Title
Applications and confounds in drug discovery and repurposing

Permalink
https://escholarship.org/uc/item/14q5b0wp

Author
Tummino, Tia Angelina

Publication Date
2023

Supplemental Material
https://escholarship.org/uc/item/14q5b0wp#supplemental

Peer reviewed|Thesis/dissertation

Applications and confounds in drug discovery and repurposing

by
Tia A. Tummino

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Brian K. Shoichet
_____
Chair

Allan Basbaum
_____

Aashish Manglik
_____

_____

_____
Committee Members

Dedication

To the underdogs.

Acknowledgements

As with many things in life, *it takes a village*, and a PhD is no different. I would not have finished this PhD if it were not for the many people who have supported me, loved me, and encouraged me along the way.

First, I would like to thank my advisor Brian Shoichet for never giving up on me, and always pushing me to do my best work. Thank you for accepting me into your lab when I didn't know anything more than that I was interested in learning and contributing something cool. Thank you for letting me cry on your red couch many, many times. Thank you for guiding me as best as my stubborn personality would let you. Lastly, thank you for letting me go to Canada every summer and to Italy for my wedding even when I had a LOT of work to do.

Next, I would like to thank my thesis committee Allan Basbaum and Aashish Manglik. Your passion for science, command of scientific literature, gentle pushing, and endless support always made me feel better and gave me fresh inspiration that I often lacked. I am so grateful for all the thesis and non-thesis meetings I have spent learning from you. I hope to support others in the future half as well as you have supported me.

Thank you to the Shoichet, Irwin, and Kokel lab members- past and present. I have learned so much from all of you. It's always hard seeing people move on, but it's been wonderful watching you all go on to succeed. To Reed, JK, BJB, and Stefan: thank you

for mentoring me and teaching me everything I know about docking. To John Irwin: you always know how to make us all laugh. Thanks for keeping things fun and rigorous. To Anat, Isha, Fangyu, Olivier, Andrii, Moira, Sijie, Yujin, Shiming, Matt, Alina, JJ, Ben, Khanh, and anyone else I may have missed: thank you for being smart, funny, kind people. I loved working beside you, chatting about science and non-science topics. Thank you for making me feel welcome. To Matt, Douglas, Cole, Darya, and Reid: thank you for being there at the start of this journey. We have officially all survived.

To my many collaborators: João, Christos, Andy, Assaf, Kaavya, Evan, Yuki, Henry, Francesca, Francois, Kris, Marco, Nevan, Ruth, everyone at Enamine and Bienta, and everyone else who helped along the way, thanks for being great partners, running assays when I desperately needed them, giving great feedback, and helping make my projects possible. I could not have done it without you. Working with and learning from you all is the highlight of my time in graduate school. To my other mentor, Deanna Kroetz, and my qualification exam committee Mike Keiser and Kaveh Ashrafi, thank you for helping me get to a point where I could succeed. Thank you for always being there for me throughout this journey. I am grateful to know you. To Rebecca, I will always tell people you were the best admin at UCSF.

To the Colorado College Neuroscience program and Department of Psychology-Bob, Kristi, Tomi-Ann, Tricia, Kevin, Jason, John, Mark, and Ann: thank you for giving me the foundations that have led me to succeed. To Lori: I have always looked up to you and relied on you for advice and support. I am so grateful to have you in my life. Thank you

for making me the woman-of-science I am today.

