33-2.38 (m, 1 H, 5'-H_b), 2.97-3.04 (m, 2 H, 1''-H), 3.18-3.32 (m, 4 H, epi-5-H, 3''-H), 3.42-3.47 (m, 0.5 H, prop-3-H_a), 3.49-3.53 (m, 0.5 H, prop-3-H_b), 3.71-3.74 (m, 0.5 H, epi-3-H), 3.83-3.87 (m, 1 H, 6'-H), 3.97-4.01 (m, 2 H, val-2-H, 3'-H), 4.04-4.09 (m, 1 H, 4'-H), 4.27 (dd, *J* = 7.6, 2.3 Hz, 0.5 H, epi-2-H), 4.31-4.36 (m, 2 H, prop-2-H, 2'-H), 5.66 (d, *J* = 3.8 Hz, 1 H, 1'-H), 5.79 (d, *J* = 7.3 Hz, 1 H, 5-H), 7.56 (d, *J* = 8.2 Hz, 1 H, 6-H); **^{13}C NMR** (126 MHz, D_2O) δ 16.95, 17.04 (C-val-4), 17.43 (C-8'''), 18.45 (C-val-4), 20.37 (C-epi-4), 25.01 (C-7'''), 25.60 (C-2''), 27.55, 27.58 (C-3'''), 27.70, 27.74, 27.95, 28.05, (C-4'''-6'''), 29.78 (C-val-3'), 32.66 (C-5'), 35.40 (C-2'''), 35.93 (C-epi-5), 36.13 (C-3''), 40.09 (C-prop-3), 44.16, 44.23 (C-1''), 49.85 (C-epi-3), 53.52 (C-prop-2), 55.56 (C-epi-2), 58.92 (C-val-2), 59.28 (C-6'), 68.98 (C-10'''), 72.57, 72.57 (C-2'), 72.87 (C-3'), 79.74 (C-4'), 86.52 (C-9'''), 91.65, 91.68 (C-1'), 102.16 (C-5), 116.28 (q, $^1J_{CF}$ = 291.9 Hz, TFA-$\underline{C}F_3$), 142.73 (C-6), 151.35 (C-2), 153.80, 153.83 (C=N), 159.07 (urea-C=O), 159.36 (C-4), 162.97 (q, $^1J_{CF}$ = 35.6 Hz, TFA-C=O), 166.13 (C-1'''), 171.82 (C-prop-1), 176.59 (C-epi-1), 177.37 (C-7'), 178.16 (C-val-1).

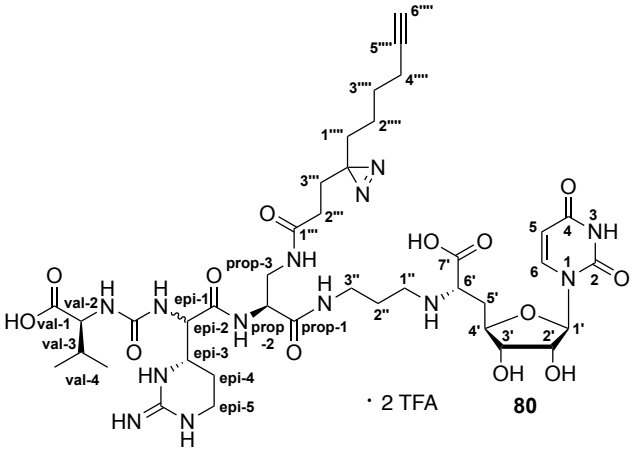
*Attempted Synthesis of Final Compound **80***

Fully protected probe (78)



Amine **76** (42 mg, 0.032 mmol, 1.0 eq.) was coupled with fatty acid **BB5-PAL** (7.1 mg, 0.036 mmol, 1.15 eq.) following Method E. After 12 hours, the solvent was removed *in vacuo* and the crude was purified by flash column chromatography (9.5 g; DCM:MeOH, 99:1) to give the product (**78**) as a white foam (33 mg, 0.022 mmol; 69% yield). The reaction and work-up of this compound were performed in the dark and using brown glass ware. $\mathbf{R}_f$ (TLC) = 0.24 (DCM:MeOH, 99:1); **MS** (ESI) m/z calc. for $C_{72}H_{119}N_{13}O_{16}SSi_2$ $[M+H]^+$ 1510.82; found 1511.09.

Diazirine probe (80)



Fully protected muraymycin intermediate **78** (33 mg, 0.022 mmol, 1.0 eq.) was dissolved in an 80% TFA in water solution (5.0 mL). The mixture was stirred at room temperature for 14 hours. The solvent was removed *in vacuo* and the crude was purified by preparative HPLC to give the TFA-salt of the product as a white solid. However, decomposition of the product was visible during HPLC purification, and even though one fraction with pure product was collected, only insufficient quantities of product **80** could be isolated (<1 mg). The reaction and work-up of this compound were performed in the dark, using brown glass ware. **MS** (ESI) m/z calc. for $C_{39}H_{59}N_{13}O_{13}$ $[M+H]^+$ 918.44; found 918.52.

10.6 References

[1] Pandey, A.; Mann, M. Proteomics to study genes and genomes. *Nature* **2000**, *405* (6788), 837-846.

[2] Norin, M.; Sundström, M. Structural proteomics: developments in structure-to-function predictions. *Trends Biotechnol.* **2002**, *20* (2), 79-84.

[3] Xu, J.; Li, X.; Ding, K.; Li, Z. Applications of Activity-Based Protein Profiling (ABPP) and Bioimaging in Drug Discovery. *Chem. Asian J.* **2020**, *15* (1), 34-41.

[4] Yang, P.; Liu, K. Activity-Based Protein Profiling: Recent Advances in Probe Development and Applications. *ChemBioChem* **2015**, *16* (5), 712-724.

[5] Jeffery, D. A.; Bogyo, M. Chemical proteomics and its application to drug discovery. *Curr. Opin. Biotechnol.* **2003**, *14* (1), 87-95.

[6] Hopf, C.; Bantscheff, M.; Drewes, G. Pathway Proteomics and Chemical Proteomics Team Up in Drug Discovery. *Neurodegener. Disc.* **2007**, *4* (2-3), 270-280.

[7] Jessani, N.; Cravatt, B. F. The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 54-59.

[8] Willems, L. I.; Overkleefts, H. S.; van Kasteren, S. I. Current Developments in Activity-Based Protein Profiling. *Bioconjugate Chem.* **2014**, *25* (7), 1181-1191.

[9] Sutton, C. W. The role of targeted chemical proteomics in pharmacology. *Br. J. Pharmacol.* **2012**, *166* (2), 457-475.

[10] Lapinsky, D. J. Tandem photoaffinity labeling-bioorthogonal conjugation in medicinal chemistry. *Bioorg. Med. Chem.* **2012**, *20* (21), 6237-6247.

[11] Battenberg, O. A.; Nodwell, M. B.; Sieber, S. A. Evaluation of α-Pyrones and Pyrimidones as Photoaffinity Probes for Affinity-Based Protein Profiling. *J. Org. Chem.* **2011**, *76* (15), 6075-6087.

[12] Li, Z.; Qian, L.; Li, L.; Bernhammer, J. C.; Huynh, H. V.; Lee, J.-S.; Yao, S. Q. Tetrazole Photoclick Chemistry: Reinbestigating Its Suitability as a Bioorthogonal Reaction and Potential Applications. *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (6), 2002-2006.

[13] Guo, H.; Li. Z. Developments of bioorthogonal handle-containing photo-crosslinkers for photoaffinity labeling. *Med. Chem. Commun.* **2017**, *8* (8), 1585-1591.

[14] Li, Z.; Hao, P.; Li. L.; Tan, C. Y.; Cheng, X.; Cheng, G. Y.; Sze, S. K.; Shen, H. M.; Yao, S. Q. Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinas inhibitors for cell- and tissue-based proteome profiling. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (33), 8551-8556.

[15] Leriche, G.; Chisholm, L.; Wagner, A. Cleavable linkers in chemical biology. *Bioorg. Med. Chem.* **2012**, *20* (2), 571-582.

[16] Martell, J.; Weerapana, E. Applications of Copper-Catalyzed Click Chemistry in Activity-Based Protein Profiling. *Molecules* **2014**, *19* (2), 1378-1393.

[17] Sieber, S. A.; Cravatt, B. F. Analytical platforms for activity-based protein profiling – exploiting the versatility of chemistry for functional proteomics. *Chem. Commun.* **2006**, (22), 2311-2319.

[18] Reardon, H. T.; Herbst, R. A.; Henry, C. L.; Herbst, D. M.; Ngo, N.; Cisar, J. S.; Weber, O. D.; Niphakis, M. J.; O'Neill, G. P. Quantification of In Vivo Target Engagement Using Microfluidic Activity-Based Protein Profiling. *SLAS Technol.* **2019**, *24* (5), 489-498.

[19] Wang, S.; Tian, Y.; Wang, M.; Wang, M.; Sun, G.-B.; Sun, X.-B. Advanced Activity-Based Protein Profiling Application Strategies for Drug Development. *Front. Pharmacol.* **2018**, *9*, 353.

[20] Chen, X.; Wong, Y. K.; Wang, J.; Zhang, J.; Lee, Y.-M.; Shen, H.-M.; Lin, Q.; Hua, Y.-C. Target identification with quantitative activity based protein profiling (ABPP). *Proteomics* **2017**, *17*, 3-4.

[21] Ong, S.-E.; Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **2006**, *1* (6), 2650-2660.

[22] van Rooden, E. J.; Florea, B. I.; Deng, H.; Baggelaar, M. P.; van Esbroeck, A. C. M.; Zhou, J.; Overkleeft, H. S.; van der Stelt, M. Mapping *in vivo* target interaction profiles of covalent inhibitors using chemical proteomics with label-free quantification. *Nat. Protoc.* **2018**, *13* (4), 752-767.

[23] Zhu, H.; Tamura, T.; Hamachi, I. Chemical proteomics for subcellular proteome analysis. *Curr. Opin. Chem. Biol.* **2019**, *48*, 1-7.

[24] Wiedner, S. D.; Anderson, L. N.; Sadler, N. C.; Chrisler, W. B.; Kodali, V. K.; Smith, R. D.; Wright, A. T. Live Cell Organell-Specific Activity-Based Protein Profiling. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (11), 2919-2922.

[25] Pichler, C. M.; Krysiak, J.; Breinbauer, R. Target identification of covalently binding drugs by activity-based protein profiling (ABPP). *Bioorg. Med. Chem.* **2016**, *24* (15), 3291-3303.

[26] Peters, E. C.; Gray, N. S. Chemical Proteomics Identifies Unanticipated Targets of Clinical Kinase Inhibitors. *ACS Chem. Biol.* **2007**, *2* (10), 661-664.

[27] Lum, K. M.; Sato, Y.; Beyer, B. A.; Plaisted, W. C.; Anglin, J. L.; Lairson, L. L.; Cravatt, B. F. Mapping Protein Targets of Bioactive Small Molecules using Lipid-Based Chemical Proteomics. *ACS Chem. Biol.* **2017**, *12* (10), 2671-2681.

[28] Hoch, D. G.; Abegg, D.; Wang, C.; Shuster, A.; Adibekian, A. Cysteine-specific Chemical Proteomics: From Target Identification to Drug Discovery. *Chimia* **2016**, *70* (11), 764-767.

[29] Whitby, L. R.; Obach, R. S.; Simon, G. M.; Hayward, M. M.; Cravatt, B. F. Quantitative chemical proteomic profiling of the *in vivo* targets of reactive drug metabolites. *ACS Chem. Biol.* **2017**, *12* (8), 2040-2050.

[30] Wright, A. T.; Cravatt, B. F. Chemical proteomic probes for profiling cytochrome P450 activities and drug interactions *in vivo*. *Chem. Biol.* **2007**, *14* (9), 1043-1051.

[31] Jessani, N.; Young, J. A.; Diaz, S. L.; Patricelli, M. P.; Varki, A.; Cravatt, B. F. Class Assignment of Sequence-Unrelated Members of Enzyme Superfamilies by Activity-Based Protein Profiling. *Angew Chem. Int. Ed. Engl.* **2005**, *44* (16), 2400-2403.

[32] Marchenko, M.; Thomson, A.; Ellis, T. N.; Knuckley, B.; Causey, C. P. Development of a clickable activity-based protein profiling (ABPP) probe for agmatine deiminases. *Bioorg. Med. Chem.* **2015**, *23* (9), 2159-2167.

[33] Uttamchandani, M.; Li, J.; Sun, H.; Yao, S. Q. Activity-Based Protein Profiling: New Developments and Directions in Functional Proteomics. *ChemBioChem* **2008**, *9* (5), 667-675.

[34] Barglow, K. T.; Cravatt, B. F. Activity-based protein profiling for the functional annotation of enzymes. *Nat. Methods.* **2007**, *4* (10), 822-827.

[35] Wong, Y.-K.; Zhang, J.; Hua, Z.-C.; Lin, Q.; Shen, H.-M.; Wang, J. Recent advances in quantitative and chemical proteomics for autophagy studies. *Autophagy* **2017**, *13* (9), 1472-1486.

[36] Nomura, D. K.; Dix, M. M.; Cravatt, B. F. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat. Rev. Cancer* **2010**, *10* (9), 630-638.

[37] Morimoto, K.; van der Hoorn, R. A. L. The Increasing Impact of Activity-Based Protein Profiling in Plant Science. *Plant Cell Physiol.* **2016**, *57* (3), 446-461.

[38] Liu, Y.; Fredrickson, J. K.; Sadler, N. C.; Nandhikonda, P.; Smith, R. D.; Wright, A. T. Advancing understanding of microbial bioenergy conversion processes by activity-based protein profiling. *Biotechnol. Biofuels* **2015**, *8*, 156.

[39] Whidbey, C.; Wright, A. T. Activity-Based Protein Profiling – Enabling Multimodal Functional Studies of Microbial Communities. *Curr. Top. Microbiol. Immunol.* **2019**, *420*, 1-12.

[40] Lentz, C. S.; Sheldon, J. R.; Crawford, L. A.; Cooper, R.; Garland, M.; Amieva, M. R.; Weerapana, E.; Skaar, E. P.; Bogyo, M. Identification of a *S. aureus* virulence factor by activity-based protein profiling (ABPP). *Nat. Chem. Biol.* **2018**, *14* (6), 609-617.

[41] Wright, M. H. Chemical Proteomics of Host-Microbe Interactions. *Proteomics* **2018**, *18* (18), e1700333.

[42] Weerapana, E.; Speers, A. E.; Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP) – a general method for mapping sites of probe modification in proteomes. *Nat. Protoc.* **2007**, *2* (6), 1414-1425.

[43] Vodovozova, E. L. Photoaffinity Labeling and Its Application in Structural Biology. *Biochemistry* **2007**, *72* (1), 1-20.

[44] Robinette, D.; Neamati, N.; Tomer, K. B.; Borchers, C. H. Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. *Expert Rev. Proteomics* **2006**, *3* (4), 399-408.

[45] Couvineau, A.; Ceraudo, E.; Tan, Y.-V.; Laburthe, M. VPAC1 receptor binding site: Contribution of photoaffinity labeling approaches. *Neuropeptides* **2010**, *44* (2), 127-132.

[46] Hamouda, A. K.; Jayakar, S. S.; Chiara, D. C.; Cohen, J. B. Photoaffinity Labeling of Nicotinic Receptors: Diversity of Drug Binding Sites! *J. Mol. Neurosci.* **2014**, *53* (3), 480-486.

[47] Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J.-S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2017**, *15*, 14.

[48] Chuang, V. T. G.; Otagiri, M. Photoaffinity Labeling of Plasma Proteins. *Molecules* **2013**, *18* (11), 13831-13859.

[49] Xia, Y.; Peng. L. Photoactivatable Lipid Probes for Studying Biomembranes by Photoaffinity Labeling. *Chem. Rev.* **2013**, *113* (10), 7880-7929.

[50] Lapinsky, D. J. Tandem photoaffinity labeling-bioorthogonal conjugation in medicinal chemistry. *Bioorg. Med. Chem.* **2012**, *20* (21), 6237-6247.

[51] Smith, E.; Collins, I. Photoaffinity labeling in target- and binding-site identification. *Future Med. Chem.* **2015**, *7* (2), 159-183.

[52] Schülke, J.-P.; McAllister, L. A.; Geoghegan, K. F.; Parikh, V.; Chappie, T. A.; Verhoest, P. R.; Schmidt, C. J.; Johnson, D. S.; Brandon, N. J. Chemoproteomics Demonstrates Target Engagement and Exquisite Selectivity of the Clinical Phosphodiesterase 10A Inhibitor M-10 in Its Native Environment. *ACS Chem. Biol.* **2014**, *9* (12), 2823-2832.

[53] MacKinnon, A. L.; Taunton, J. Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr. Protoc. Chem. Biol.* **2009**, *1*, 55-73.

[54] Brunner, J. New Photolabeling And Crosslinking Methods. *Annu. Rev. Biochem.* **1993**, *62*, 483-514.

[55] Deng, Y.; Couch, B. A.; Koleske, A. J.; Turk, B. E. A peptide photoaffinity probe specific for the active conformation of the Abl tyrosine kinase. *ChemBioChem* **2012**, *13* (17), 2510-2512.

[56] Li, G.; Liu, Y.; Yu, X.; Li, X. Multivalent Photoaffinity Probe for Labeling Small Molecule Binding Proteins. *Bioconjugate Chem.* **2014**, *25* (6), 1172-1180.

[57] Tan, X.-D.; Pan, M.; Gao, S.; Zheng, Y.; Shi, J.; Li, Y.-M. A diubiquitin-based photoaffinity probe for profiling K27-linkage targeting deubiquitinases. *Chem. Commun* **2017**, *53* (73), 10208-10211.

[58] Nury, C.; Czarny, B.; Cassar-Lajeunesse, E.; Georgiadis, D.; Bregant, s.; Dive, V. A Pan Photoaffinity Probe for Detecting Active Forms of Matrix Metalloproteinases. *ChemBioChem* **2013**, *14* (1), 107-114.

[59] Sakurai, K.; Ozawa, S.; Yamada, R.; Yasui, T.; Mizuno, S. Comparison of the Reactivity of Carbohydrate Photoaffinity Probes with Different Photoreactive Groups. *ChemBioChem* **2014**, *15* (10), 1399-1403.

[60] Rodriguez, A. C.; Kohler, J. J. Recognition of diazirine-modified O-GlcNAc by human O-GlcNAcase. *Med. Chem. Commun.* **2014**, *5* (8), 1227-1234.

[61] Farley, R. A.; Goldman, D. W.; Bayley, H. Identification of Regions of the Catalytic Subunit of (Na-K)-ATPase Embedded within the Cell Membrane. *J. Biol. Chem.* **1980**, *255* (3), 860-864.

[62] Masuda, K.; Koizumi, A.; Misaka, T.; Hatanaka, Y.; Abe, K.; Tanaka, T.; Ishiguro, M.; Hashimoto, M. Photoactive ligands probing the sweet taste receptor. Design and synthesis of highly potent diazirinyl D-phenylalanine derivatives. *Bioorg. Med. Chem. Lett.* **2010**, *20* (3), 1081-1083.

[63] Standring, D. N.; Knowles, J. R. Photoaffinity Labeling of Lactate Dehydrogenase by the Carbene Derived from the 3-Diazirino Analogue of Nicotinamide Adenine Dinucleotide. *Biochemistry* **1980**, *19* (12), 2811-2816.

[64] Husain, S. S.; Nirthanan, S.; Ruesch, D.; Solt, K.; Cheng, Q.; Li, G.-D.; Arevalo, E.; Olsen, R. W.; Raines, D. E.; Forman, S. A.; Cohen, J. B.; Miller, K. W. Synthesis of Trifluoromethylaryl Diazirine and Benzophenone Derivatives of Etomidate that Are Potent General Anesthetics and Effective Photolabels for Probing Sites on Ligand-Gated Ion Channels. *J. Med. Chem.* **2006**, *49* (16), 4818-4825.

[65] Chee, G.-L.; Yalowich, J. C.; Bodner, A.; Wu, X.; Hasinoff, B. B. A diazirine-based photoaffinity etoposide probe for labeling topoisomerase II. *Bioorg. Med. Chem.* **2010**, *18* (2), 830-838.