Thank you to my many friends at UCSF, in PSPG, other graduate programs, and other labs. To Bianca, Annamarie, Mikayla, Dina, Kendra, Chase, Ki, Meghan and our honorary member Jenny: thanks for being the best PSPG pals ever. Grad school was so much fun with you by my side. To Greg Lee: thanks for being my first mentor at UCSF and for helping me become a better scientist and always making me laugh. And of course, thank you for not attending the Spark event where I eventually met Ryan, I owe it all to you. To Kyle, Rachel, Nick, Holly, and everyone else at UCSF: thanks for beer hours, chit-chat, and funny stories that make the days go by quickly. To Elissa and Izzy: thanks for the long lunches that turn into Starbucks runs, the dog photos, and for being great friends to me no matter how much I complain. I will miss working with you both so closely, all the giggles, and especially all the spreetz' (Aperol honestly deserves its own acknowledgement). To Megan and Capria: thank you both for being by my side, cheering me on through everything that life has thrown us these last six years. You are both inspiring, strong women who have given me some of the best friendship I have ever been showered with. I can't wait to continue watching you both succeed and flourish, as well as continuing all the many ridiculous and fun shenanigans that we get up to. I love being M & the ia's with you.

Thank you to my friends and family outside of UCSF, in Colorado and beyond: Aunt Jenn, Uncle Frank, Jamie, Matt, Brynleigh and Braylon, Frankie, Danielle, Melissa, Thomas, and Archie, Kayce, Kenny, Beau and Bear, Jonah, Drew, Ryan, Nicole, Krissy,

Reilly, April and Richie, and Nate: I am so grateful for your love, support, and friendships and all the fun distractions you've provided me with over the years. Thank you for not asking too often when I was graduating, for being understanding when I forget to call or text you back, and for pretending to care when I tried to discuss science with you. Here's to more free time, Russian river floats, baseball games, hikes, and camping trips with you all in the future!

Thank you to the Muir and Moffat clans. You are the best in-laws a girl could ask for. It's been a blast being so close to you all during this time of my life. And thank you for the many trips to Tahoe that have kept me sane along the way. A special thank you to my brother-in-law Recneps, who will always make me smile no matter how bad my day was.

Thank you to Ohana- Trish, Sally, Sylvain, Dominique, Chloé, and Gaëlle- who took me in and have made me a part of the family. When I was growing up I didn't know what "graduate school" was, or that there was any other option besides being a medical doctor. Thank you for helping me find my way and loving me like your own. I can't wait to give that unconditional love and support to the next generation, Sébastien et Zoé. I promise to work on my French now that I will have more free time.

Thank you to Winnie the Pooh, my work-from-home partner and eerily human doodle. It's hard to know how much you can love something until you have a living being who relies on you for survival but having you around has made the long hours feel shorter

and the lonely work-from-home years feel full. I promise more hikes, hose-play, swimming, and walks to the dog park are in your future.

Lastly, thank you to "Nayr nugget". I honestly would have given up on this whole pursuit if it wasn't for you. Thank you for believing in me, especially when I don't believe in myself. Thank you for teaching me how to be a better scientist, prism graph maker, and overall better person. Thank you for forcing me to have some work-life balance and for working with me as a team. It's hard to put into words how much you mean to me, but I love spending life at 6-sigma with you. I will always be grateful to UCSF for bringing you into my life. I love you.

Contributions

The research in the following thesis chapters were performed under the guidance of Dr. Brian Shoichet. Additional guidance and contributions were provided by Dr. Allan Basbaum, Dr. Alexandros Makriyannis, Dr. Kaavya Krishna Kumar, Dr. Francois Pognan, Dr. Marco Vignuzzi, Dr. Francesca Moretti, Dr. Nevan Krogan, Dr. Aashish Manglik, Dr. Matt O' Meara, Dr. Andrew Kruse, and many other mentors and collaborators.

Chapter 1 is written by Tia A. Tummino.

Chapter 2 is reproduced in full with permission from the publication:

Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O' Meara, M. J., Monel, B., ... & Shoichet, B. K. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science*, *373*, 541-547.

Chapter 3 is reproduced in full with permission from the publication:

O'Donnell, H. R., **Tummino, T. A.**, Bardine, C., Craik, C. S., & Shoichet, B. K. (2021). Colloidal aggregators in biochemical SARS-CoV-2 repurposing screens. *Journal of Medicinal Chemistry, 64*, 17530-17539.

Chapter 4 is reproduced in full with permission from the publication:

Alon, A., Lyu, J., Braz, J. M., **Tummino, T. A.**, Craik, V., O'Meara, M. J., ... & Kruse, A. C. (2021). Structures of the σ2 receptor enable docking for bioactive ligand discovery.

Nature, *600*, 759-764.

Chapter 5 is reproduced in part from a manuscript in preparation:

Tummino, T. A., Iliopoulos-Tsoutsouvas, C., Braz, J. M., O'Brien, E. S., Stein, R. M., Craik, V., ... & Shoichet, B. K. Structure-based discovery of cannabinoid-1 receptor agonists with reduced side effects.

Chapter 6 is written by Tia A. Tummino.

"Close your eyes and clone yourself,

build your heart an army,

to defend your innocence,

while you do everything wrong."

– John Mayer, *Age of Worry*

Applications and confounds in drug discovery and repurposing

Tia A. Tummino

Abstract

The process of discovering a new drug is always evolving with the knowledge, technologies, and needs of the time. This information should be used to guide your search and to separate legitimate drug candidates from artifacts and suboptimal leads. In fact, it has been said that a *Drug Hunter*'s job is not to find the best molecule, but to find a reason why every molecule is not the best molecule. The focus of this dissertation is firstly the application of computational drug discovery and repurposing to identify new treatments for diseases. Secondly, it is the mechanistic understanding of two artifacts common in early-stage drug discovery and repurposing that if used appropriately, should remove potential false-positive screening hits from being pursued as lead candidates.

Chapter 1 describes the large-scale docking technology developed in the lab and how it can be used to discover new drugs for protein targets of interest to a particular disease. It further describes the utility of drug repurposing and how it was used during the COVID-19 pandemic to search for novel antivirals. Briefly, it introduces how ligands discovered in drug repurposing screens were ultimately found to be acting through mechanisms that confounded their antiviral activities.

Chapter 2 demonstrates how compounds that induce a phenomenon known as drug-induced phospholipidosis are not legitimate antivirals, and that this effect is a

confound in cell-based antiviral repurposing screens. This shared mechanism underlies the activity of many σ_1 and σ_2 ligands, among others, that were pursued as potential antivirals early in the COVID-19 pandemic. Counter-screening for this activity will help save time, money, and resources from being spent on drugs that have no legitimate promise as antiviral drugs.

Chapter 3 identifies colloidal aggregation as another mechanism by which many compounds show up as false-positive screening hits in biochemical drug repurposing screens. Importantly, we demonstrate that by reducing the formation of colloids in screening assays, we can remove false-positive enzymatic activity of multiple ligands that otherwise appear to be inhibitors of viral proteins.

Chapter 4 demonstrates a legitimate use for σ_2 ligands as potential therapeutics, importantly controlling for both phospholipidosis and aggregation as confounding factors in their activity. We demonstrate with novel selective ligands that σ_2 receptor ligands are antiallodynic in neuropathic pain models, and that their effects are time-dependent, replicating similar phenotypes of other σ_2 ligands from the literature.

Chapter 5 applies the large-scale docking technique on the lipid-binding G-protein coupled cannabinoid-1 (CB1) receptor. Here, we demonstrate the concept of "new chemistry for new biology" by first identifying a novel CB1 agonist and then finding that it has strongly analgesic properties but lacks two of the major cannabinoid side-effects: sedation and catalepsy.

Table of Contents

List of Figures

List of Tables

Chapter 1:Innovations and challenges in drug discovery and repurposing

Innovations and challenges in drug discovery and repurposing

Tia A. Tummino[1,2]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco (UCSF), San Francisco, CA, USA.
[2]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA.