[66] Dubinsky, L.; Delago, A.; Amara, N.; Krief, P.; Rayo, J.; Zor, T.; Kravchenko, V. V.; Meijler, M. M. Species selective diazirine positioning in tag-free photoactive quorum sensing probes. *Chem. Commun.* **2013**, *49* (52), 5826-5828.

[67] Das, J. Identification of alcohol-binding site(s) in proteins using diazirine-based photoaffinity labeling and mass spectrometry. *Chem. Biol. Drug Des.* **2019**, *93* (6), 1158-1165.

[68] Kuboe, S.; Yoda, M.; Ogata, A.; Kitade, Y.; Tomari, Y.; Ueno, Y. Diazirine-containing RNA photocrosslinking probes for the study of siRNA-protein interactions. *Chem. Commun.* **2010**, *46* (39), 7367-7369.

[69] Nakamoto, K.; Ueno, Y. Diazirine-Containing RNA Photo-Cross-Linking Probes for Capturing microRNA Targets. *J. Org. Chem.* **2014**, *79* (6), 2463-2472.

[70] Nakamoto, K.; Akao, Y.; Ueno, Y. Diazirine-containing tag-free RNA probes for efficient RISC-loading and photoaffinity labeling of microRNA targets. *Bioorg. Med. Chem. Lett.* **2018**, *28* (17), 2906-2909.

[71] Song, C.-X.; He, C. Bioorthogonal Labeling of 5-Hydroxymethylcytosine in Genomic DNA and Diazirine-Based DNA Photo-Cross-Linking Probes. *Acc Chem. Res.* **2011**, *44* (9), 709-717.

[72] Zhang, H.; Song, Y.; Zou, Y.; Ge, Y.; An, Y.; Ma, Y.; Zhu, Z.; Yang, C. J. A diazirine-based photoaffinity probe for facile and efficient aptamer-protein covalent conjugation. *Chem. Commun.* **2014**, *50* (38), 4891-4894.

[73] Soethoudt, M.; Alachouzos, G.; van Rooden, E. J.; Moya-Garzón, M. D.; van den Berg, R. J.B.H.N.; Heitman, L. H.; van der Stelt, M. Development of a Cannabinoid-Based Photoaffinity Probe to Determine the $\Delta^{8/9}$-Tetrahydrocannabinol Protein Interaction Landscape in Neuroblastoma Cells. *Cannabis Cannabinoid Res.* **2018**, *3* (1), 136-151.

[74] Yang, T.; Liu, Z.; Li, X. D. Developing diazirine-based chemical probes to identify histone modification 'readers' and 'erasers'. *Chem. Sci.* **2015**, *6* (2), 1011-1017.

[75] Müskens, F. M.; Ward, R. J.; Herkt, D.; van de Langemheen, H.; Tobin, A. B.; Liskamp, R. M. J.; Miligan, G. Design, Synthesis, and Evaluation of a Diazirine Photoaffinity Probe for Ligand-Based Receptor Capture Targeting G Protein-Coupled Receptors. *Mol. Pharmacol.* **2019**, *95* (2), 196-209.

[76] Vervacke, J. S.; Funk, A. L.; Wang, Y.-C.; Strom, M.; Hrycyna, C. A.; Distefano, M. D. Diazirine-containing photoactivatable isoprenoid: synthesis and application in studies with isoprenylcysteine carboxyl methyltransferase. *J. Org. Chem.* **2014**, *79* (5), 1971-1978.

[77] Lachaal, M.; Rampal, A. L.; Lee, W.; Shi, Y.-w.; Jung, C. Y. GLUT1 Transmembrane Glucose Pathway. Affinity labeling with a transportable D-glucose diazirine. *J. Biol. Chem.* **1996**, *271* (9), 5225-5230.

[78] Thuy-Boun, P. S.; Wolan, D. W. A glycal-based photoaffinity probe that enriches sialic acid binding proteins. *Bioorg. Med. Chem. Lett.* **2019**, *29* (18), 2609-2612.

[79] Gunaratne, G. S.; Su, P.; Marchant, J. S.; Slama, J. T.; Walseth, T. F. 5-Azido-8-ethynyl-NAADP: A bifunctional, clickable photoaffinity probe for the identification of NAADP receptors. *Biochim. Biophys. Acta Mol. Cell Res.* **2019**, *1866* (7), 1180-1188.

[80] Huang, W.; Sun, W.; Song, Z.; Yu, Y.; Chen, X.; Zhang, Q. Incorporation of a fluorous diazirine group into phosphatidylinositol 4,5-bisphosphate to illustrate its interaction with ADP-ribosylation factor 1. *Org. Biomol. Chem.* **2012**, *10* (27), 5197-5201.

[81] An, J.; Zhai, G.; Guo, Z.; Bai, X.; Chen, P.; Dong, H.; Tian, S.; Ai, D.; Zhang, Y.; Zhang, K. Combinatorial Peptide Ligand Library-Based Photoaffinity Probe for the Identification of Phosphotyrosine-Binding Domain Proteins. *Anal. Chem.* **2019**, *91* (5), 3221-3226.

[82] Hashimoto, M.; Hatanaka, Y. A novel biotinylated diazirinyl ceramide analogue for photoaffinity labeling. *Bioorg. Med. Chem. Lett.* **2008**, *18* (2), 650-652.

[83] Wu, B.; Jayakar, S. S.; Zhou, X.; Titterton, K.; Chiara, D. C.; Szabo, A. L.; Savechenkov, P. Y.; Kent, D. E.; Cohen, J. B.; Forman, S. A.; Miller, K. W.; Bruzik, K. S. Inhibitable Photolabeling by Neurosteroid Diazirine Analog in the β3-Subunit of Human Heteropentameric Type A GABA Receptors. *Eur. J. Med. Chem.* **2019**, *162*, 810-824.

[84] Park, J.-j.; Sadakane, Y.; Masuda, K.; Tomohiro, T.; Nakano, T.; Hatanaka, Y. Synthesis of Diazirinyl Photoprobe Carrying a Novel Cleavable Biotin. *ChemBioChem* **2005**, *6* (5), 814-818.

[85] Feltes, M.; Moores, S.; Gale, S. E.; Krishnan, K.; Mydock-McGrane, L.; Covey, D. E.; Ory, D. S.; Schaffer, J. E. Synthesis and characterization of diazirine alkyne probes for the study of intracellular cholesterol trafficking. *J. Lipid Res.* **2019**, *60* (3), 707-716.

[86] Yan, C.; Kepa, J. K.; Siegel, D.; Stratford, I. J.; Ross, D. Dissecting the role of multiple reductases in bioactivation and cytotoxicity of the antitumor agent RH1. *Mol. Pharmacol.* **2008**, *74* (6), 1657-1665.

[87] Halloran, M. W.; Lumb, J.-P. Recent Applications of Diazirines in Chemical Proteomics. *Chem. Eur. J.* **2019**, *25* (19), 4885-4898.

[88] Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **2012**, *20* (2), 554-570.

[89] Iacobucci, C.; Götze, M.; Piotrowski, C.; Arlt, C.; Rehkamp, A.; Ihling, C.; Hage, C.; Sinz, A. Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents. *Anal. Chem.* **2018**, *90* (4), 2805-2809.

[90] Das, J. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chem. Rev.* **2011**, *111* (8), 4405-4417.

[91] Glachet, T.; Marzag, H.; Rosa, N. S.; Colell, J. F. P.; Zhang, G.; Warren, W. S.; Franck, X.; Theis, T.; Reboul, V. Iodonitrene in Action: Direct Transformation of Amino Acids into Terminal Diazirines and

$^{15}N_2$-Diazirines and Their Application as Hyperpolarized Markers. *J. Am. Chem. Soc.* **2019**, *141* (34), 13689-13696.

[92] Hill, J. R.; Robertson, A. A. B. Fishing for Drug Targets: A Focus on Diazirine Photoaffinity Probe Synthesis. *J. Med. Chem.* **2018**, *61* (16), 6945-6963.

[93] Wang, L.; Tachrim, Z. P.; Kurokawa, N.; Ohashi, F.; Sakihama, Y.; Hashidoko, Y.; Hashimoto, M. Base-Mediated One-Pot Synthesis of Aliphatic Diazirines for Photoaffinity Labeling. *Molecules* **2017**, *22* (8), 1389.

[94] Ichiishi, N.; Moore, K. P.; Wassermann, A. M.; Wolkenberg, S. E.; Krska, S. W. Reducing Limitation in Probe Design: The Development of a Diazirine-Compatible Suzuki-Miyaura Cross Coupling Reaction. *ACS Med. Chem. Lett.* **2018**, *10* (1), 56-61.

[95] Ahad, A. M.; Jensen, S. M.; Jewett, J. C. A traceless Staudinger reagent to deliver diazirines. *Org. Lett.* **2013**, *15* (19), 5060-5063.

[96] Hutchings, M. I.; Truman, A. W.; Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **2019**, *51*, 72-80.

[97] Gould, K. Antibiotics: from prehistory to the present day. *J. Antimicrob. Chemother.* **2016**, *71* (3), 572-575.

[98] Khardori, N.; Stevaux, C.; Ripley, K. Antibiotics: From the Beginning to the Future: Part 1. *Indian J. Pediatr.* **2020**, *87* (1), 39-42.

[99] Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic discovery: history, methods and perspectives. *Int. J. Antimicrob. Agents* **2019**, *53* (4), 371-382.

[100] da Cunha, B. R.; Fonseca, L. P.; Calado, C. R. C. Antibiotic Discovery: Where Have We Come from, Where Do We Go? *Antibiotics* **2019**, *8* (2), 45.

[101] Badshah, S. L.; Ullah, A. New developments in non-quinolone-based antibiotics for the inhibition of bacterial gyrase and topoisomerase IV. *Eur. J. Med. Chem.* **2018**, *152*, 393-400.

[102] Heide, L. New aminocoumarin antibiotics as gyrase inhibitors. *Int. J. Med. Microbiol.* **2014**, *304* (1), 31-36.

[103] Buttner, M. J.; Schäfer, M.; Lawson, D. M.; Maxwell, A. Structural insights into simocyclinone as an antibiotic, effector ligand and substrate. *FEMS Microbiol. Rev.* **2018**, *42* (1), fux055.

[104] Tse-Dinh, Y.-C. Targeting bacterial topoisomerase I to meet the challenge of finding new antibiotics. *Future Med. Chem.* **2015**, *7* (4), 459-471.

[105] Mosaei, H.; Harbottle, J. Mechanisms of antibiotics inhibiting bacterial RNA polymerase. *Biochem. Soc. Trans.* **2019**, *47* (1), 339-350.

[106] Hong, W.; Zeng, J.; Xie, J. Antibiotic drugs targeting bacterial RNAs. *Acta Pharm. Sin. B* **2014**, *4* (4), 258-265.

[107] Colameco, S.; Elliot, M. A. Non-coding RNAs as antibiotic targets. *Biochem. Pharmacol.* **2017**, *133*, 29-42.

[108] Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **2013**, *12* (5), 371-387.

[109] Chopra, S.; Reader, J. tRNAs as Antibiotic Targets. *Int. J. Mol. Sci.* **2014**, *16* (1), 321-349.

[110] Vondenhoff, G. H. M.; Van Aerschot, A. Aminoacyl-tRNA synthetase inhibitors as potential antibiotics. *Eur. J. Med. Chem.* **2011**, *46* (11), 5227-5236.

[111] Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **2014**, *12* (1), 35-48.

[112] Fernández-Villa, D.; Aguilar, M. R.; Rojo, L. Folic Acid Antagonists: Antimicrobial and Immunomodulating Mechanisms and Applications. *Int. J. Mol. Sci.* **2019**, *20* (20), 4996.

[113] Yao, J.; Rock, C. O. Bacterial Fatty Acid Metabolism in Modern Antibiotic Discovery. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2017**, *1862* (11), 1300-1309.

[114] Choi, M.; Karunaratne, K.; Kohen, A. Flavin-Dependent Thymidylate Synthase as a New Antibiotic Target. *Molecules* **2016**, *21* (5), 654.

[115] Wencewicz, T. A. New antibiotics from Nature's chemical inventory. *Bioorg. Med. Chem.* **2016**, *24* (24), 6227-6252.

[116] Nasiri, M. J.; Haeili, M.; Ghazi, M.; Goudarzi, H.; Pormohammad, A.; Fooladi, A. A. I.; Feizabadi, M. M. New Insights in to the Intrinsic and Acquired Drug Resistance Mechanisms in Mycobacteria. *Front. Microbiol.* **2017**, *8*, 681.

[117] Hedstrom, L.; Liechti, G.; Goldberg, J. B.; Gollapalli, D. R. The antibiotic potential of prokaryotic IMP dehydrogenase inhibitors. *Curr. Med. Chem.* **2011**, *18* (13), 1909-1918.

[118] Awuni, E. Status of Targeting MreB for the Development of Antibiotics. *Front. Chem.* **2020**, *7*, 884.

[119] Sass, P.; Brötz-Oesterhelt, H. Bacterial cell division as a target for new antibiotics. *Curr. Opin. Microbiol.* **2013**, *16* (5), 522-530.

[120] Carro, L. Protein-protein interactions in bacteria: a promising and challenging avenue towards the discovery of new antibiotics. *Beilstein J. Org. Chem.* **2018**, *14*, 2881-2896.

[121] Kohanski, M. A.; Dwyer, D. J.; Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **2010**, *8* (6), 423-435.

[122] Clatworthy, A. E.; Pierson, E.; Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* **2007**, *3* (9), 541-548.

[123] Ghai, I.; Ghai, S. Understanding antibiotic resistance via outer membrane permeability. *Infect. Drug Resist.* **2018**, *11*, 523-530.

[124] Yilmaz, C.; Özcengiz, G. Antibiotics: Pharmacokinetics, toxicity, resistance and multidrug efflux pumps. *Biochem. Pharmacol.* **2017**, *133*, 43-62.

[125] Reeve, S. M.; Lombardo, M. N.; Anderson, A. C. Understanding the structural mechanisms of antibiotic resistance sets the platform for new discovery. *Future Microbiol.* **2015**, *10* (11), 1727-1733.

[126] Blair, J. M. A.; Webber, M. A.; Baylay, A. J.; Ogbolu, D. O.; Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **2015**, *13* (1), 42-51.

[127] Lobanovska, M.; Pilla, G. Penicillin's Discovery and Antibiotic Resistance: lessons for the Future? *Yale J. Biol. Med.* **2017**, *90* (1), 135-145.

[128] Hall, C. W.; Mah, T.-F. Molecular mechanisms of biofilm-based antibiotic resistance and tolerance in pathogenic bacteria. *FEMS Microbiol. Rev.* **2017**, *41* (3), 276-301.

[129] Beggs, G. A.; Brennan, R. G.; Arshad, M. MarR family proteins are important regulators of clinically relevant antibiotic resistance. *Protein Sci.* **2020**, *29* (3), 647-653.

[130] Felden, B.; Cattoir, V. Bacterial Adaptation to Antibiotics through Regulatory RNAs. *Antimicrob. Agents Chemother.* **2018**, *62* (5), e02503-17.

[131] Hu, Y.; Gao, G. F.; Zhu, B. The antibiotic resistome: gene flow in environments, animals and human beings. *Front. Med.* **2017**, *11* (2), 161-168.

[132] Perry, J.; Waglechner, N.; Wright, G. The Prehistory of Antibiotic Resistance. *Cold Spring Harb. Perspect. Med.* **2016**, *6* (6), a025197.

[133] Sultan, I.; Rahman, S.; Jan, A. T.; Siddiqui, M. T.; Mondal, A. H.; Haq, Q. M. R. Antibiotics, Resistome and Resistance Mechanisms: A Bacterial Perspective. *Front. Microbiol.* **2018**, *9*, 2066.

[134] Gama, J. A.; Zilhão, R.; Dionisio, F. Impact of plasmid interactions with the chromosome and other plasmids on the spread of antibiotic resistance. *Plasmid* **2018**, *99*, 82-88.

[135] Babakhani, S.; Oloomi, M. Transposons: the agents of antibiotic resistance in bacteria. *J. Basic Microbiol.* **2018**, *58* (11), 905-917.

[136] Hershberg, R. Antibiotic-Independent Adaptive Effects of Antibiotic Resistance Mutations. *Trends Genet.* **2017**, *33* (8), 521-528.

[137] Schroeder, M.; Brooks, B. D.; Brooks, A. E. The Complex Relationship between Virulence and Antibiotic Resistance. *Genes* **2017**, *8* (1), 39.

[138] Ghosh, D.; Veeraraghavan, B.; Elangovan, R.; Vivekanandan, P. Antibiotic Resistance and Epigenetics: More to It than Meets the Eye. *Antimicrob. Agents Chemother.* **2020**, *64* (2), e02225-19.

[139] Brauner, A.; Fridman, O.; Gefen, O.; Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* **2016**, *14* (5), 320-330.

[140] Chellat, M. F.; Raguž, L.; Riedl, R. Targeting Antibiotic Resistance. *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (23), 6600-6626.

[141] González-Bello, C. Antibiotic adjuvants – A strategy to unlock bacterial resistance to antibiotics. *Bioorg. Med. Chem. Lett.* **2017**, *27* (18), 4221-4228.

[142] Eiamphungporn, W.; Schaduangrat, N.; Malik, A. A.; Nantasenamat, C. Tackling the Antibiotic Resistance Caused by Class A β-Lactamases through the Use of β-Lactamase Inhibitory Protein. *Int. J. Mol. Sci.* **2018**, *19* (8), 2222.

[143] Graf, F. E.; Palm, M.; Warringer, J.; Farewell, A. Inhibiting conjugation as a tool in the fight against antibiotic resistance. *Drug Dev. Res.* **2019**, *80* (1), 19-23.

[144] Wang, Y.; Yang, Y.; Shi, Y.; Song, H.; Yu, C. Antibiotic-Free Antibacterial Strategies Enabled by Nanomaterials: Progress and Perspectives. *Adv. Mater* **2020**, *32* (18), e1904106.

[145] Larsen, E. M.; Johnson, R. J. Microbial esterases and ester prodrugs: An unlikely marriage for combatting antibiotic resistance. *Drug Dev. Res.* **2019**, *80* (1), 33-47.

[146] Rao, S. C. V.; De Waelheyns, E.; Economou, A.; Anné, J. Antibiotic targeting of the bacterial secretory pathway. *Biochim. Biophys. Acta* **2014**, *1843* (8), 1762-1783.

[147] Munguia, J.; Nizet, V. Pharmacological Targeting of the Host-Pathogen Interaction: Alternatives to Classical Antibiotics to Combat Drug-Resistant Superbugs. *Trends Pharmacol. Sci.* **2017**, *38* (5), 473-488.

[148] Monserrat-Martinez, A.; Gambin, Y.; Sierecki, E. Thinking Outside the Bug: Molecular Targets and Strategies to Overcome Antibiotic Resistance. *Int. J. Mol. Sci.* **2019**, *20* (6), 1255.

[149] Porfírio, S.; Carlson, R. W.; Azadi, P. Elucidating Peptidoglycan Structure: An Analytical Toolset. *Trends Microbiol.* **2019**, *27* (7), 607-622.

[150] Vollmer, W.; Blanot, D.; de Pedro, M. A. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **2008**, *32* (2), 149-167.

[151] Desmarais, S. M.; De Pedro, M. A.; Cava, F.; Huang, K. C. Peptidoglycan at its peaks: how chromatographic analyses can reveal bacterial cell-wall structure and assembly. *Mol. Microbiol.* **2013**, *89* (1), 1-13.

[152] Boudreau, M. A.; Fisher, J. F.; Mobashery, S. Messenger Functions of the Bacterial Cell Wall-derived Muropeptides. *Biochemistry* **2012**, *51* (14), 2974-2990.

[153] Smith, C. A. Structure, Function and Dynamics in the *mur* Family of Bacterial Cell Wall Ligases. *J. Mol. Biol.* **2006**, *362* (4), 640-655.

[154] Park, J. T. Uridine-5'-Pyrophosphate Derivatives. I. Isolation from Staphylococcus aureus. *J. Biol. Chem.* **1952**, *194* (2), 877-884.