1.1 The Dream of Discovering Drugs

When I entered graduate school, I knew that I wanted to learn how to discover drugs. In particular, I wanted my work to help people who suffered from diseases of the nervous system— like depression, anxiety, PTSD, neurodegeneration, pain, or addiction. Coming from a Neuroscience background, I was mostly familiar with traditional phenotypic drug discovery approaches: you have an animal— probably a rat or a mouse, or in some cases humans experimenting on themselves— you treat it with a compound and see how it changes the behavior or some other readout of activity. Then, you can go back and figure out how the compound works, oftentimes uncovering a new aspect of biology, neurocircuitry, or cellular signaling at the same time.[1] When I entered graduate school, I became much more familiar with more target-based drug discovery approaches, which start with a protein target that is important in disease and you screen libraries of molecules *in vitro* against it to find your "magic bullet"[2,3]. Regardless of which approach[4] I took, I was sure that if I found the right set of tools, I could push this field forward during my time in graduate school.

What I was blissfully unaware of was, firstly, how difficult it is to do good drug

2

discovery, especially as a student at an academic institution. Secondly, I had yet to develop an appreciation for *computational* drug discovery- taking the physical animals and cells out of the equation, instead using a combination of experimental and predicted atomistic models of proteins and ligands to guide our search of chemical space. These themes shaped my ideas and work throughout my PhD and will be present throughout the following chapters of this dissertation.

1.2 Recent Innovations in Drug Discovery

I joined the Shoichet lab during a magical time, as multiple innovations had recently transformed the field of computational drug discovery. Firstly, and particularly for G-protein coupled receptors (GPCRs) which are the target of many CNS drugs, the structural biology revolution made getting high-quality atomic-level information about GPCRs increasingly possible.[5,6] This work was granted the 2012 Nobel Prize in Chemistry[7,8], and by the time I entered the field in 2018, an embarrassment of riches surrounded me. It seemed that every day there was a new important protein structure solved— and once the protein structure was solved— we could use this information to find new drugs.[9] By the time I graduated, this information became even easier to access without even solving experimental structures due to the advent of AlphaFold AI software which was able to predict the folds of proteins where structures had not yet been, or weren't yet able, to be resolved.[10]

Importantly, how this information can be used to find new drugs became an area of interest for me and is what led me to the Shoichet lab. Dr. Brian Shoichet, and many

that came before him including Dr. Tack Kuntz, spent their careers developing physics-based search algorithms[11] (termed "docking"[12,13], from here on out) to computationally approximate the free energy of a ligand binding in a protein cavity binding site without testing every ligand experimentally.[14–16] What set this method (DOCK3.7/3.8) apart from other docking programs (DOCK6, AutoDock, Glide, etc.) is a careful balance of physics-based accuracy with computational speed of the calculations, making it possible to know if a ligand might bind or not in one second or less.

At the time I joined the lab, docking was well-established, though new developments to make the algorithm faster and more physically accurate were always being tested.[17] The major innovation at the time, however, was the development and application of large-scale make-on-demand chemical libraries for virtual screening.[18] The basis for this approach is that chemical space is vast, exceeding numbers of stars in the universe[19], yet most of these molecules have not yet been synthesized and therefore are not included in chemical libraries. Further, molecules that do exist in chemical libraries are often structurally similar to known biogenic molecules, creating a feedback loop of the types of "new" drugs able to be found.[20,21] So, in collaboration with Enamine, a chemical company based in Ukraine, the lab combined their make-on-demand chemical libraries with our virtual screening tools. This work consisted of virtually enumerating hundreds of millions to billions of molecules that could theoretically be made using existing chemical building blocks and simple chemical reactions for direct use in large-scale docking campaigns.[18,22]

The beauty of the technique is that by docking more molecules that are dissimilar to existing known molecules, we can find novel chemotypes that act upon our favorite protein targets in different ways, uncovering novel biological outcomes at the level of the protein, resulting signaling pathways, and sometimes even at the level of behavioral profiles.[23] Previous work in the lab on multiple important drug targets, including the mu opioid receptor[24], the alpha 2A adrenergic receptor[25], and the serotonin transporter[26] exemplify such findings. A similar approach, here looking at the σ_2 and cannabinoid-1 receptors became the focus of my work in **Chapters 4 & 5**.

1.3 Drug Repurposing as an Alternate Approach.

In addition to *de novo* drug discovery, part of my work focused on drug repurposing. Drug repurposing is an approach where you use an existing drug that has already passed FDA scrutiny to treat a disease it wasn't developed to treat[27]. Typically, drug repurposing is used when you don't know much about the underlying biology of a disease, or if there is an urgent need to find a treatment as quickly as possible. This approach became particularly appealing during the COVID-19 pandemic[28] which struck the world in my third year of graduate school. Using maps of human-SARS-CoV-2 protein-protein interactions, our goal was to try and computationally identify FDA-approved drugs that could disrupt interactions between human host proteins that were being hijacked by SARS-CoV-2 during viral infection. Surprisingly, our work identified many drugs that target the σ_1 and σ_2 receptors— proteins typically thought of as being involved in CNS processes and the target of many "dirty" drugs— as being potentially repurposable as antivirals.[29,30] However, the mechanism of how these proteins were involved in SARS-

CoV-2 infection was unclear, and the drugs we identified had no structure-activity relationship to support their antiviral effects coming from engagement of these targets. Work understanding how these, and other, drugs were ultimately confounding drug repurposing projects led to **Chapters 2 & 3** of this dissertation.

1.4 References

1. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507 (2011).

2. Swinney, D. C. Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clin Pharmacol Ther* **93**, 299–301 (2013).

3. Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat. Rev. Cancer* **8**, 473–480 (2008).

4. Al-Ali, H. The evolution of drug discovery: from phenotypes to targets, and back. *MedChemComm* **7**, 788–798 (2016).

5. Blundell, T. L. & Chaplin, A. K. The resolution revolution in X-ray diffraction, Cryo-EM and other Technologies. *Prog. Biophys. Mol. Biol.* **160**, 2–4 (2021).

6. Zhou, X. E., Melcher, K. & Xu, H. E. Structural biology of G protein-coupled receptor signaling complexes. *Protein Sci.* **28**, 487–501 (2019).

7. Clark, R. B. Profile of Brian K. Kobilka and Robert J. Lefkowitz, 2012 Nobel Laureates in Chemistry. *Proc. Natl. Acad. Sci.* **110**, 5274–5275 (2013).

8. Rasmussen, S. G. F. *et al.* Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).

9. Congreve, M., Graaf, C. de, Swain, N. A. & Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **181**, 81–91 (2020).

10. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

11. Pagadala, N. S., Syed, K. & Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **9**, 91–102 (2017).

12. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **7**, 146–157 (2011).

13. Kuntz, I. D., Meng, E. C. & Shoichet, B. K. Structure-Based Molecular Design. *Acc. Chem. Res.* **27**, 117–123 (1994).

14. Shoichet, B. K., Leach, A. R. & Kuntz, I. D. Ligand solvation in molecular docking. *Proteins Struct. Funct. Bioinform.* **34**, 4–16 (1999).

15. Shoichet, B. K. & Kuntz, I. D. Protein docking and complementarity. *J. Mol. Biol.* **221**, 327–346 (1991).

16. Shoichet, B. K., Kuntz, I. D. & Bodian, D. L. Molecular docking using shape descriptors. *J. Comput. Chem.* **13**, 380–397 (1992).

17. Gu, S., Smith, M. S., Yang, Y., Irwin, J. J. & Shoichet, B. K. Ligand Strain Energy in Large Library Docking. *J. Chem. Inf. Model* **61**, 4331–4341 (2021).

18. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

19. Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996).

20. Irwin, J. J. *et al.* ZINC20 A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).

21. Lyu, J., Irwin, J. J. & Shoichet, B. K. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **19**, 712–718 (2023).

22. Tingle, B. I. *et al.* ZINC-22 A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *J. Chem. Inf. Model.* **63**, 1166–1176 (2023).

23. Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **59**, 4103–4120 (2016).

24. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185 (2016).

25. Fink, E. A. *et al.* Structure-based discovery of nonopioid analgesics acting through the α2A-adrenergic receptor. *Science* **377**, eabn7065 (2022).