[155] Manat, G.; Roure, S.; Auger, R.; Bouhss, A.; Barreteau, H.; Mengin-Lecreulx, D.; Touzé, T. Deciphering the Metabolism of Undecaprenyl-Phosphate: The Bacterial Cell-Wall Unit Carrier at the Membrane Frontier. *Microb. Drug Resist.* **2014**, *20* (3), 199-214.

[156] Teng, K.-H.; Liang, P.-H. Undecaprenyl diphosphate synthase, a *cis*-prenyltransferase synthesizing lipid carrier for bacterial cell wall biosynthesis. *Mol. Membr. Biol.* **2012**, *29* (7), 267-273.

[157] Laddomada, F.; Miyachiro, M. M.; Dessen, A. Structural Insights into Protein-Protein Interactions Involved in Bacterial Cell Wall Biogenesis. *Antibiotics* **2016**, *5* (2), 14.

[158] Meeske, A. J.; Sham, L.-T.; Kimsey, H.; Koo, B.-M.; Gross, C. A.; Bernhardt, T. G.; Rudner, D. Z. MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis. Proc. Natl. Acad. Sci. USA.* **2015**, *112* (20), 6437-6442.

[159] Mohammadi. T.; van Dam, V.; Sijbrandi, R.; Vernet, T.; Zapun, A.; Bouhss, A.; Diepeveen de Bruin, M.; Nguzen-Distèche, M.; de Kruijff, B.; Breukink, E. Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J.* **2011**, *30* (8), 1425-1432.

[160] Cho, H.; Wivagg, C. N.; Kapoor, M.; Barry, Z.; Rohs, P. D. A.; Suh, H.; Marto, J. A.; Garner, E. C.; Bernhardt, T. G. Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. *Nat. Microbiol.* **2016**, *1*, 16172.

[161] Do, T.; Page, J. E.; Walker, S. Uncovering the activities, biological roles, and regulation of bacterial cell wall hydrolases and tailoring enzymes. *J. Biol. Chem.* **2020**, *295* (10), 3347-3361.

[162] Sanders, A. N.; Pavelka, M. S. Phenotypic analysis of *Escherichia coli* mutants lacking L,D-transpeptidases. *Microbiology* **2013**, *159* (Pt 9), 1842-1852.

[163] Workman, S. D.; Worrall, L. J.; Strynadka, N. C. J. Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. *Nat. Commun.* **2018**, *9* (1), 1159.

[164] Bugg, T. D. H.; Braddick, D.; Dowson, C. G.; Roper, D. I. Bacterial cell wall assembly: still an attractive antibacterial target. *Trends Biotechnol.* **2011**, *29* (4), 167-173.

[165] Han, H.; Yang, Y.; Olesen, S. H.; Becker, A.; Betzi, S.; Schönbrunn, E. The Fungal Product Terreic Acid is a Covalent Inhibitor of the Bacterial Cell Wall Biosynthetic Enzyme UDP-*N*-acteylglucosamine 1-carboxyvinyltransferase (MurA). *Biochemistry* **2010**, *49* (19), 4276-4282.

[166] Sova, M.; Kovac, A.; Turk, S.; Hrast, M.; Blanot, D.; Gobec, S. Phosphorylated hydroxyethylamines as novel inhibitors of the bacterial cell wall biosynthesis enzymes MurC to MurF. *Bioorg. Chem.* **2009**, *37* (6), 217-222.

[167] Francisco, G. D.; Li, Z.; Albright, J. D.; Eudy, N. H.; Katz, A. H.; Petersen, P. J.; Labthavikul, P.; Singh, G.; Yang, Y.; Rasmussen, B. A.; Lin, Y.-I.; Mansour, T. S. Phenyl thiazolyl urea and carbamate derivatives as new inhibitors of bacterial cell-wall biosynthesis. *Bioorg. Med. Chem. Lett.* **2004**, *14* (1), 235-238.

[168] Doig, P.; Boriack-Sjodin, P. A.; Dumas, J.; Hu, J.; Itoh, K.; Johnson, K.; Kazmirski, S.; Kinoshita, T.; Kuroda, S.; Sato, T.-o.; Sugimoto, K.; Tohyama, K.; Aoi, H.; Wakamatsu, K.; Wang, H. Rational design of inhibitors of the bacterial cell wall synthetic enzyme GlmU using virtual screening and lead-hopping. *Bioorg. Med. Chem.* **2014**, *22*, (21), 6256-6269.

[169] Mesleh, M. F.; Rajaratnam, P.; Conrad, M.; Chandrasekaran, V.; Liu, C. M.; Pandya, B. A.; Hwang, Y. S.; Rye, P. T.; Muldoon, C.; Becker, B.; Zuegg, J.; Meutermans, W.; Moy, T. I. Targeting Bacterial Cell

Wall Peptidoglycan Synthesis by Inhibition of Glycosyltransferase Activity. *Chem. Biol. Drug Des.* **2016**, *87* (2), 190-199.

[170] Yamaguchi, T.; Blázquez, B.; Hesek, D.; Lee, M.; Llarrull, L. I.; Boggess, B.; Oliver, A. G.; Fisher, J. F.; Mobashery, S. Inhibitors for Bacterial Cell-Wall Recycling. *ACS Med. Chem. Lett.* **2012**, *3* (3), 238-242.

[171] Smithen, D. A.; Forget, S. M.; McCormick, Syvitski, R. T.; Jakeman, D. L. Polyphosphate-containing bisubstrate analogues as inhibitors of a bacterial cell wall thymidylyltransferase. *Org. Biomol. Chem.* **2015**, *13* (11), 3347-3350.

[172] Bucher, T.; Oppenheimer-Shaanan, Y.; Savidor, A.; Bloom-Ackermann, Z.; Kolodkin-Gal, I. Disturbance of the bacterial cell wall specifically interferes with biofilm formation. *Environ. Microbiol. Rep.* **2015**, *7* (6), 990-1004.

[173] Ikeda, M.; Wachi, M.; Jung, H. K.; Ishino, F.; Matsuhashi, M. The *Escherichia coli mraY* Gene Encoding UDP-*N*-Acetylmuramoyl-Pentapeptide: Undecaprenyl-Phosphate Phospho-*N*-Acetylmuramoyl-Pentapeptide Transferase. *J. Bacteriol.* **1991**, *173* (3), 1021-1026.

[174] Boyle, D. S.; Donachie, W. D. *mraY* Is an Essential Gene for Cell Growth in *Escherichia coli*. *J. Bacteriol.* **1998**, *180* (23), 6429-6432.

[175] Bouhss, A.; Menging-Lecreulx, D.; Le Beller, D.; van Heijenoort, J. Topological analysis of the MraY protein catalyzing the first membrane step of peptidoglycan synthesis. *Mol. Microbiol.* **1999**, *34* (3), 576-585.

[176] Al-Dabbagh, B.; Henry, X.; El Ghachi, M.; Auger, G.; Blanot, D.; Parquet, C.; Mengin-Lecreulx, D.; Bouhss, A. Active Site Mapping of MraY, a Member of the Polyprenyl-phosphate *N*-Acetylhexosamine 1-Phosphate Transferase Superfamily, Catalyzing the First Membrane Step of Peptidoglycan Biosynthesis. *Biochemistry* **2008**, *47* (34), 8919-8928.

[177] Chung, B. C.; Zhao, J.; Gillespie, R. A.; Kwon, D.-Y.; Guan, Z.; Hong, J.; Zhou, P.; Lee, S.-Y. Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. *Science* **2013**, *341* (6149), 1012-1016.

[178] Chung, B. C.; Mashalidids, E. H.; Tanino, T.; Kim, M.; Matsuda, A.; Hong, J.; Ichikawa, S.; Lee, S.-Y. Structural insights into inhibition of Lipid I production in bacterial cell wall synthesis. *Nature* **2016**, *533* (7604), 557-560.

[179] Hakulinen, J. K.; Hering, J.; Brändén, G.; Chen, H.; Snijder, A.; Ek, M.; Johansson, P. MraY-antibiotic complex reveals details of tunicamycin mode of action. *Nat. Chem. Biol.* **2017**, *13* (3), 265-267.

[180] Mashalidis, E. H.; Kaeser, B.; Terasawa, Y.; Katsuyama, A.; Kwon, D.-Y.; Lee, K.; Hong, J.; Ichikawa, S.; Lee, S.-Y. Chemical logic of MraY inhibition by antibacterial nucleoside natural products. *Nat. Commun.* **2019**, *10* (1), 2917.

[181] Liu, Y.; Rodrigues, J. P. G. L. M.; Bonvin, A. M. J. J.; Zaal, E. A.; Berkers, C. R.; Heger, M.; Gawarecka, K.; Swlezewska, E.; Breukink, E.; Egmond, M. R. New Insight into the Catalytic Mechanism of Bacterial MraY from Enzyme Kinetics and Docking Studies. *J. Biol. Chem.* **2016**, *291* (29), 15057-15068.

[182] Al-Dabbagh, B.; Olatunji, S.; Crouvoisier, M.; El Ghachi, M.; Blanot, D.; Mengin-Lecreulx, D.; Bouhss, A. Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate *N*-acetylhexosamine 1-phosphate transferase superfamily. *Biochimie* **2016**, *127*, 249-257.

[183] Winn, M.; Goss, R. J. M.; Kimura, K.-i.; Bugg, T. D. H. Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis. *Nat. Prod. Rep.* **2010**, *27* (2), 279-304.

[184] Funabashi, M.; Baba, S.; Takatsu, T.; Kizuka, M.; Ohata, Y. Tanaka, M.; Nonaka, K.; Spork, A. P.; Ducho, C.; Chen, W.-C. L.; Van Lanen, S. G. Structure-Based Gene Targeting Discovery of a Novel Bacterial Translocase I Inhibitor, Sphaerimicin. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (44), 11607-11611.

[185] Rodolis, M. T.; Mihalyi, A.; O'Reilly, A.; Slikas, J.; Roper, D. I.; Hancock, R. E. W.; Bugg, T. D. H. Identification of a Novel Inhibition Site in Translocase MraY Based upon the Site of Interaction with Lysis Protein E from Bacteriophage ϕX174. *ChemBioChem* **2014**, *15* (9), 1300-1308.

[186] Rodolis, M. T.; Mihalyi, A.; Ducho, C.; Eitel, K.; Gust, B.; Goss, R. J. M.; Bugg, T. D. H. Mechanism of action of the uridyl peptide antibiotics: an unexpected link to a protein-protein interaction site in translocase MraY. *Chem. Commun.* **2014**, *50* (86), 13023-13025.

[187] Auberger, N.; Frlan, R.; Al-Dabbagh, B.; Bouhss, A.; Crouvoisier, M.; Gravier-Pelletier, C.; Le Merrer, Y. Synthesis and biological evaluation of potential new inhibitors of the bacterial transferase MraY with a β-ketohosphonate structure. *Org. Biomol. Chem.* **2011**, *9* (24), 9301-8312.

[188] Mihalyi, A.; Jamshidi, S.; Slikas, J.; Bugg, T. D. H. Identification of novel inhibitors of phosphor-MurNAc-pentapeptide translocase MraY from library screening: Isoquinoline alkaloid michellamine B and xanthene dye phloxine B. *Bioorg. Med. Chem.* **2014**, *22* (17), 4566-4571.

[189] Cao, J.; Yi, F.; Tian, Q.; Dang, G.; Si, W.; Liu, S.; Yu, S. Targeting the gram-negative bacteria peptidoglycan synthase MraY as a new approach for monoclonal antibody anti-bacterial activity. *Hum. Vaccin. Immunother.* **2017**, *13* (9), 2086-2091.

[190] Trautman, E. P.; Crawford, J. M. A New Nucleoside Antibiotic Chokes Bacterial RNA Polymerase. *Biochemistry* **2017**, *56* (37), 4923-4924.

[191] Ushimaru, R.; Liu, H.-w. Biosynthetic Origin of the Atypical Stereochemistry in the Thioheptose Core of Albomycin Nucleoside Antibiotics. *J. Am. Chem. Soc.* **2019**, *141* (6), 2211-2214.

[192] Bugg, T. D. H.; Kerr, R. V. Mechanism of action of nucleoside antibacterial natural product antibiotics. *J. Antibiot.* **2019**, *72* (12), 865-876.

[193] McDonald, L. A.; Barbieri, L. R.; Carter, G. T.; Lenoy, E.; Lotvin, J.; Petersen, P. J.; Siegel, M. M.; Singh, G.; Williamson, R. T. Structures of the Muraymycins, Novel Peptidoglycan Biosynthesis Inhibitors. *J. Am. Chem. Soc.* **2002**, *124* (35), 10260-10261.

[194] Spork, A. P.; Ducho, C. Novel 5'-deoxy nucleosyl amino acid scaffolds for the synthesis of muraymycin analogues. *Org. Biomol. Chem.* **2010**, *8* (10), 2323-2326.

[195] Cui, Z.; Wang, X.; Koppermann, S.; Thorson, J. S.; Ducho, C.; Van Lanen, S. G. Antibacterial Muraymycins from Mutant Strains of *Streptomyces* sp. NRLL 30471. *J. Nat. Prod.* **2018**, *81* (4), 942-948.

[196] Cheng, L.; Chen, W.; Zhai, L.; Xu, D.; Huang, T.; Lin, S.; Zhou, X.; Deng, Z. Identification of the gene cluster involved in muraymycin biosynthesis from Streptomyces sp. NRRL 30471. *Mol. Biosyst.* **2011**, *7* (3), 920-927.

[197] Cui, Z.; Liu, X.; Overbay, J.; Cai, W.; Wang, X.; Lemke, A.; Wiegmann, D.; Niro, G.; Thorson, J. S.; Ducho, C.; Van Lanen, S. G. Enzymatic Synthesis of the Ribosylated Glycyl-Uridine Disaccharide Core of Peptidyl Nucleoside Antibiotics. *J. Org. Chem.* **2018**, *83* (13), 7239-7249.

[198] Shiraishi, T.; Kuzuyama, T. Recent advances in the biosynthesis of nucleoside antibiotics. *J. Antibiot.* **2019**, *72* (12), 913-923.

[199] Cui, Z.; Wang, X.-C.; Liu, X.; Lemke, A.; Koppermann, S.; Ducho, C.; Rohr, J.; Thorson, J. S.; Van Lanen, S. G. Self-Resistance during Muraymycin Biosynthesis: a Complementary Nucleotidyltransferase and Phosphotransferase with identical Modification Sites and Distinct Temporal Order. *Antimicrob. Agents Chemother.* **2018**, *62* (7), e00193-18.

[200] Serpi, M.; Ferrari, V.; Pertusati, F. Nucleoside Derived Antibiotics to Fight Microbial Drug Resistance: New Utilities for an Established Class of Drugs? *J. Med. Chem.* **2016**, *59* (23), 10343-10382.

[201] Mitachi, K.; Aleiwi, B. A.; Schneider, C. M.; Siricilla, S.; Kurosu, M. Stereocontrolled Total Synthesis of Muraymycin D1 Having a Dual Mode of Action against *Mycobacterium tuberculosis. J. Am. Chem. Soc.* **2016**, *138* (39), 12975-12980.

[202] Yamashita, A.; Norton, E.; Petersen, P. J.; Rasmussen, B. A.; Singh, G.; Yang, Y.; Mansour, T. S.; Ho, D. M. Muraymycins, Novel Peptidoglycan Biosynthesis Inhibitors: Synthesis and SAR of Their Analogues. *Bioorg Med. Chem. Lett.* **2003**, *13* (19), 3345-3350.

[203] Tanino, T.; Hirano, S.; Ichikawa, S.; Matsuda, A. Synthetic study of muraymycins using Ugi-four component reaction. *Nucleic Acids Symp. Ser.* **2008**, (52), 557-558.

[204] Tanino, T.; Ichikawa, S.; Shiro, M.; Matsuda, A. Total Syntehsis of (-)-Muraymycin D2 and Its Epimer. *J. Org. Chem.* **2010**, *75* (5), 1366-1377.

[205] Hirano, S.; Ichikawa, S.; Matsuda, A. Synthesis of Caprazamycin Analogues and Their Structure-Activity Relationship for Antibacterial Activity. *J. Org. Chem.* **2008**, *73* (2), 569-577.

[206] Spork, A. P.; Büschleb, M.; Ries, O.; Wiegmann, D.; Boettcher, S.; Mihalyi, A.; Bugg, T. D. H.; Ducho, C. Lead Structures for New Antibacterials: Stereocontrolled Synthesis of a Bioactive Muraymycin Analogue. *Chem. Eur. J.* **2014**, *20* (47), 15292-15297.

[207] Spork, A. P.; Wiegmann, D.; Granitzka, M.; Stalke, D.; Ducho, C. Stereoselective Synthesis of Uridine-Derived Nucleosyl Amino Acids. *J. Org. Chem*. **2011**, *76* (24), 10083-10098.

[208] Ries, O.; Büschleb, M.; Granitzka, M.; Stalke, D.; Ducho, C. Amino acid motifs in natural products: synthesis of *O*-acylated derivatives of (2*S*,3*S*)-3-hydroxyleucine. *Beilstein J. Org. Chem.* **2014**, *10*, 1135-1142.

[209] Koppermann, S.; Cui, Z.; Fischer, P. D.; Wang, X.; Ludwig, J.; Thorson, J. S.; Van Lanen, S. G.; Ducho, C. Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics. *ChemMedChem* **2018**, *13* (8), 779-784.

[210] Tanino, T.; Al-Dabbagh, B.; Mengin-Lecreulx, D.; Bouhss, A.; Oyama, H.; Ichikawa, S.; Matsuda, A. Mechanistic Analysis of Muraymycin Analogues: A Guide to the Design of MraY Inhibitors. *J. Med. Chem.* **2011**, *54* (24), 8421-8439.

[211] Spork, A. P. „Synthetische Untersuchungen zur Nucleosid-Einheit von Muraymycin-Antibiotika.“ Ph.D. thesis, Georg-August-Universität Göttingen, **2012**.

[212] Leyerer, K. “Muraymycin Nucleoside Antibiotics: Novel SAR Insights and Synthetic Approaches.” Ph.D. thesis, Saarland University, Saarbrücken, **2018**.

[213] Heib, A.; Niro, G.; Weck, S. C.; Koppermann, S.; Ducho, C. Muraymycin Nucleoside Antibiotics: Structure-Activity Relationship for Variations in the Nucleoside Unit. *Molecules* **2019**, *25* (1), 22.

[214] Wiegmann, D.; Koppermann, S.; Ducho, C. Aminoribosylated Analogues of Muraymycin Nucleoside Antibiotics. *Molecules* **2018**, *23* (12), 3085.

[215] Spork, A. P.; Koppermann, S.; Wohnig, S. S. N.; Linder, R.; Ducho, C. Analogues of Muraymycin Nucleoside Antibiotics with Epimeric Uridine-Derived Core Structures. *Molecules* **2018**, *23* (11), 2868.

[216] Tanino, T.; Ichikawa, S.; Al-Dabbagh, B.; Bouhss, A.; Oyama, H.; Matsuda, A. Synthesis and Biological Evaluation of Muraymycin Analogues Active against Anti-Drug-Resistant Bacteria. *ACS Med. Chem. Lett.* **2010**, *1* (6), 258-262.

[217] Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2005**, *2* (4), 261-267.

[218] Frigerio, M.; Santagostino, M.; Sputore, S. A User-Friendly Entry to 2-Iodoxybenzoid Acid (IBX). *J. Org. Chem.* **1999**, *64* (12), 4537-4538.

[219] More, J. D.; Finney, N. S. A Simple and Advantageous Protocol for the Oxidation of Alcohols with *o*-Iodoxybenzoid Acid (IBX). *Org. Lett.* **2002**, *4* (17), 3001-3003.

[220] Büschleb, M. „Synthese von Capreomycidin- und Epicapreomycidin-haltigen Naturstoffbausteinen.“ Ph.D. thesis, Georg-August-Universität Göttingen, Göttingen, **2012**.

[221] Büschleb, M.; Granitzka, M.; Stalke, D.; Ducho, C. A biomimetic domino reaction for the concise synthesis of capreomycidine and epicapreomycidine. *Amino Acids* **2012**, *43* (6), 2313-2328.

[222] Zhao, M.; Li, J.; Mano, E.; Song, Z.; Tschaen, D. M.; Gravowski, E. J. J.; Reider, P. J. Oxidation of Primary Alcohols to Carboxylic Acids with Sodium Chlorite Catalyzed by TEMPO and Bleach. *J. Org. Chem.* **1999**, *64*, 2564-2566.

[223] Allegretti, P. A.; Ferreira, E. M. Generation of α,β-unsaturated platinum carbenes from homopropargylic alcohols: rearrangements to polysubstituted furans. *Org. Lett.* **2011**, *13* (21), 5924-5927.

[224] Harris, M. R.; Konev, M. O.; Jarvo, E. R. Enantiospecific intramolecular Heck reactions of secondary benzylic ethers. *J. Am. Chem. Soc.* **2014**, *136* (22), 7825-7828.

[225] Wisniewska, H. M.; Swift, E. C.; Jarvo, E. R. Functional-group-tolerant, nickel-catalyzed cross-coupling reaction for enantioselective construction of tertiary methyl-bearing stereocenters. *J. Am. Chem. Soc.* **2013**, *135* (24), 9083-9090.

[226] Schaus, S. E.; Brånalt, J.; Jacobsen, E. N. Total Synthesis of Muconin by Efficient Assembly of Chiral Building Blocks. *J. Org. Chem.* **1998**, *63* (15), 4876-4877.

[227] Jin, T.; Himuro, M.; Yamamoto, Y. Brønsted acid-catalyzed cascade cycloisomerization of enynes via acetylene cations and sp3-hybridized C-H bond activation. *J. Am. Chem. Soc.* **2010**, *132* (16), 5590-5591.

[228] Kadam, A.; Nguyen, M.; Kopach, M.; Richardson, P.; Gallou, F.; Wan, Z.-K.; Zhang, W. Comparative performance evaluation and systematic screening of solvents in a range of Grignard reactions. *Green Chem.* **2013**, *15*, 1880-1888.

[229] Ruminiski, P.; Penning, T. D.; Jiang, L.; Devadas, B.; Rogers, T.; VanCamp, J.; Yuan, C.; Pharmacia Corporation. *Lactone Integrin Antagonists*. Patent No. US 2002/0045645 A1, Apr. 18, **2002**.

[230] Peng, T.; Hang, H. C. Bifunctional Fatty Acid Chemical Reporter for Analyzing S-palmitoylated Membrane Protein-Protein Interactions in Mammalian Cells. *J. Am. Chem. Soc.* **2015**, *137* (2), 556-559.

[231] Legros, J.; Figadère, B. Iron-promoted C-C bond formation in the total synthesis of natural products and drugs. *Nat. Prod. Rep.* **2015**, *32* (11), 1541-1555.

[232] Cheng, X.; Li, L.; Uttamchandani, M.; Yao, S. Q. In Situ Proteome Profiling of C75, a Covalent Bioactive Compound with Potential Anticancer Activities. *Org. Lett.* **2014**, *16* (5), 1414-1417.

[233] Lemke, A. „Catalytic and Biocatalytic Methods for the Efficient Synthesis of Biologically Relevant Non-Proteinogenic Amino Acids." Ph.D. thesis, Georg-August-Universität Göttingen, **2013**.

[234] Kambe, T.; Correia, B. E.; Niphakis, M. J.; Cravatt, B. F. Maping the protein interaction landscape for fully functionalized small-molecule probes in human cells. *J. Am. Chem. Soc.* **2014**, *136* (30), 10777-10782.

[235] MacKinnon, A. L.; Garrison, J. L.; Hegde, R. S.; Taunton, J. Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J. Am. Chem. Soc.* **2007**, *129* (47), 14560-14561.

[236] Vervacke, J. S.; Funk, A. L.; Wang, Y.-C.; Strom, M.; Hrycyna, C. A.; Distefano, M. D. Diazirine-Containing Photoactivatable Isoprenoid: Synthesis and Application in Studies with Isoprenylcysteine Carboxyl Methyltransferase. *J. Org. Chem.* **2014**, *79* (5), 1971-1978.

[237] Schütz, C. „Inhibitoren der bakteriellen Translocase and Collagenase als potentielle Antibiotika." Master thesis, Universität des Saarlandes, **2016**.

[238] Al-Warhi, T. I.; Al-Hazimi, H. M.A.; El-Faham, A. Recent development in peptide coupling reagents. *J. Saudi Chem. Soc.* **2012**, *16*, 97-116.

[239] Behrendt, R.; White, P.; Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **2016**, *22* (1), 4-27.

[240] Carpino, L. A.; Williams, J. R.; Łopusiński, A. Polymeric De-blocking Agents for the Fluoren-9-ylmethoxycarbonyl (FMOC) Amino-protecting Group. *J. Chem. Soc. Chem. Commun.* **1978**, *11*, 450-451.

[241] Wohnig, S.; Spork, A. P.; Koppermann, S.; Mieskes, G.; Gisch, N.; Jahn, R.; Ducho, C. Total Synthesis of Dansylated Park's Nucleotide for High-Throughput MraY Assays. *Chem. Eur. J.* **2016**, *22* (49), 17813-17819.

[242] Leyerer, K. „Arbeiten zur Synthese neuer Analoga der Muraymycin-Antibiotika." Master thesis, Universität Paderborn, **2013**.

11 Development of Protein Degraders from Small Molecule Protein Inhibitors

11.1 Introduction

11.1.1 Targeted Protein Degradation – A Promising Tool of Chemical Biology

Protein degradation is an important regulatory mechanism of protein function in human cells, disposing of unnecessary or damaged proteins as part of constant turn-over cycles[1,2]. After the discovery of the lysosome-mediated protein degradation mechanism in the middle of the last century, the ubiquitin-proteasome system was described in the late 1970s and early 1980s[1], and accounts for the largest cellular protein degradation machinery[2].

Artificial Induction of Proteasomal Degradation of a Protein of Interest
Within the cellular system, proteins are commonly tagged with ubiquitin to initiate their proteasomal degradation process[3,4] (Figure 11.1.1). Briefly, ubiquitin is bound by E1 enzymes, which transfer the ubiquitin to the active site of E2 enzymes[3]. The ubiquitin-carrying E2 enzymes are then able to associate with E3 ligases (binding specific proteins), which brings the E2 enzyme and a target protein into close proximity facilitating ubiquitination of the target, where ubiquitin is attached to lysin motifs on the target protein's surface[3]. Since ubiquitin (a small protein unit) itself carries several lysins that can be ubiquitinated, monoubiquitination can lead to the formation of polyubiquitin chains[3].

Over 600 different E3 ligases are encoded in the human genome, each with their specific targets and functions[3]. Generally, E3 ligases are divided into three main families: the RING (really interesting new gene), the HECT (homologous to the E6AP carboxyl terminus), and the RBR (RING-between-RING) E3 ligases[3]. Notably, in contrast to the main catalytic mechanism of the RING (including the CRLs; cullin RING ligases) family (which is described above), the HECT and RBR E3 ligases transfer the ubiquitin units *via* an additional intermediate step[3]. After E3 ligase-mediated ubiquitination of the target protein, the proteasome – a protease – degrades the tagged protein[5]. By forming a hollow cylinder, encasing the active site, the proteasome only disassembles polypeptide chains which are guided through its core, forming a highly specific and regulated protein degrading machinery[5].

In chemical biology, the ability to control protein expression or stability within an intact cellular system poses a valuable tool and can provide essential information in the study of disease biology as well as target validation studies and drug development processes. Hence, a variety of genetic techniques, such as the CRISPR/Cas9 gene editing approach, and translational regulators, including RNAi, are often applied, which inhibit the expression of a protein of interest[6]. While these are powerful tools, they usually require prolonged periods until their effects are noticeable, which can create the opportunity for the studied cellular system to develop compensation mechanisms[6]. Furthermore, these techniques have also been reported to potentially have off-target effects[6]. Considering these limitations, non-genetic, small molecule-based strategies could offer an interesting alternative. For example, small molecule-mediated targeted protein degradation could

enable directed and fast removal of a protein of interest. In addition, since small molecule agents would only have to bind, and not necessarily inhibit, a target protein, it was further envisioned that this approach could be used to degrade proteins that are historically challenging to inhibit with common covalent or reversible inhibitor strategies[3].

In the early 1990s, selective estrogen receptor α (ERα) binders were identified to cause degradation of their target, thus being one of the first reported small molecules with targeted protein degradation effects[7]. Later, a strategy aiming at the introduction of hydrophobic patches on a protein's surface to induce its degradation due to the resemblance to a partially unfolded protein was developed[3]. This so called "hydrophobic tagging" strategy uses protein of interest-HaloTag fusion proteins, which are bound by hydrophobic probes, leading to the selective degradation of the target-fusion protein[3]. The same concept was also applied to adamantane-linked known ligands of a protein, which introduce the lipophilic adamantane moiety *via* small molecule-directed binding of a target[3]. However, with the identification of E3 ligase ligands a novel approach was investigated, in which an E3 ligase binder and a binder of a protein of interest would be linked to initiate the proteasomal degradation machinery to degrade that target protein[8].

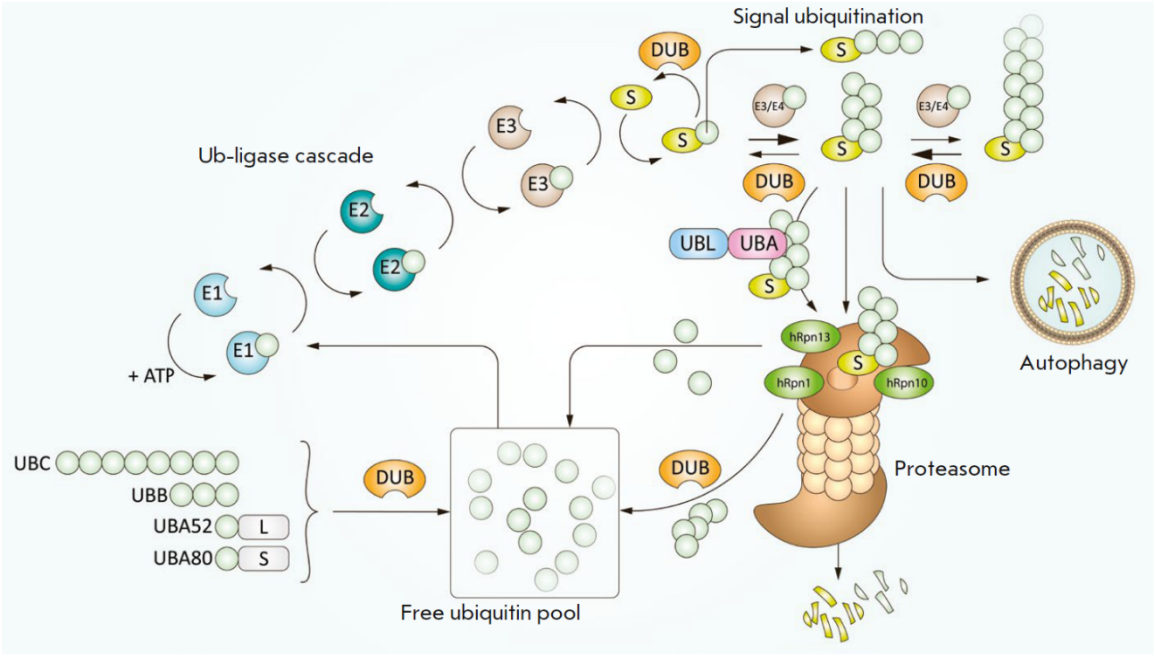


Figure 11.1.1 | The Proteasomal Degradation Machinery. Schematic overview of the proteasomal degradation of proteins. Free ubiquitin is bound by E1 enzymes in an ATP-dependent manner. From E1 enzymes, ubiquitin is transferred to the active site of E2 enzymes. E3 ligases recruit substrate proteins to ubiquitin-carrying E2 enzymes, which either ubiquitinate the E3 ligase substrate directly or by transferring the ubiquitin unit onto the E3 ligase in an intermediate step. Monoubiquitination can lead to polyubiquitination, which can either influence the substrate protein's function or lead to its degradation. While the ubiquitin-induced protein degradation is mostly carried out by the proteasome, in some cases the tagged proteins are degraded *via* autophagy mechanisms. Taken from Kudriaeva *et al*., *Acta Naturae* **2020**, *12* (1), 18-32[4].

Bivalent Small Molecule Protein Degraders

One of the most successfully applied E3 ligase ligands for targeted protein degradation are thalidomide and its analogs[8]. In the late 1950s, thalidomide was a novel agent to treat morning sickness in pregnant women[9]. While the compound appeared to be safe in mouse models – although its mode of action was not known –, serious side effects were seen in form of severe birth defects in children whose mothers were taking thalidomide during their pregnancies[9]. Consequently, thalidomide was removed from the market and further research was performed to investigate the question of thalidomide's mechanism[8]. Interestingly, in the meantime, thalidomide was proven to be an effective treatment for leprosy and multiple myeloma[8]. However, it was not until Ito *et al.* reported that thalidomide directly binds to cereblon (CRBN), an E3 ligase, in 2010 that thalidomide's effects were slowly uncovered[10].

Structural studies further elucidated how thalidomide and its analogs, including the less toxic lenalidomide and pomalidomide (also commonly referred to as "IMiDs" because of their immunomodulatory effects)[8], bind CRBN *via* their glutarimide rings, while the phthalimide moieties are solvent exposed[11] (Figure 11.1.2). Subsequent studies further demonstrated IMiD-mediated changes in CRBN's substrate specificity, now including transcription factors such as Ikaros (IKZF1) and Aiolos (IKZF3), explaining the immunomodulatory phenotype[8].

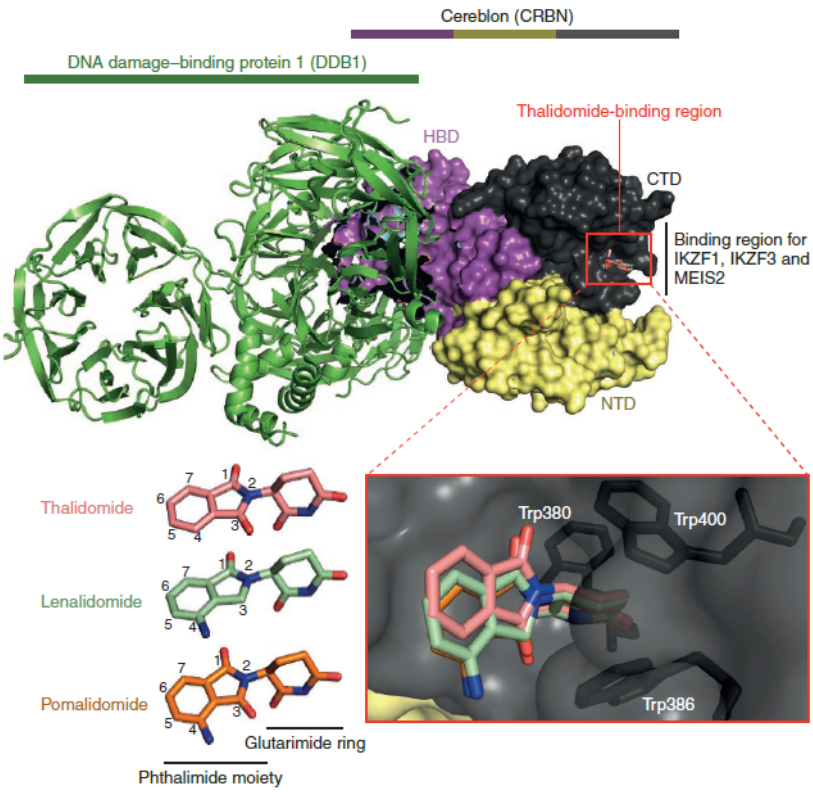


Figure 11.1.2 | Co-crystal Structure of CRBN in Complex with Thalidomide and Analogs. The co-crystal structure of CRBN CTD is shown in black, helical-bundle domain (HBD) in purple, N-terminal domain (NTD) in yellow, and the DNA damage-binding protein 1 (DDB1) in green, which is part of the ubiquitin ligase complex. The E3 ligase CRBN is further shown with its bound ligands thalidomide, lenalidomide, and pomalidomide. Taken from Bhogaraju *et al.*, *Struct. Mol. Biol.* **2014**, *21* (9), 739-740[11].

While IMiDs directly induce targeted protein degradation effects, approaches of linking these E3 ligase ligands with other protein binders to specifically facilitate the ubiquitination and subsequent degradation of a target of interest were thus investigated[12]. These bivalent degrader molecules are also referred to as PROTACs (proteolysis-targeting chimera)[12]. A first study providing *proof of concept* of this method was published in 2001, when a first PROTAC recruiting Skp1-cullin-F box (SCF) containing Hrt1 E3 ligase was shown to induce degradation of methionine aminopeptidase 2 (MetAp-2)[12]. Further studies demonstrated the degradation of various targets, mediated through several different E3 ligases. For example, aryl hydrocarbon receptor nuclear translocator (ARNT)[13] and ERα[14] were degraded *via* the Von Hippel-Lindau (VHL) E3 ligase, an MDM2-recruiting PROTAC induced androgen receptor degradation, and a cellular inhibitor of apoptosis protein 1 (cIAP1)-binding PROTAC mediated the proteasomal degradation of cellular retinol- and retinoic acid-binding proteins (CRABP-I and II)[12]. However, all of these early PROTACs lacked sufficient cellular efficacy, thus requiring high micromolar concentrations to show cellular effects[12].

The first highly potent degrader molecules that were reported are CRBN-recruiting PROTACs[8]. In 2015, Winter, Buckley and coworkers, as well as Lu *et al.*, led by James E. Bradner and Craig M. Crews, respectively, published the development of bromodomain-containing protein 4 (BRD4) degraders based on the BRD4 inhibitor JQ1 and thalidomide analogs as CRBN binders, only differing in the linker moiety[15,16] (Figure 11.1.3). While exhibiting potencies of around 100 nM in cells[15,16], the group of James E. Bradner was further able to show efficacy of their PROTAC in a murine xenograft model[16]. Notably, murine CRBN has to be humanized by the exchange of Ile391 for a valine, which otherwise prevents the degradation of target proteins[17].

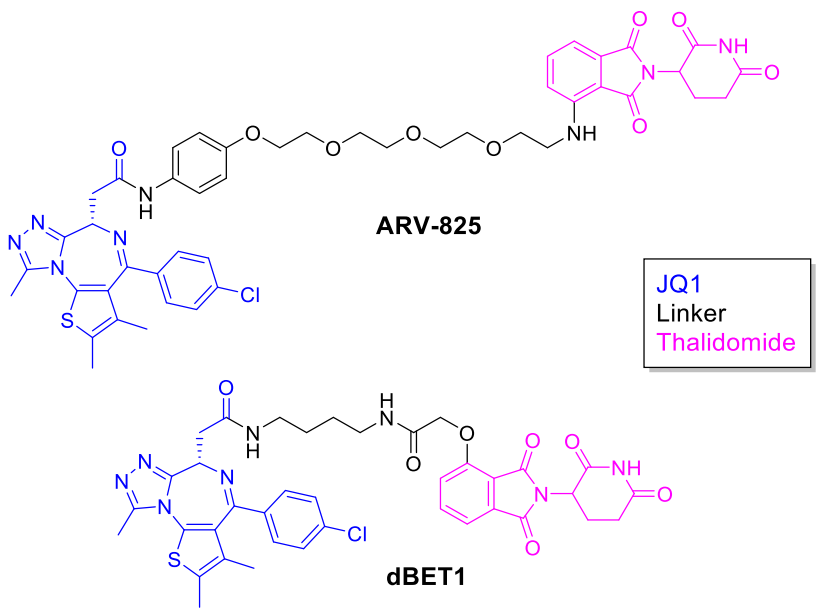


Figure 11.1.3 | Structures of the First CRBN-based BRD4 Degraders. Structures of the two BRD4 degraders reported from the labs of Crews (ARV-825) and Bradner (dBET1). Both compounds are derived from the established BRD4 inhibitor JQ1 (blue) and the CRBN recruiter thalidomide (magenta).