26. Singh, I. *et al.* Structure-based discovery of conformationally selective inhibitors of the serotonin transporter. *Cell* **186**, 2160-2175.e17 (2023).

27. Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).

28. Edwards, A. What Are the Odds of Finding a COVID-19 Drug from a Lab Repurposing Screen? *J. Chem. Inf. Model* **60**, 5727–5729 (2020).

29. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

30. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020).

Gloss to Chapter 2

If you were to have asked me at the start of grad school if part of my work would focus on antiviral drug repurposing to try and save us from a worldwide pandemic, I would have looked at you like you had two heads. However, when Brian asked for volunteers to help find repurposing candidates that might help us deal with COVID-19, I jumped at the opportunity to adjust my focus to meet the need of the time. What came out of that initial project, however, was very intriguing. Why are antidepressants, antipsychotics, antihistamines, and antimalarials showing antiviral activity? Why would both σ_1 and σ_2, which have little structural similarity and are not genetically related, be involved? How can we separate the activity of each receptor from one another when so little is known about their biological functions and so many of their ligands bind both receptors? The further we dug into these questions, the more confused we got. Finally, after testing nearly 100 ligands, some of which we ourselves had discovered, and no structure-activity relationship emerged, we knew we were in trouble.

Rather than throwing the data in the garbage, we decided to try and understand why *these* ligands in particular were showing up as antiviral screening hits, which led us to the phenomenon of drug-induced phospholipidosis. I will forever be grateful for the lessons that I learned during this project, and especially for having to face that my initial hypothesis was incorrect. This work has made me a much more skeptical scientist, always hunting for ways my data may be misleading me or misrepresenting reality.

Chapter 2:Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2

Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2

Tia A. Tummino[1,2,3,4]†, Veronica V. Rezelj[5]†, Benoit Fischer[6]†, Audrey Fischer[6]†, Matthew J. O'Meara[7], Blandine Monel[8], Thomas Vallet[5], Kris M. White[9,10], Ziyang Zhang[3,4,11,12], Assaf Alon[13], Heiko Schadt[6], Henry R. O'Donnell[1], Jiankun Lyu[1,3,4], Romel Rosales[9,10], Briana L. McGovern[9,10], Raveen Rathnasinghe[9,10,14], Sonia Jangra[9,10], Michael Schotsaert[9,10], Jean-René Galarneau[15], Nevan J. Krogan[3,4,11,16], Laszlo Urban[15], Kevan M. Shokat[3,4,11,12], Andrew C. Kruse[13], Adolfo García-Sastre[9,10,17,18], Olivier Schwartz[8], Francesca Moretti*[6], Marco Vignuzzi[5]*, Francois Pognan[6]*, Brian K. Shoichet[1,3,4]*

†Contributed equally.
* Corresponding authors: bshoichet@gmail.com, francois.pognan@novartis.com, marco.vignuzzi@pasteur.fr, francesca.moretti@novartis.com

[1]Department of Pharmaceutical Chemistry, University of California San Francisco (UCSF), San Francisco, CA, USA.
[2]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA.
[3]Quantitative Biosciences Institute (QBI), UCSF, San Francisco, CA, USA.
[4]QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA.
[5]Institut Pasteur, Viral Populations and Pathogenesis Unit, CNRS UMR 3569, 75724 Paris, Cedex 15, France.
[6]Novartis Institutes for BioMedical Research, Preclinical Safety, Basel, Switzerland.
[7]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA.
[8]Institut Pasteur, Virus and Immunity Unit, CNRS UMR 3569, 75724 Paris, Cedex 15, France.
[9]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[10]Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount

Sinai, New York, NY, USA.

[11]Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, CA, USA.

[12]Howard Hughes Medical Institute, UCSF, San Francisco, CA, USA.

[13]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA.

[14]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[15]Novartis Institutes for BioMedical Research, Preclinical Safety, Cambridge, MA, USA.