These studies ultimately initiated a whole movement in which medicinal chemists and chemical biologists started to hunt for small molecule-based targeted protein degraders. In the past five years, many targets were shown to be degradable by small molecule degraders, recruiting CRBN, VHL, and various other E3 ligases. Table 11.1.1 gives an overview of some examples of targeted protein degraders reported in the literature, highlighting the variety of targets that were found degradable by one or more E3 ligases using small molecules.

Table 11.1.1 | Examples of Recently Reported Small Molecule-based Protein Degraders[a]

Target Protein	E3 Ligase	Ref.
ALK	VHL	[18]
AMPKγ	CRBN	[19]
APP	CRBN	[20]
AR	VHL	[21]
Bcl-2	CRBN	[22]
BCL-X_L	VHL	[23]
BCR-ABL1	VHL	[24]
BRAF-V600E	CRBN	[25]
BRD4	DCAF16	[26]
BRD7/9	VHL	[27]
BTK	CRBN	[28]
CDK2/5	CRBN	[29]
CDK6	CRBN/VHL/IAP	[30,31]
CDK9	CRBN	[32]
CK2	CRBN	[33]
EGFR	CRBN	[34]
ERα	VHL	[35]
FAK	CRBN/VHL	[36,37]
FKBP12	DCAF16	[38]
HDAC6	CRBN/VHL	[39,40]
IGF-1R/Src	CRBN	[41]
IRAK4	VHL	[42]
JAK	IAP	[43]
Mcl-1	CRBN	[22]
MDM2	CRBN	[44]
MEK	VHL	[45]
PDEδ	CRBN	[46]
RIPK2	VHL/IAP/CRBN	[47]
SHP2	VHL	[48]
Sirt2	CRBN	[49]
SMARCA2/4	VHL	[50]
STAT3	CRBN	[51]
Tau	Keap1	[52]
TBK1	VHL	[53]
TrkC	CRBN	[26]

[a] ALK, anaplastic lymphoma kinase; APP, amyloid precursor protein; AR, androgen receptor; BTK, Bruton's tyrosine kinase; CK2, casein kinase 2; DCAF16, DDB1 and CUL4-associated factor 16; IAP, inhibitor of apoptosis protein; IGF-1R, insulin-like growth factor 1 receptor; IRAK4, interleukin-1 receptor-associated kinase 4; Keap1, Kelch like ECH-associated protein 1; RIPK2, receptor interacting serine/threonine kinase 2; SHP2, Src homology 2 domain-containing phosphatase 2; Sirt2, sirtuin 2; SMARCA2/4, SWI/SNF related, matrix-associated, actin-dependent regulator of chromatin A2; TBK1, TANK-binding kinase 1; TrkC, tropomyosin receptor kinase C.

The Mode of Action and Properties of Small Molecule Protein Degraders

Similar to the endogenous process of proteasomal protein degradation in human cells, small molecule-based degraders facilitate the binding of a protein of interest to an E3 ligase[54] (Figure 11.1.4). In the case that this artificial connection between an E3 ligase and a target protein allows for optimal ternary complex formation, the target protein is ubiquitinated by the E2 enzyme, which is recruited by the E3 ligase[54]. These ubiquitin-tagged proteins are further recognized by the proteasomal degradation machinery and converted into oligopeptide fragments[54].

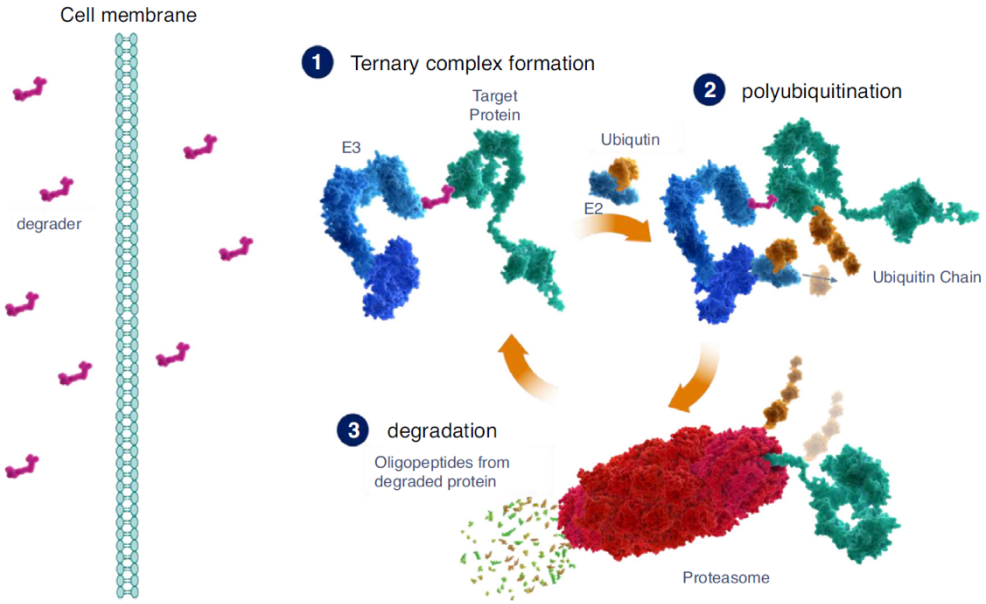


Figure 11.1.4 | The Mode of Action of Small Molecule Protein Degraders. A bivalent small molecule degrader (magenta) binds an E3 ligase complex (blue) on one side and a protein of interest (green) on the other. By facilitating optimal ternary complex formation between the two proteins, the small molecule degrader enables ubiquitination of the target protein by the nearby E2 enzyme. After polyubiquitination, tagged proteins are recognized by the proteasome and subsequently degraded, while the small-molecule degrader is released to initiate another cycle of targeted protein degradation. Adapted from Zhang *et al*., *Drug Discov. Today Technol.* **2019**, *31*, 53-60[54].

The different stages of this mechanism can be validated with competition experiments that suppress the degradation effects of the protein degrader[55]. For example, by adding an excess of ligand for either the E3 ligase or the target protein, the formation of the ternary complex can be prevented[55]. Similarly, adding neddylation inhibitors (an important step in the ubiquitination process) or proteasome inhibitors, such as bortezomib, can prevent ubiquitination or proteasomal degradation of the target, respectively, to further provide evidence of a ubiquitin/proteasome-dependent mode of action of the heterobifunctional protein degrader[55]. Moreover, assays that apply fluorescent visualization techniques to trace ternary complex formation can be used to study the localization[56] and kinetics[57] of this mechanism in a cellular context. Notably, small molecule-based degraders, which naturally facilitate the same mechanism without carrying a distinct E3 ligase- or target protein-binding moiety, are referred to as "molecular glues"[58]. As mentioned above, one famous example of such a compound is thalidomide itself, which binds a tiny pocket of the

E3 ligase CRBN and is able to facilitate the PPIs between CRBN and target proteins[58] (Figure 11.1.5).

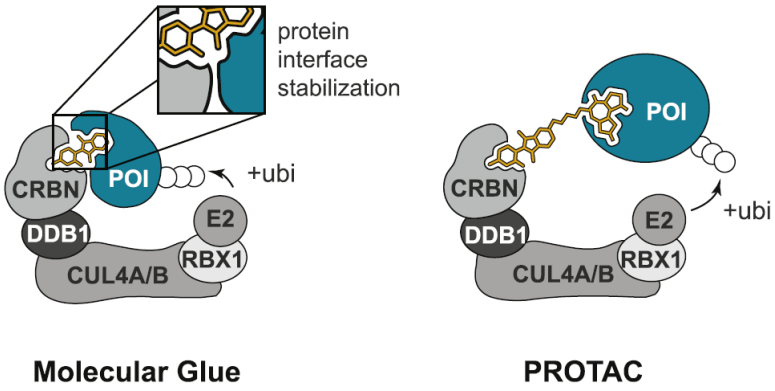


Figure 11.1.5 | Molecular Glues vs. Heterobifunctional Targeted Protein Degraders. Shown is a cartoon of the mechanism by which molecular glue compounds (left) and bivalent targeted protein degraders (PROTACs; right) facilitate the complex formation between an E3 ligase (here: CRBN) and a protein of interest (POI). Adapted from Hanzl *et al.*, *Curr. Opin. Chem. Biol.* **2020**, *56*, 35-41[58].

When regarding targeted protein degraders as potential therapeutics or probes for *in vivo* treatments, it is important to note that, because of their rather large structure, these compounds commonly defy Lipinsky's rule of five[59]. With average molecular weights of about 700-900 Da, heterobifunctional degraders need to be thoroughly optimized with respect to their solubility and other key physicochemical properties to ensure optimal bioavailability of the probe[59,60]. For example, due to the described mode of action, degraders only require a transient interaction with their target to cause protein degradation[60]. Thus, target protein-binding moieties should be compact to keep the overall size of the conjugate as small as possible, even if some binding affinity is sacrificed[60]. Furthermore, switching to other E3 ligase ligands also impacts the physicochemical properties of a targeted protein degrader[60]. Lastly, the linker region poses a highly crucial part of the degrader molecule[60]. While its length determines degradation efficacy by enabling optimal ternary complex formation[61], the linker composition is another critical feature, as it influences pharmacokinetic/pharmacodynamic (PK/PD) parameters of a degrader and directs its bioavailability[60]. Since the optimization of the linker moiety is typically a non-rational process of intense medicinal chemistry campaigns, *in silico* methods have been developed to serve as a guidance tool in choosing linker lengths that can potentially promote ternary complex formation[62].

Interestingly, the tight interplay of optimal linker properties to support ternary complex formation dependent on the E3 ligase, which is recruited by the degrader, can also lead to a surprisingly selective degrader molecule derived from non-selective or even promiscuous protein-binding scaffolds[63,64]. Furthermore, the unique mechanism of degradation can provide advantages over traditional inhibition effects, such as improved efficacy and potency, due to the complete removal of the target instead of mere catalytic inhibition[65]. These features, coupled with the potential to target otherwise "undruggable" targets, emphasize the immense opportunities this approach has to offer[65]. As a research tool, targeted protein degraders have been shown to be highly useful in facilitating relatively fast

protein degradation, which provides another attractive method to investigate disease biology[65]. However, the development of such probes is highly complex and time-consuming, since there is no good method available yet to confidently predict the design of a degrader and linker constitution, and overall physicochemical properties need to be optimized and can vary dramatically for each individual target protein and recruited E3 ligase[65]. Nevertheless, the first heterobifunctional targeted protein degraders have made it into clinical trials, which is a major step of this relatively young approach towards being used as an established therapeutic in the clinic[66].

In recent years, targeted protein degradation strategies that use other degradation machineries within the human cell have been applied[67] (Figure 11.1.6). For instance, lysosome targeting chimeras (LYTACs), which generally combine an antibody of a target protein with a small glycan-like moiety, enable endosome/lysosome-mediated degradation of their target[67]. In contrast to PROTACs, LYTACs are also able to degrade extracellular and membrane-bound targets[67]. Furthermore, autophagy-targeting chimeras (AUTACs) have been reported, which function similar to PROTACs *via* ubiquitination of their target protein[67]. However, AUTACs initiate a specific ubiquitination pattern that is recognized by the autophagy pathway[67]. Moreover, autophagosome-tethering compounds (ATTECs) have been described, which lead to autophagy-mediated degradation of a protein of interest without ubiquitinating the target, but by directly conjugating a protein to the autophagosome[67].

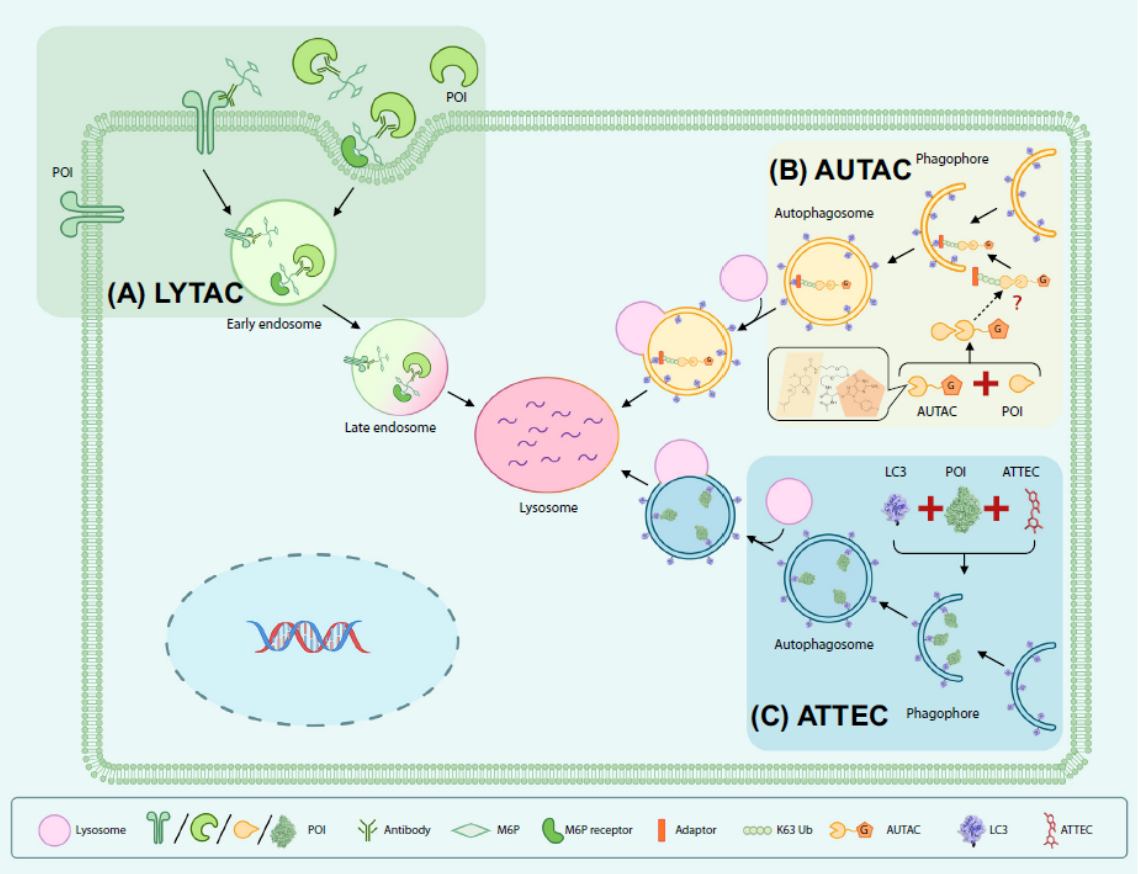


Figure 11.1.6 | Novel Targeted Degradation Approaches. Overview of recently reported targeted degradation approaches, involving the endosome/lysosome (LYTAC) and the autophagosome (AUTAC/ATTEC) machinery. Taken from Ding *et al.*, *Trends Pharmacol. Sci.* **2020**, *41* (7), 464-474[67].

Targeted Protein Degradation Using the Degradation Tag (dTAG) System

In 2018, Nabet, Roberts, Buckley, and coworkers reported the development of a novel technique to induce targeted protein degradation *via* a specific degradation tag (dTAG) system[68] (Figure 11.1.7). The group was able to show that by expressing fusion proteins of a protein of interest with the FKBP12^{F36V} mutant, it is possible to imitate the effects of a targeted protein degrader by degrading the fusion protein using FKBP12^{F36V}-specific degrader, dTAG-13[68]. This compound takes advantage of the "bump-and-hole" strategy, where a small change in the binding site leads to a highly selective probe, which induces CRBN-dependent FKBP12 mutant degradation, but cannot bind wild-type FKBP[69]. While the initial publication demonstrated the universal application of this novel technique by degrading BRD4, KRAS^{G12V}, HDAC1, MYC, EZH2 (enhancer of zeste 2 homolog 2), and PLK1, as well as its ability to be a useful tool in mice[68], further studies applying this method have since shown promising results, highlighting the value of this intriguing chemical biological tool[69].

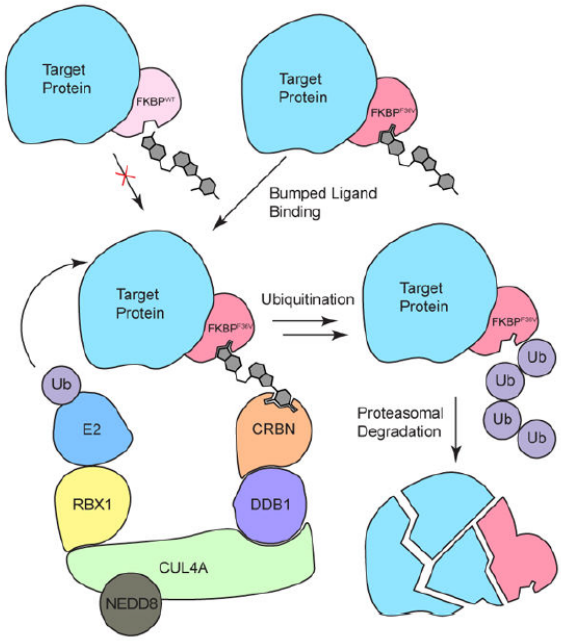


Figure 11.1.7 | The dTAG Approach for Targeted Protein Degradation. In the dTAG approach, a selective degrader molecule, which specifically targets an FKBP bump-hole mutant, is used to facilitate CRBN-mediated degradation of a fusion protein with a protein of interest. Adapted from Nabet *et al*., *Nat. Chem. Biol.* **2018**, *14* (5), 431-441[68].

11.2 Development of Potential PI5P4K Degraders

11.2.1 Aim of This Project

With the development and characterization of potent and selective PI5P4K inhibitor scaffolds in Chapter 5, another approach to target these lipid kinases was pursued. It was envisioned that, similar to the desthiobiotin-labeled probes in Chapter 5, THZ-P1-2 could be linked to an E3 ligase-recruiting moiety *via* a flexible linker. Since it was shown that PI5P4Ks seem to have not only kinase but also important scaffolding functions, a protein degrader would have the advantage of completely silencing PI5P4K activity. The overall goal therefore was to design, synthesize, and characterize a small library of THZ-P1-2-based, CRBN-recruiting protein degraders with varying linker lengths and compositions. Notably, while some of the developed degraders are currently investigated in other projects, the depiction of full compound structures in this chapter is avoided. Moreover, the synthetic route for the core scaffold has been already introduced in Chapter 5.

Design of PI5P4K-targeting Protein Degraders

When this project was initiated, the reversible CVM-based PI5P4K inhibitor scaffold was not yet fully characterized and optimized. In contrast, the covalent pan-PI5P4K inhibitor, THZ-P1-2, had already been identified as a selective and potent inhibitor of these lipid kinases, covalently labeling a cysteine on a disordered loop outside of the ATP-binding pocket. Moreover, desthiobiotin-labeled analogs of this inhibitor had already been successfully designed and synthesized, which feature a long linker extending from the covalent warhead.

Consequently, in a first approach, linkers of different length and composition were planned to be introduced in place of the electrophilic warhead (Figure 11.2.1). Removal of the Michael-acceptor moiety would potentially further improve the efficacy of the degraders, since one targeted protein degrader molecule is typically able to degrade many target protein molecules in a catalytic or sub-stoichiometric manner, whereas a covalent binder would only be able to target one target molecule. In addition to that, the reversible THZ-P1-2 analog (THZ-P1-2-R) exhibited weak PI5P4K activity, supporting the assumption that, even after removing the covalent warhead, the reversible scaffold would have sufficient PI5P4K binding affinity to facilitate degradation. Besides the linker composition, it was planned to also vary the attachment point of the linker itself. For this, some degraders were based on the full-length THZ-P1-2 scaffold, while others were designed based on a slightly truncated scaffold without the aminobenzoic acid moiety. This would help in reducing the overall size of the degraders, while keeping all essential features for PI5P4K binding intact (see reported co-crystal structure in Chapter 5).

Furthermore, two additional, related scaffolds of PI5P4K inhibitors were chosen as starting points for potential PI5P4K degraders, based on their biochemical potency. Since more elaborated SAR studies on the THZ-P1-2 scaffold (described in Chapter 5) were not yet completed when this project was started, using the additional scaffolds as reference was aimed at avoiding scaffold-specific bias.

Finally, thalidomide was chosen as the E3 ligase ligand for the first set of compounds, as this E3 ligase had been shown before to facilitate the degradation of many target proteins, while being a relatively small and easily accessible molecule.

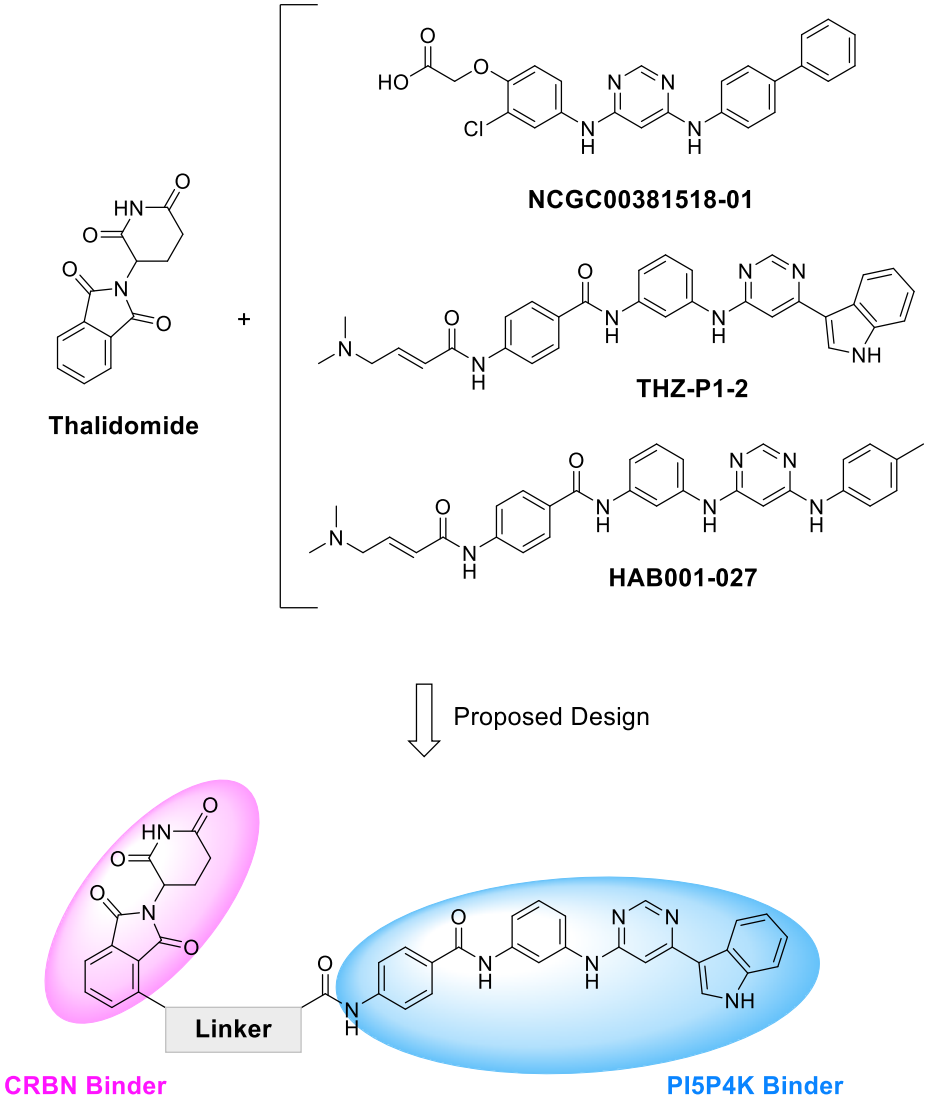


Figure 11.2.1 | Design Strategy to Access Potential PI5P4K Degraders. Based on the validated PI5P4K inhibitor THZ-P1-2 and related analogs NCGC00381518-01 and HAB001-027, potential targeted protein degrader molecules were designed. By introducing a linker, the PI5P4K binder (blue) was connected to an E3 ligase ligand (here: thalidomide; CRBN-binder; magenta), as shown in the general example of the degrader design at the bottom.

11.2.2 Results and Discussion

In this chapter, a brief overview is given on first efforts in the development of potential PI5P4K degraders. Since the synthetic routes for the PI5P4K inhibitors, THZ-P1-2 and HAB001-027 (compound 9 in Chapter 5), are already described in Chapter 5, NCGC00381518-01 intermediates were kindly provided by Fleur Ferguson in the lab, and further structural details of the degrader designs are not disclosed. Due to ongoing projects, no in-depth discussion of the synthetic efforts is provided. Rather, the results of cellular activities and the general strategy in the development of the designed probes is presented.

Biochemical Characterization of Potential PI5P4K Degraders

After initial synthesis of the designed potential PI5P4K degraders (Figure 11.2.2), the compounds were evaluated for PI5P4Kα and β inhibitory activity, as described in Chapters 5.3 and 5.4 (performed in collaboration with Matthew Hall's lab at NCATS). In total, seven different degraders were synthesized and characterized (Figure 11.2.3).

Surprisingly, while it was expected that removal of the covalent warhead would lead to a decrease in PI5P4K activity to some degree, most of the designed degraders completely lost PI5P4K activity. Since THZ-P1-2-R maintained PI5P4K inhibitory activity, it is likely that the linker moieties cause unfavorable target binding. However, **PIPDeg1** and **PIPDeg7** showed promising PI5P4K activity with IC_{50} values in the low-micromolar range (Figure 11.2.3). Because of their catalytic mode of action, targeted protein degraders typically do not require strong target affinities to induce degradation. It is therefore likely that the weak PI5P4K affinity of these compounds might still be sufficient for target protein degradation.

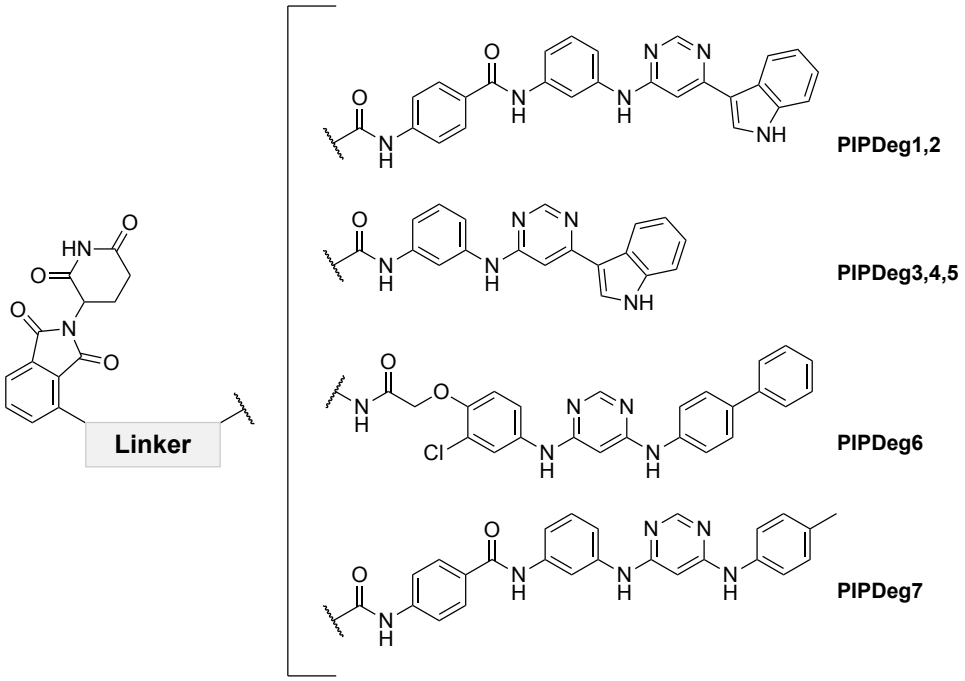


Figure 11.2.2 | General Structure of Synthesized Potential PI5P4K Degraders.

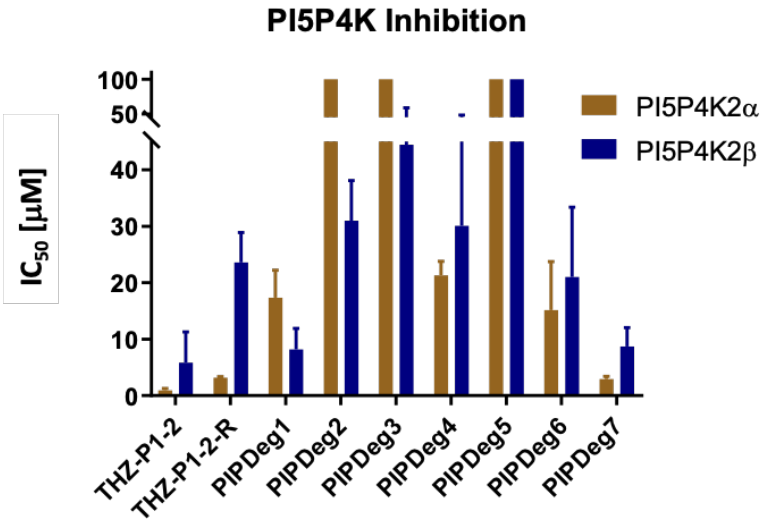


Figure 11.2.3 | Biochemical PI5P4K Inhibitory Activities of Potential Degraders. Biochemical inhibitory activities of potential PI5P4K degraders were determined in an ADP-Glo and Transcreener assay set-up for PI5P4Kα and β, respectively (as described in Chapter 5). THZ-P1-2 and its reversible analog, THZ-P1-2-R, are included for reference.

In the meantime, the cellular engagement of CRBN for the degraders was evaluated in HEK293T cells (performed in collaboration with Eric Fischer's lab at HMS/DFCI). While this assay provides valuable information about the ability of the designed molecules to engage the E3 ligase CRBN, it mainly functions as a validation tool to confirm cellular permeability of the degraders. The results of this assay are summarized in Table 11.2.1.

Generally, all designed degraders showed CRBN engagement in cells. Interestingly, degraders based on the same scaffold exhibited a noticeable range of cellular activity, pointing towards a strong effect of the linker composition influencing cellular efficacy. **PIPDeg1** and **PIPDeg7** were again amongst the most active analogs of the designed degraders in this assay, further providing promising initial data for these potential PI5P4K degraders.

Table 11.2.1 | Cellular CRBN Engagement Assay for Potential PI5P4K Degraders

Compound ID	IC_{50} [μM]
PIPDeg1	0.8
PIPDeg2	11.5
PIPDeg3	3.1
PIPDeg4	7.9
PIPDeg5	11.7
PIPDeg6	1.3
PIPDeg7	2.9

Evaluation of PI5P4K Degradation

In a final step, Jurkat cells were treated with the potential PI5P4K degraders at various concentrations, ranging from 10 nM to 10 µM, for 24 hours (Figure 11.2.4). The absolute protein concentration of all three PI5P4K isoforms was detected by Western blot analysis (data was kindly provided by Benjamin Fram in the lab). In summary, none of the tested degrader molecules showed PI5P4K degradation activity. While this could be due to the lack in affinity for the target protein for most of the tested degraders, even decent PI5P4K binders, such as **PIPDeg1** and **PIPDeg7**, did not show any activity in this cell line. These are also the compounds that showed sufficient cellular CRBN-binding and did not exhibit cellular uptake issues. Thus, it is highly likely that the linker might not facilitate optimal ternary complex formation in this case. It was further ruled out that PI5P4Ks might not be degradable, since some in-house data (not shown) of CRBN-recruiting, promiscuous degraders suggests degradability of PI5P4K.

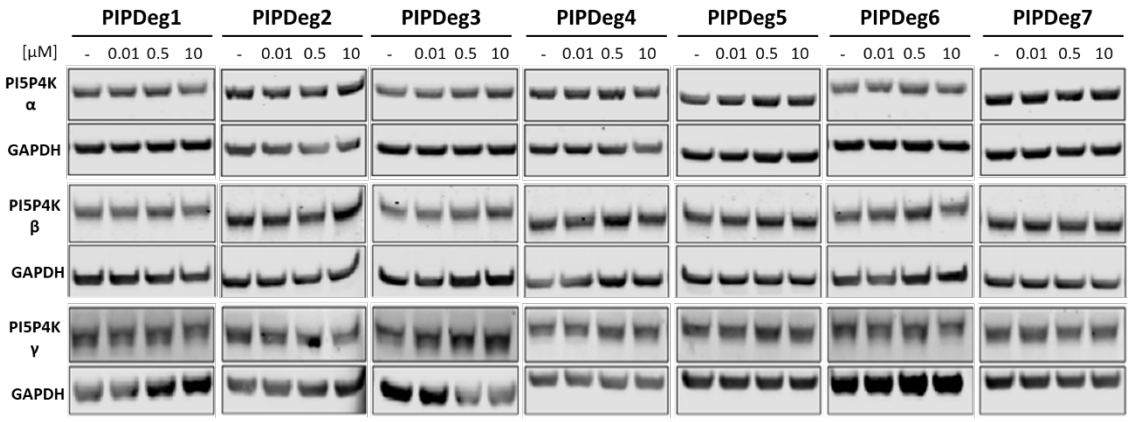


Figure 11.2.4 | PI5P4K Degrader Activity in Jurkat Cells. Jurkat cells were treated with increasing concentrations of potential PI5P4K degrader compound or DMSO for 24 hours. Total protein levels of PI5P4Kα, β, and γ were detected by Western blot analysis.

Following up on the hypothesis that the linkers used in this first batch of potential PI5P4K degraders were not able to support a ternary complex formation between PI5P4Ks and CRBN, an *in silico* docking of the two proteins, including their ligands, was performed (in collaboration with Eric Fischer's lab, HSM/DFCI). While the results (Figure 11.2.5) suggest that linkers, connecting the thalidomide analog (here: pomalidomide) and THZ-P1-2, probably need to be of eight carbons in length or longer, it is important to consider that the model did not include the flexible loop, which is normally covalently labeled by THZ-P1-2. Since this loop would be localized directly at the interface between CRBN and PI5P4K, it is likely that this disordered loop hinders ternary complex formation.

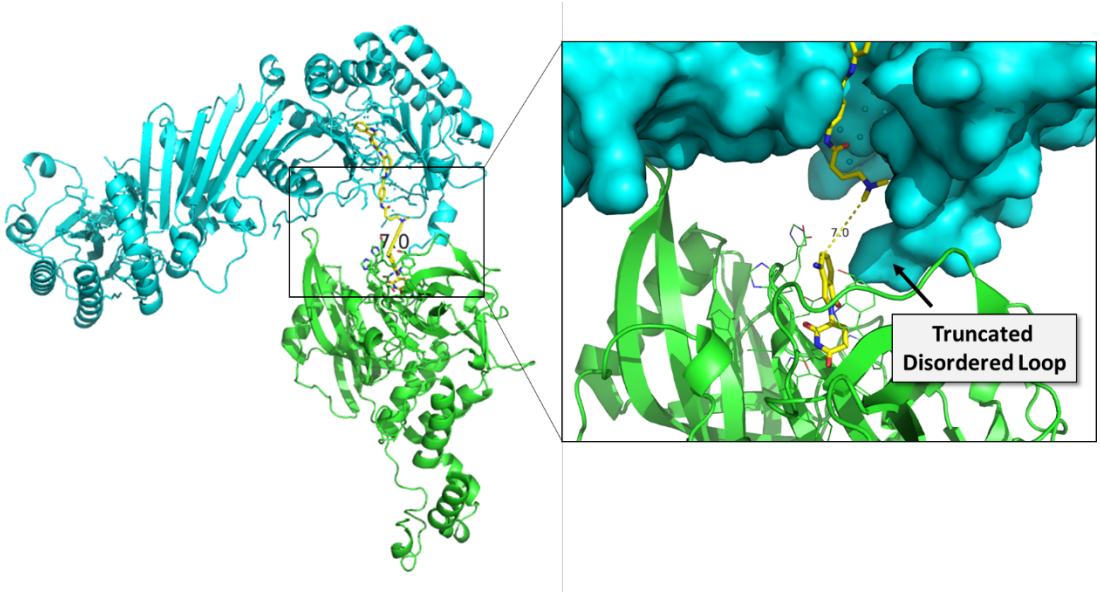


Figure 11.2.5 | Model of Ternary Complex Formation Between CRBN and PI5P4Kα. The co-crystal structures of PI5P4Kα (turquoise) with THZ-P1-2 (yellow sticks) and CRBN (green) with pomalidomide (yellow sticks) were docked and optimized for ternary complex formation features. The distance of the closest connection between pomalidomide and THZ-P1-2 was calculated as 7.0 Å.

11.2.3 Conclusion

In this project, the goal was to develop targeted protein degraders on the basis of a potent, covalent PI5P4K inhibitor, THZ-P1-2. While desthiobiotin-labeled probes of THZ-P1-2, with a linker extending from the covalent warhead tail of the scaffold, had been successfully applied in streptavidin-mediated pulldown assays (Chapter 5), so far none of the designed THZ-P1-2-based degrader molecules was able to induce PI5P4K degradation.

Many linker compositions caused complete loss of PI5P4K inhibitory activity of the compounds already in biochemical assays, indicating unfavorable compound features for binding the ATP-binding pocket. Cellular evaluation of the degraders' efficacy revealed that none of the compounds was able to sufficiently facilitate ternary complex formation between CRBN and PI5P4K. Further computational analysis suggested that the flexible loop, which is covalently labeled by THZ-1-2, is potentially jeopardizing the PPI, and would thus explain the observed results.

While some of the data in the lab suggested that PI5P4Ks are indeed degradable by CRBN, a different linker attachment site would be needed to circumvent issues caused by the disordered loop. Moreover, since this project was initially started, other potent PI5P4K-targeting scaffolds have been developed in the lab and might serve as better starting points for the development of PI5P4K-selective degraders. Another option would be to consider different E3 ligases, as this might change the required linker length and composition.

This study thus provided valuable insights into potential strategies for the development of PI5P4K degraders. These would be highly interesting tools, as PI5P4Ks are not only active kinases but also perform essential scaffolding functions, which would not be inhibited by a conventional kinase inhibitor (see Chapter 5). Further efforts towards the development of PI5P4K degraders are therefore currently ongoing in the lab.

11.3 Development of Potential Pin1 Degraders

11.3.1 Aim of This Project

In Chapter 9, the development and characterization of the potent and selective Pin1 inhibitor, sulfopin, is described. Similar to the PI5P4K inhibitors, a desthiobiotin-labeled analog of sulfopin was developed, which was an essential tool in characterizing this covalent inhibitor. Consequently, the possibility of designing sulfopin-based degrader molecules was investigated. The overall goal of this project was therefore to design and synthesize a small initial compound library based on a commercially available sulfopin analog to elucidate effects of different linkers. In a second step, a set of full-length sulfopin-based compounds, which should be analyzed in cellular assays, were to be synthesized. Since the discussed compounds are currently further characterized, no specific structures are provided.

Design Strategy for Potential Pin1 Degraders

As sulfopin is a covalent fragment inhibitor of the isomerase Pin1 with a small molecular weight, the linker for the desthiobiotin-labeled probes had to be thoroughly optimized, leading to sulfopin-DTB, as shown in Chapter 9. Consequently, a similar linker approach was chosen for the design of potential Pin1 degraders, introducing various linker lengths and compositions, extending from the *tert*-butyl moiety (Figure 11.3.1). Since the covalency of the fragment, due to the chloroacetamide warhead, is essential for target affinity, a covalent targeted protein degrader approach was investigated.

While the chemistry to access respective sulfopin analogs for the synthesis of potential Pin1 degraders was optimized, a commercially available analog (Pin1-3-Acid) was used for an initial screen for linker lengths and compositions. Finally, the CRBN ligand thalidomide was chosen as the E3 ligase-recruiting moiety due to its small molecular weight, chemical accessibility, and CRBN's reported ability to degrade a variety of proteins.

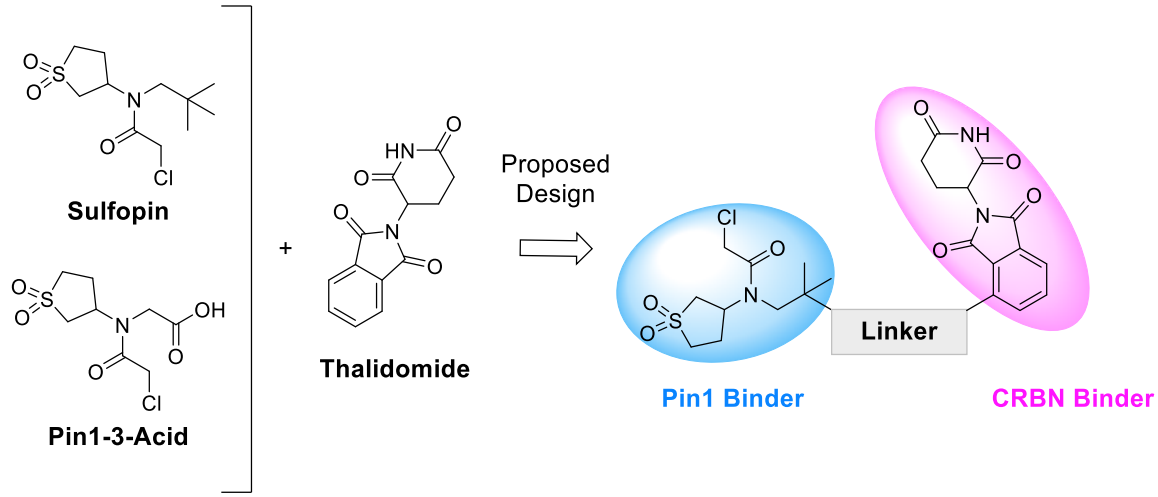


Figure 11.3.1 | Design Strategy for Potential Pin1 Degraders. Based on the validated Pin1 inhibitor sulfopin, potential targeted protein degrader molecules were designed. By introduction of a linker, the Pin1 binder (blue) was connected to an E3 ligase ligand (here: thalidomide; CRBN-binder; magenta), as shown in the general example of the degrader design at the right. To test initial linker designs, the commercially available sulfopin analog, Pin1-3-Acid, was used as the Pin1-binding moiety instead.

11.3.2 Results and Discussion

In this chapter, an overview of the first steps towards Pin1 degraders is given. Since the discussed compounds are currently further characterized, no specific structures are shown in this chapter. Thus, experimental procedures for the synthetic routes of the designed compounds are not disclosed. Nevertheless, this chapter provides an introduction to the development of Pin1 degraders and their characterization.

Determination of Target Affinity

Figure 11.3.2 summarizes the designed and synthesized sulfopin analogs for targeted degradation purposes of this chapter. As the target-binding moiety is a small fragment, addition of a thalidomide molecule *via* a linker drastically changes the properties of the parent compound. Thus, cLogP values are provided as this is an important feature for the following discussion. In total, six degraders were synthesized with the commercially available sulfopin analog, followed by five degrader molecules with the full-length sulfopin core structure.

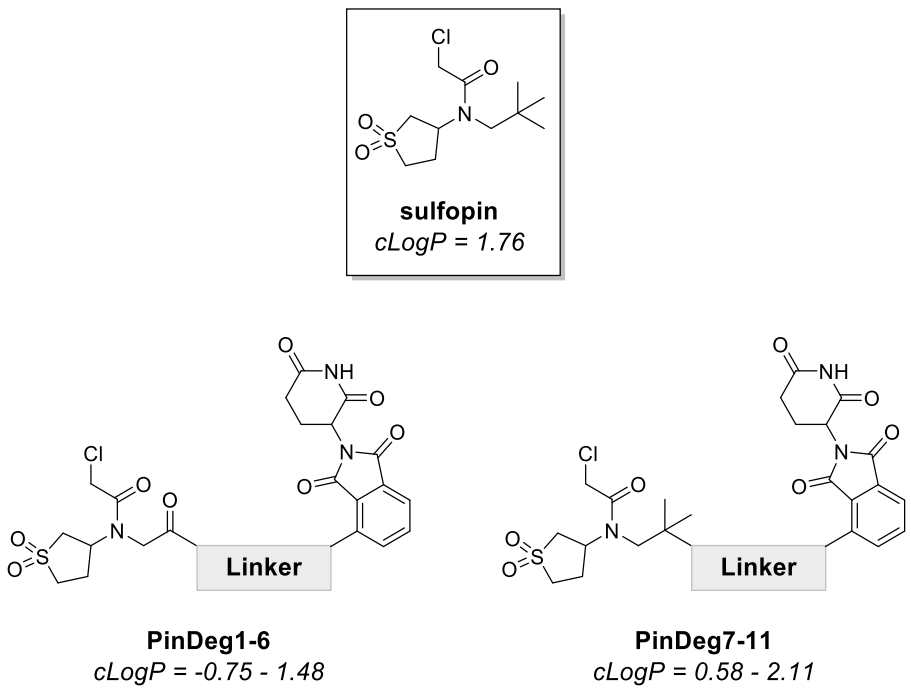


Figure 11.3.2 | Structural Overview of Potential Pin1 Degraders. Based on the characterized Pin1 inhibitor sulfopin, potential Pin1 degraders, as depicted at the bottom, were designed and synthesized. For further characterization, cLogP values are provided with sulfopin as reference.

In order to determine the compounds' Pin1 affinity, a competition pulldown experiment using the sulfopin-DTB probe (Chapter 9) in lysate from PATU-8988T cells was performed (Figure 11.3.3). After pre-incubation with the covalent degrader molecules, sulfopin-DTB was added, irreversibly labeling Pin1 that has not been bound by the potential degrader molecules. In general, all compounds that have the full sulfopin scaffold show potent Pin1 affinity, while the sulfopin analog-derived compounds only exhibit weak Pin1 binding. Interestingly, even though this assay does not require the compounds to be cell-permeable, compounds with a higher cLogP value showed more potent Pin1 labeling. This effect was

very prominent among the less potent analogs, based on the commercially available, sulfopin-related scaffold.

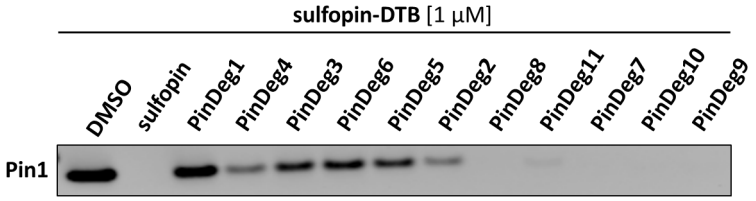


Figure 11.3.3 | Pin1 Affinity of Potential Pin1 Degraders. PATU-8988T cell lysate was treated with 10 µM sulfopin or degrader, or DMSO. The lysate samples were incubated at 4 °C overnight and subsequently treated with sulfopin-DTB for 1 hour. The desthiobiotin-labeled Pin1 was then purified by streptavidin-mediated pulldown and analyzed by Western blot analysis.

In the next step, the compounds were tested in a cellular CRBN binding assay using HEK293T cells (Table 11.3.1). Most compounds showed no to very low cellular CRBN binding affinity. Since some of the potential degraders are highly polar, these results were not completely surprising. However, some designed compounds have comparable cLogP values to sulfopin or higher and nonetheless show poor cellular CRBN binding. A potential explanation for this might be that sulfopin's small size enhances cellular uptake, even though its cLogP is rather low. By increasing the molecules' size, the degraders might therefore lose their ability to be cell-permeable. Notably, while a positive result in this assay has been a good indication for potent CRBN binding and thus cellular permeability, a compound with a negative result might still be able to permeate into the cell in a sufficient amount to induce degradation.

Table 11.3.1 | Cellular CRBN Engagement Assay for Potential Pin1 Degraders

Compound ID	IC_{50} [µM]	Compound ID	IC_{50} [µM]
PinDeg1	>10	**PinDeg7**	>10
PinDeg2	>10	**PinDeg8**	>10
PinDeg3	4.5	**PinDeg9**	>10
PinDeg4	>10	**PinDeg10**	>10
PinDeg5	>10	**PinDeg11**	~10.1
PinDeg6	>10		

Cellular Characterization of Potential Pin1 Degraders

Considering the results from the target engagement studies above, the most promising degrader molecules with the full-length sulfopin scaffold were chosen for cellular evaluation of their potential Pin1 degradation ability (Figure 11.3.4). PATU-8988T cells were treated with 10 µM degrader or DMSO for five hours. The cells were lysed and lysates incubated with sulfopin-DTB. This way, both total cellular Pin1 levels and compound-labeled Pin1 could be analyzed by Western blot analysis.

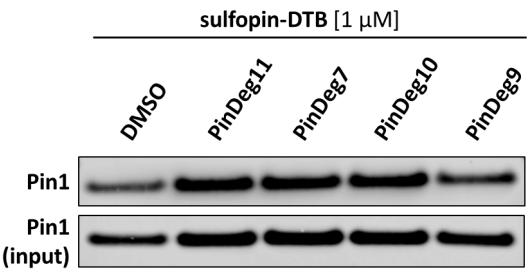


Figure 11.3.4 | Cellular Efficacy of Potential Pin1 Degraders. PATU-8988T cells were treated with 10 μM degrader or DMSO for 5 hours. The cell lysates were then incubated with sulfopin-DTB for 1 hour and desthiobiotin-labeled Pin1 enriched by streptavidin-mediated pulldown. Pin1 input (showing total Pin1 levels) and labeled Pin1 from the pulldown were analyzed by Western blot.

While none of the compounds showed Pin1 degradation effects, **PinDeg9** exhibited weak competition with the sulfopin-DTB probe, indicating a sufficient cellular concentration of this degrader in PATU-8988T cells. Consequently, even though the compound seems to bind Pin1 in cells, no productive ternary complex formation with CRBN is accomplished. To further validate **PinDeg9** activity in cells, a follow-up experiment was performed in which **PinDeg9** was incubated at different concentrations, ranging from 100 nM to 10 μM, for either eight or 24 hours (Figure 11.3.5). As expected, **PinDeg9** did not show any Pin1 degradation effects in this assay set-up. Interestingly, due to the low molecular weight of Pin1 itself, binding of the degrader probe caused a visible shift of the Pin1 band at high concentrations (10 μM) in the Western blot analysis. These results further verified that **PinDeg9** labels Pin1 at a concentration of 10 μM in cells.

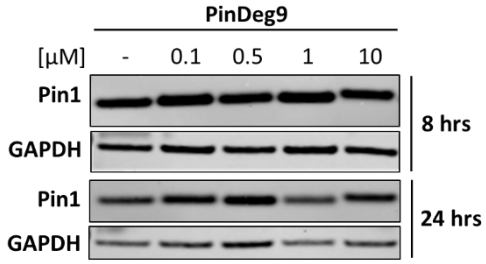


Figure 11.3.5 | Cellular Efficacy of PinDeg9. PATU-8988T cells were treated with different concentrations of PinDeg9 or DMSO for 8 or 24 hours. Total Pin1 levels were analyzed by Western blot analysis.

In the meantime, a molecular docking of CRBN and Pin1 was performed to simulate ternary complex formation mediated by a CRBN-based degrader, which can provide valuable information for the rational design of future Pin1 degraders (in collaboration with Eric Fischer's lab, HSM/DFCI) (Figure 11.3.6). Since sulfopin binds to a shallow pocket on Pin1, the resulting data suggests that linkers of as short as three to four carbon atoms should be able to facilitate efficient ternary complex formation.

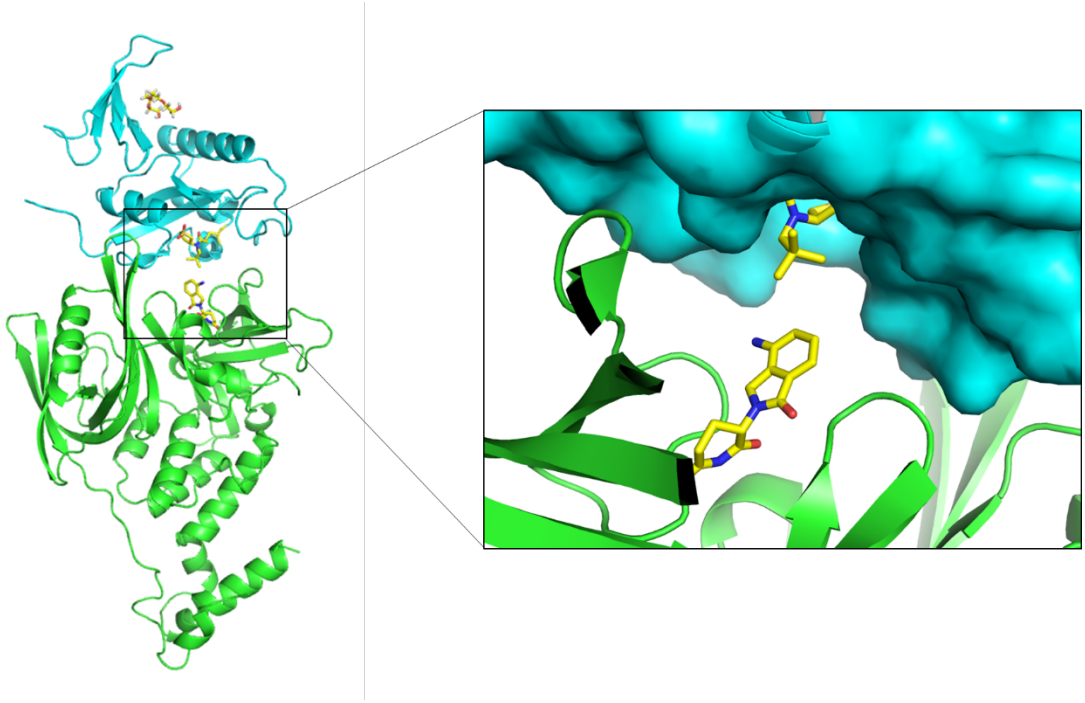


Figure 11.3.6 | Model of Ternary Complex Formation Between CRBN and Pin1. The co-crystal structures of Pin1 (turquoise) with sulfopin (yellow sticks) and CRBN (green) with pomalidomide (yellow sticks) were docked and optimized for ternary complex formation features. The distance of the closest connection between pomalidomide and sulfopin was calculated as ~3.3 Å.

11.3.3 Conclusion

With the development of sulfopin as a potent and selective, covalent Pin1 inhibitor, the goal of this project was to design, synthesize, and characterize sulfopin-based degraders. So far, none of the designed compounds was able to yield Pin1 degradation.

Due to the small size and high polarity of the parent scaffold, the main issue of these compounds has been a lack in cellular permeability. Thus, future designs of potential Pin1 degraders based on this scaffold need to be thoroughly optimized for higher cLogP values and cellular permeability. One option to achieve this goal would be to switch to a different E3 ligase. For example, reported VHL ligands are commonly much more hydrophobic than thalidomide and might help in increasing lipophilicity of the probes. Since the linker length should be relatively short, long, lipophilic linkers would likely not be sufficient to balance out the hydrophilicity of thalidomide and sulfopin to obtain a degrader molecule that exhibits good cell permeability and optimal Pin1 degrader properties.

In summary, this chapter provides the first results in the efforts to develop a Pin1-targeting protein degrader. These results nicely show the issue of cellular uptake for heterobifunctional degraders. Due to the fragment size of the parent compound, it turned out to be quite challenging to balance degrader lipophilicity and cellular permeability-determining properties, while providing optimal structural features for ternary complex formation. This data is thus highly valuable for further efforts towards the development of Pin1 degraders, which are currently ongoing in the lab.

11.3.4 Experimental

Cell Culture and Reagents
PATU-8988T cells were obtained from DSMZ and cultured in Dulbecco's modified Eagle's medium (DMEM) (including 10% fetal bovine serum (FBS; Sigma Aldrich) and 1% penicillin/streptomycin) at 37 °C. Incubators were humidified and exhibited 5% of CO_2 content. The cells were frequently tested for the absence of Mycoplasma infection.

Immunoblotting
Cell were collected and pelleted at 4 °C (300 g) for 5 min. After washing the cell pellet with 5 mL cold PBS buffer, cells were re-suspended in cell lysis buffer (50 mM Hepes, pH 7.4, 1 mM EDTA (ethylenediaminetetraacetic acid), 10% glycerol, 1 mM TCEP (tris(2-carboxyethyl)phosphine), 150 mM EDTA, 0.5% NP-40, and protease inhibitor tables (Roche)). Subsequent clarification of the resulting lysate at 14,000 rpm for 15 min at 4 °C yielded the desired cell lysate, which were quantified by BCA assay (Pierce).

Lysates were loaded into Bolt 4-12% Bis-Tris Gels (Thermo Fisher) and run at 90 V for 1.5 hours. Fully developed gels were blotted, using the iBlot Gel Transfer at P3 for 7 min. The nitrocellulose membranes were then blocked in Odyssey blocking buffer (LICOR Biosciences) for 1 hour at room temperature. The blots were further incubated with respective antibodies at 4 °C overnight, using a 20% Odyssey blocking buffer in 1X tris buffered saline with tween (TBST) buffer mixture.

After washing the membranes three times with 1X TBST buffer, the blots were incubated for another 1 hour at room temperature with a secondary, IRDye goat, anti-rabbit or goat, anti-mouse antibody (LICOR; 1:1,000), using a 20% Odyssey blocking buffer in 1X TBST buffer mixture. The membranes were washed three times with 1X TBST buffer, before being analyzed with an ODYSSEY Infrared Imaging System (LICOR).

Lysate Pulldown with Sulfopin-DTB
Lysate samples from lysed cell experiments or lysates, pre-treated with a compound of interest at the indicated concentrations, were then treated using the following protocol. PATU-8988T cell lysates (500 µL per sample) were treated with sulfopin-DTB (1 µM) for 1 hour at 4 °C. Streptavidin-coated beads (Thermo scientific) were diluted with lysis buffer (1:1). 30 µL of the bead slurry was added to each lysate sample and incubated for 1.5 hours at 4 °C. Beads were further washed with washing buffer (50 mM Hepes, pH 7.5, 10 mM NaCl, 1 mM EDTA, 10% glycerol) four times, pelleted, and dried. After addition of 30 µL of 2X LDS buffer (containing 5% β-mercaptoethanol) per sample, the beads were boiled at 95 °C for 5 min, cooled on ice for at least 30 min, and analyzed by Western blot.

Cellular Target Engagement Assay
PATU-8988T cells were plated in 6-well plates (100,000 cells per well) in 2 mL of media. A day after plating, cells were treated with DMSO or at the indicated concentrations with a compound of interest and for the indicated amount of time. Cells were then collected, lysed, and either treated with sulfopin-DTB for a pulldown experiment or directly used for immunoblotting, after being diluted with 4X LDS buffer (including 10% β-mercaptoethanol) in a ratio of 3:1, boiled (95 °C for 5 min), and cooled.

11.4 References

[1] Ciechanover, A. Intracellular protein degradation: From a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Best Pract. Res. Clin. Haematol.* **2017**, *30* (4), 341-355.

[2] Bondeson, D. P.; Crews, C. M. Targeted Protein Degradation by Small Molecules. *Annu. Rev. Pharmacol. Toxicol.* **2017**, *57*, 107-123.

[3] Cromm, P. M.; Crews, C. M. Targeted Protein Degradation: from Chemical Biology to Drug Discovery. *Cell Chem. Biol.* **2017**, *24* (9), 1181-1190.

[4] Kudriaeva, A. A.; Sokolov, A. V.; Belogurov, A. A. Stochastics of Degradation: The Autophagic-Lysosomal System of the Cell. *Acta Naturae* **2020**, *12* (1), 18-32.

[5] Prabha, C. R.; Mukherjee, S.; Raman, R.; Kulkarni, S. The ends and means of artificially induced target protein degradation. *Appl. Microbiol. Biotechnol.* **2012**, *96* (5), 1111-1123.

[6] Röth, S.; Fulcher, L. J.; Sapkota, G. P. Advances in targeted degradation of endogenous proteins. *Cell Mol. Life Sci.* **2019**, *76* (14), 2761-2777.

[7] Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J. Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U S A.* **2001**, *98* (15), 8554-8559.

[8] Collins, I.; Wang, H.; Caldwell, J. J.; Chopra, R. Chemical approaches to targeted protein degradation through modulation of the ubiquitin-proteasome pathway. *Biochem. J.* **2017**, *474* (7), 1127-1147.

[9] Vargesson, N. Thalidomide-Induced Teratogenesis: History and Mechanisms. *Birth Defects Res. C. Embryo Today* **2015**, *105* (2), 140-156.

[10] Ito, T.; Ando, H.; Suzuki, T.; Ogura, T.; Hotta, K.; Imamura, Y.; Yamaguchi, Y.; Handa, H. Identification of a Primary Target of Thalidomide Teratogenicity. *Science* **2010**, *327* (5971), 1345-1350.

[11] Bhogaraju, S.; Dikic, I. A peek into the atomic details of thalidomide's clinical effects. *Struct. Mol. Biol.* **2014**, *21* (9), 739-740.

[12] Pettersson, M.; Crews, C. PROteolysis Targeting Chimeras (PROTACs) – Past, Present and Future. *Drug Discov. Today Technol.* **2019**, *31*, 15-27.

[13] Lee, H.; Puppala, D.; Choi, E.-Y.; Swanson, H.; Kim, K.-B. Targeted Degradation of the Aryl Hydrocarbon Receptor by the PROTAC Approach: A Useful Chemical Genetic Tool. *ChemBioChem* **2007**, *8* (17), 2058-2062.

[14] Cyrus, K.; Wehenkel, M.; Choi, E.-Y.; Lee, H.; Swanson, H.; Kim. K.-B. Jostling for Position: Optimizing Linker Location in the Design of Estrogen Receptor-targeting PROTACs. *ChemMedChem* **2010**, *5* (7), 979-985.

[15] Lu, J.; Qian, Y.; Altieri, M.; Dong, H.; Wang, J.; Raina, K.; Hines, J.; Winkler, J. D.; Crew, A. P.; Coleman, K.; Crews, C. M. Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem. Biol.* **2015**, *22* (6), 755-763.

[16] Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E. Selective Target Protein Degradation via Phthalimide Conjugation. *Science* **2015**, *348* (6241), 1376-1381.

[17] Wu, T.; Yoon, H.; Xiong, Y.; Dixon-Clarke, S. E.; Nowak, R. P.; Fischer, E. S. Targeted protein degradation as a powerful research tool in basic biology and drug target discovery. *Nat. Struct. Mol. Biol.* **2020**, *27* (7), 605-614.

[18] Powell, C. E.; Gao, Y.; Tan, L.; Donovan, K. A.; Nowak, R. P.; Loehr, A.; Bahcall, M.; Fischer, E. S.; Jänne, P. A.; Goerge, R. E.; Gray, N. S. Chemically Induced Degradation of Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* **2018**, *61* (9), 4249-4255.

[19] Yang, S.-J.; Jeon, S.-J.; Van Nguyen, T.; Deshaies, R. J.; Park, C.-S.; Lee, K. M. Ubiquitin-dependent proeasomal degradation of AMPK gamma subunit by Cereblon inhibits AMPK activity. *Biochim. Biophys. Acta Mol. Cell Res.* **2020**, *1867* (8), 118729.

[20] Kurihara, T.; Asahi, T.; Sawamura, N. Cereblon-mediated degradation of the amyloid precursor protein via the ubiquitin-proteasome pathway. *Biochem. Biophys. Res. Commun.* **2020**, *524* (1), 236-241.

[21] Salami, J.; Alabi, S.; Willard, R. R.; Vitale, N. J.; Wang, J.; Dong, H.; Jin, M.; McDonnell, D. P.; Crew, A. P.; Neklesa, T. K.; Crews, C. M. Androgen receptor degradation by the proteolysis-targeting chimera ARCC-4 outperforms enzalutamide in cellular models of prostate cancer drug resistance. *Commun. Biol.* **2018**, *1*, 100.

[22] Wang, Z.; He, N.; Guo, Z.; Niu, C.; Song, T.; Guo, Y.; Cao, K.; Wang, A.; Zhu, J.; Zhang, X.; Zhang, Z. Proteolysis Targeting Chimeras for the Selective Degradation of Mcl-1/Bcl-2 Derived from Nonselective Target Binding Ligands. *J. Med. Chem.* **2019**, *62* (17), 8152-8163.

[23] Khan, S.; Zhang, X.; Lv. D.; Zhang, Q.; He, Y.; Zhang, P.; Liu, X.; Thummuri, D.; Yuan, Y.; Wiegand, J. S.; Pei, J.; Zhang, W.; Sharma, A.; McCurdy, C. R.; Kuruvilla, V. M.; Baran, N.; Ferrando, A. A.; Kim, Y.-m.; Rogojina, A.; Houghton, P. J.; Huang, G.; Hromas, R.; Konopleva, M.; Zheng, G.; Zhou, D. A selective BCL-X_L PROTAC degrader achieves safe and potent antitumor activity. *Nat. Med.* **2019**, *25* (12), 1938-1947.

[24] Burslem, G. M.; Schultz, A. R.; Bondeson, D. P.; Eide, C. A.; Stevens, S. L. S.; Druker, B. J.; Crews, C. M. Targeting BCR-ABL1 in Chronic Myeloid Leukemia by PROTAC-mediated Targeted Protein Degradation. *Cancer Res.* **2019**, *79* (18), 4744-4753.

[25] Han, X.-R.; Chen, L.; Wie, Y.; Yu, W.; Chen, Y.; Zhang, C.; Jiao, B.; Shi, T.; Sun, L.; Zhang, C.; Xu, Y.; Lee, M. R.; Luo, Y.; Plewe, M. B.; Wang, J. Discovery of Selective Small Molecule Degraders of BRAF-V600E. *J. Med. Chem.* **2020**, *63* (8), 4069-4080.

[26] Zhao, B.; Burgess, K. TrkC-Targeted Kinase Inhibitors And PROTACs. *Mol. Pharm.* **2019**, *16* (10), 4313-4318.

[27] Zoppi, V.; Hughes, S. J.; Maniaci, C.; Testa, A.; Gmaschitz, T.; Wieshofer, C.; Koegl, M.; Riching, K. M.; Daniels, D. L.; Spallarossa, A.; Ciulli, A. Iterative Design and Optimization of Initially Inactive Proteolysis Targeting Chimeras (PROTACs) Identify VZ185 as a Potent, Fast, and Selective von Hippel-Lindau (VHL) Based Dual Degrader Probe of BRD9 and BRD7. *J. Med. Chem.* **2019**, *62* (2), 699-726.

[28] Dobrovolsky, D.; Wang, E. S.; Morrow, S.; Leahy, C.; Faust, T.; Nowak, R. P.; Donovan, K. A.; Yang, G.; Li, Z.; Fischer, E. S.; Treon, S. P.; Weinstock, D. M.; Gray, N. S. Bruton tyrosine kinase degradation as a therapeutic strategy for cancer. *Blood* **2019**, *133* (9), 952-961.

[29] Teng, M.; Jian, J.; He, Z.; Kwiatkowski, N. P.; Donovan, K. A.; Mills, C. E.; Victor, C.; Hatcher, J. M.; Fischer, E. S.; Sorger, P. K.; Zhang, T.; Gray, N. S. Development of CDK2 and CDK5 Dual Degrader TMX-2172. *Angew. Chem. Int. Ed. Engl.* **2020**, *59* (33), 13865-13870.

[30] Brand, M.; Jiang, B.; Bauer, S.; Donovan, K. A.; Liang, Y.; Wang, E. S.; Nowak, R. P.; Yuan, J. C.; Zhang, T.; Kwiatkowski, N.; Müller, A. C.; Fischer, E. S.; Gray, N. S.; Winter, G. E. Homolog-selective degradation as a strategy to probe the function of CKD6 in AML. *Cell Chem. Biol.* **2019**, *26* (2), 300-306.

[31] Anderson, N. A.; Cryan, J.; Ahmed, A.; Dai, H.; McGonagle, G. A.; Rozier, C.; Benowitz, A. B. Selective CDK6 degradation mediated by cereblon, VHL, and novel IAP-recruiting PROTACs. *Bioorg. Med. Chem. Lett.* **2020**, *30* (9), 127106.

[32] Robb, C. M.; Contreras, J. I.; Kour, S.; Taylor, M. A.; Abid, M.; Sonawane, Y. A.; Zahid, M.; Murry, D. J.; Natarajan, A.; Rana, S. Chemically induced degradation of CDK9 by a proteolysis targeting chimera (PROTAC). *Chem. Commun.* **2017**, *53* (54), 7577-7580.

[33] Chen, H.; Chen, F.; Liu, N.; Wang, X.; Gou, S. Chemically induced degradation of CK2 by proteolysis targeting chimeras based on a ubiquitin-proteasome pathway. *Bioorg. Chem.* **2018**, *81*, 536-544.

[34] He, K.; Zhang, Z.; Wang, W.; Zheng, X.; Wang, X.; Zhang, X. Discovery and biological evaluation of proteolysis targeting chimeras (PROTACs) as an EGFR degraders based on osimertinib and lenalidomide. *Bioorg Med. Chem. Lett.* **2020**, *30* (12), 127167.

[35] Roberts, B. L.; Ma, Z.-X.; Gao, A.; Leisten, E. D.; Yin, D.; Xu, W.; Tang, W. Two-Stage Strategy for Development of Proteolysis Targeting Chimeras and its Application for Estrogen Receptor Degraders. *ACS Chem. Biol.* **2020**, *15* (6), 1487-1496.

[36] Gao, H.; Zheng, C.; Du, J.; Wu, Y.; Sun, Y.; Han, C.; Kee, K.; Rao, Y. FAK-targeting PROTAC as a Chemical Tool for the Investigation of Non-Enzymatic FAK Function in Mice. *Protein Cell* **2020**, *11* (7), 534-539.

[37] Cromm, P. M.; Samarasinghe, K. T. G.; Hines, J.; Crews, C. M. Addressing Kinase-Independent Functions of Fak via PROTAC-Mediated Degradation. *J. Am. Chem. Soc.* **2018**, *140* (49), 17019-17026.

[38] Zhang, X.; Crowley, V. M.; Wucherpfennig, T. G.; Dix, M. M.; Cravatt, B. F. Electrophilic PROTACs that degrade nuclear proteins by engaging DCAF16. *Nat. Chem. Biol.* **2019**, *15* (7), 737-746.

[39] Yang, H.; Lv, W.; He, M.; Deng, H.; Li, H.; Wu, W.; Rao, Y. Plasticity in designing PROTACs for selective and potent degradation of HDAC6. *Chem. Commun.* **2019**, *55* (98), 14848-14851.

[40] Smalley, J. P.; Adams, G. E.; Millard, C. J.; Song, Y.; Norris, J. K. S.; Schwabe, J. W. R.; Cowley, S. M.; Hodgkinson, J. T. PROTAC-mediated degradation of class I histone deacetylase enzymes in corepressor complexes. *Chem. Commun.* **2020**, *56* (32), 4476-4479.

[41] Manda, S.; Lee, N. K.; Oh, D.-C.; Lee, J. Design, Synthesis, and Biological Evaluation of Proteolysis Targeting Chimeras (PROTACs) for the Dual Degradation of IGF-1R and Src. *Molecules* **2020**, *25* (8), 1948.

[42] Nunes, J.; McGonagle, G. A.; Eden, J.; Kiritharan, G.; Touzet, M.; Lewell, X.; Emery, J.; Eidam, H.; Harling, J. D.; Anderson, N. A. Targeting IRAK4 for Degradation with PROTACs. *ACS Med. Chem. Lett.* **2019**, *10* (7), 1081-1085.

[43] Shah, R. R.; Redmond, J. M.; Mihut, A.; Menon, M.; Evans, J. P.; Murphy, J. A.; Bartholomew, M. A.; Coe, D. M. Hi-JAK-ing the ubiquitin system: The design and physicochemical optimisation of JAK PROTACs. *Bioorg. Med. Chem.* **2020**, *28* (5), 115326.

[44] Li, Y.; Yang, J.; Aguilar, A.; McEachern, D.; Przybranowski, S.; Liu, L.; Yang, C.-Y.; Wang, M.; Han, X.; Wang, S. Discovery of MD-224 as a First-in-Class, Highly Potent and Efficacious PROTAC MDM2 Degrader Capable of Achieving Complete and Durable Tumor Regression. *J. Med. Chem.* **2019**, *62* (2), 448-466.

[45] Vollmer, S.; Cunoosamy, D.; Lv, H.; Feng, H.; Li, X.; Nan, Z.; Yang, W.; Perry, M. W. D. Design, Synthesis, and Biological Evaluation of MEK PROTACs. *J. Med. Chem.* **2020**, *63* (1), 157-162.

[46] Winzker, M.; Friese, A.; Koch, U.; Janning, P.; Ziegler, S.; Waldmann, H. Development of a PDEδ-Targeting PROTACs that Impair Lipid Metabolism. *Angew. Chem. Int. Ed. Engl.* **2020**, *59* (14), 5595-5601.

[47] Mares, A.; Miah, A. H.; Smith, I. E. D.; Rackham, M.; Thawani, A. R.; Cryan, J.; Haile, P. A.; Votta, B. J.; Beal, A. M.; Capriotti, C.; Reilly, M. A.; Fisher, D. T.; Zinn, N.; Bantscheff, M.; MacDonald, T. T.; Vossenkamper, A.; Dace, P.; Churcher, I.; Benowitz, A. B.; Watt, G.; Denyer, J.; Scott-Stevens, P.; Harling, J. D. Extended pharmacodynamic responses observed upon PROTAC-mediated degradation of RIPK2. *Comun. Biol.* **2020**, *3* (1), 140.

[48] Wang, M.; Lu, J.; Wang, M.; Yang, C.-Y.; Wang, S. Discovery of SHP2-D26 as a First, Potent, and Effective PROTAC Degrader of SHP2 Protein. *J. Med. Chem.* **2020**, *63* (14), 7510-7528.

[49] Schiedel, M.; Herp, D.; Hammelmann, S.; Swyter, S.; Lehotzky, A.; Robaa, D.; Oláh, J.; Ovádi, J.; Sippl, W.; Jung, M. Chemically Induced Degradation of Sirtuin 2 (Sirt2) by a Proteolysis Targeting Chimera (PROTAC) Based on Sirtuin Rearranging Ligands (SirReals). *J. Med. Chem.* **2018**, *61* (2), 482-491.

[50] Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steuerer, S.; Karolyi-Oezguer, J.; Riedmueller, C.; Gmaschitz, T.; Wachter, J.; Dank, C.; Galant, M.; Sharps, B.; Rumpel, K.; Traxler, E.; Gerstberger, T.; Schnitzer, R.; Petermann, O.; Greb, P.; Weinstabl, H.; Bader, G.; Zoephel, A.; Weiss-Puxbaum, A.; Ehrenhöfer-Wölfer, K.; Wöhrle, S.; Boehmelt, G.; Rinnenthal, J.; Arnhof, H.; Wiechens, N.; Wu, M.-Y.; Owen-Hughes, T.; Ettmayer, P.; Pearson, M.; McConnell, D. B.; Ciulli, A. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **2019**, *15* (7), 672-680.

[51] Zhou, H.; Bai, L.; Xu, R.; Zhao, Y.; Chen, J.; McEachern, D.; Chinnaswamy, K.; Wen, B.; Dai, L. Kumar, P.; Yang, C.-Y.; Liu, Z.; Wang, M.; Liu, L.; Meagher, J. L.; Yi, H.; Sun, D.; Stuckey, J. A.l; Wang, S. Structure-Based Discovery of SD-36 as a Potent, Selective, and Efficacious PROTAC Degrader of STAT3 Protein. *J. Med. Chem.* **2019**, *62* (24), 11280-11300.

[52] Silva, M. C.; Ferguson, F. M.; Cai, Q.; Donovan, K. A.; Nandi, G.; Patnaik, D.; Zhang, T.; Huang, H.-T.; Lucente, D. E.; Dickerson, B. C.; Mitchison, T. J.; Fischer, E. S.; Gray, N. S.; Haggarty, S. J. Targeted degradation of aberrant tau in frontotemporal dementia patient-derived neuronal cell models. *eLife* **2019**, *8*, e45457.

[53] Crew, A. P.; Raina, K.; Dong, H.; Qian, H.; Wang, J.; Vigil, D.; Serebrenik, Y. V.; Hamman, B. D.; Morgan, A.; Ferraro, C.; Siu, K.; Neklesa, T. K.; Winkler, J. D.; Coleman, K. G.; Crews, C. M. Identification and Characterization of Von Hippel-Lindau-Recruiting Proteolysis Targeting Chimeras (PROTACs) of TANK-Binding Kinase 1. *J. Med. Chem.* **2018**, *61* (2), 583-598.

[54] Zhang, Y.; Loh, C.; Chen, J.; Mainolfi, N. Targeted protein degradation mechanisms. *Drug Discov. Today Technol.* **2019**, *31*, 53-60.

[55] Wu, T.; Yoon, H.; Xiong, Y.; Dixon-Clarke, S. E.; Nowak, R. P.; Fischer, E. S. Targeted protein degradation as a powerful research tool in basic biology and drug target discovery. *Nat. Struct. Mol. Biol.* **2020**, *27* (7), 605-614.

[56] Kaji, T.; Koga, H.; Kuroha, M.; Akimoto, T.; Hayata, K. Characterization of cereblon-dependent targeted protein degrader by visualizing the spatiotemporal ternary complex formation in cells. *Sci. Rep.* **2020**, *10* (1), 3088.

[57] Riching, K. M.; Mahan, S.; Corona, C. R.; McDougall, M.; Vasta, J. D.; Robers, M. B.; Urh, M.; Daniels, D. L. Quantitative Live-Cell Kinetic Degradation and Mechanistic Profiling of PROTAC Mode of Action. *ACS Chem. Biol.* **2018**, *13* (9), 2758-2770.

[58] Hanzl, A.; Winter, G. E. Targeted protein degradation: current and future challenges. *Curr. Opin. Chem. Biol.* **2020**, *56*, 35-41.

[59] Cantrill, C.; Chaturvedi, P.; Rynn, C.; Schaffland, J. P.; Walter, I.; Wittwer, M. B. Fundamental aspects of DMPK optimization of targeted protein degraders. *Drug Discov. Today* **2020**, *25* (6), 969-982.

[60] Chamberlain, P. P.; Hamann, L. G. Development of targeted protein degradation therapeutics. *Nat. Chem. Biol.* **2019**, *15* (10), 937-944.

[61] Cyrus, K.; Wehenkel, M.; Choi, E.-Y.; Han, H.-J.; Lee, H.; Swanson, H.; Kim, K.-B. Impact of linker length on the activity of PROTACs. *Mol. Biosyst.* **2011**, *7* (2), 359-364.

[62] Drummond, M. L.; Williams, C. I. *In Silico* Modeling of PROTAC-Mediated Ternary Complexes: Validation and Application. *J. Chem. Inf. Model* **2019**, *59* (4), 1634-1644.

[63] Müller, M. P.; Rauh, D. Try Me: Promiscuous Inhibitors Still Allow for Selective Targeted Protein Degradation. *Cell Chem. Biol.* **2018**, *25* (1), 4-6.

[64] Chamberlain, P. P. Targeted Protein Degradation for Kinase Selectivity. *Cell Chem. Biol.* **2019**, *26* (3), 307-308.

[65] Kostic, M.; Jones, L. H. Critical Assessment of Targeted Protein degradation as a Research Tool and Pharmacological Modality. *Trends Pharmacol. Sci.* **2020**, *41* (5), 305-317.

[66] Mullard, A. First targeted protein degrader hits the clinic. *Nat. Rev. Drug Discov.* **2019**; DOI: 10.1038/d41573-019-00043-6. (*online ahead of print*)

[67] Ding, Y.; Fei, Y.; Lu, B. Emerging New Concepts of Degrader Technologies. *Trends Pharmacol. Sci.* **2020**, *41* (7), 464-474.

[68] Nabet, B.; Roberts, J. M.; Buckley, D. L.; Paulk, J.; Dastjerdi. S.; Yang, A.; Leggett, A. L.; Erb, M. A.; Lawlor, M. A.; Souza, A.; Scott, T. G.; Vittori, S.; Perry, J. A.; Qi, J.; Winter, G. E.; Wong, K.-K.; Gray, N. S.; Bradner, J. E. The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **2018**, *14* (5), 431-441.

[69] Yesbolatova, A.; Tominari, Y.; Kanemaki, M. T. Ligand-induced genetic degradation as a tool for target validation. *Drug Discov. Today Technol.* **2019**, *31*, 91-98.