[16]Gladstone Institute of Data Science and Biotechnology, J. David Gladstone Institutes, San Francisco, CA 94158, USA.

[17]Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[18]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

2.1 Abstract

Repurposing drugs as treatments for COVID-19, the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has drawn much attention. Beginning with sigma receptor ligands and expanding to other drugs from screening in the field, we became concerned that phospholipidosis was a shared mechanism underlying the antiviral activity of many repurposed drugs. For all of the 23 cationic amphiphilic drugs we tested, including hydroxychloroquine, azithromycin, amiodarone, and four others already in clinical trials, phospholipidosis was monotonically correlated with antiviral efficacy. Conversely, drugs active against the same targets that did not induce phospholipidosis were not antiviral. Phospholipidosis depends on the physicochemical properties of drugs and does not reflect specific target-based activities— rather, it may be considered a toxic confound in early drug discovery. Early detection of phospholipidosis could eliminate these artifacts, enabling a focus on molecules with therapeutic potential.

2.2 Introduction

The outbreak of COVID-19 has inspired multiple drug repurposing screens to find antiviral therapeutics that can be rapidly brought to the clinic[1]. To date, over 1,974 drugs and investigational drugs have reported to have in vitro activity against SARS-CoV-2[1] (**Fig. 2.1**). Since almost all of these act against human targets, and might be unlikely to be viable against a novel virus[2], the question of mechanism of action arises.

Our interest in this question was motivated by the discovery that human sigma receptors were candidate targets for modulating SARS-CoV-2 infection[3], and that drugs and reagents like chloroquine, haloperidol, clemastine, and PB28—all with nanomolar affinity against one or both sigma receptors—had cellular antiviral IC_{50} values in the 300 nM to 5 µM range. Subsequently, we investigated over 50 different molecules with a wide range of affinities at these receptors. While this found molecules with relatively potent antiviral activity, structure activity relationships (SAR) found little correlation between receptor potency and antiviral efficacy in cells (**Fig. 2.S1-2.S3, Supplementary File 1**). Whereas drugs like amiodarone, sertraline, and tamoxifen had mid-to high-nM antiviral IC_{50}s, other sigma-active compounds, such as melperone and DTG, were equipotent on target without measurable antiviral activity. Intriguingly, the antiviral sigma drugs were all cationic at physiological pH and relatively hydrophobic, while those that were inactive against the virus were often smaller and more polar. This cationic-amphiphilic character was shared by many of the hits emerging from other phenotypic screens (**Fig. 2.1, 2.S4**), suggesting it was this physico-chemical property that might explain cellular antiviral activity, instead of a specific on-target activity[4].

If the cationic-amphiphilic nature of these molecules led to antiviral activity in vitro, rather than their individual target-based activities, one would expect this physical property to reflect a shared cellular mechanism. Indeed, cationic amphiphilic drugs (CADs) can provoke phospholipidosis in cells and organs[5]. This side effect is characterized by the formation of vesicle-like structures in susceptible cells and "foamy" or "whorled" membranes[5,6], and is thought to arise by CAD disruption of lipid homeostasis. CADs accumulate in intracellular compartments such as endosomes and lysosomes where they can directly or indirectly inhibit lipid processing[5]. Modulation of these same lipid processing pathways is critical for viral replication[7], and inhibiting phospholipid production has previously been associated with inhibition of coronavirus replication[8]. CADs have in vitro activity against multiple viruses including Severe Acute Respiratory Syndrome, Middle East Respiratory Syndrome, Ebola, Zika, Dengue, and filoviruses[9], though CAD-induction of phospholipidosis has only been proposed as an antiviral mechanism for Marburg virus[10]. Finally, among the drugs that are best-known to induce phospholipidosis are amiodarone[11] and chloroquine[12,13], which are potent inhibitors of SARS-CoV-2 replication in vitro[14–16], while drugs from SARS-CoV-2 phenotypic screens, such as chlorpromazine[17] and tamoxifen[16], can also induce phospholipidosis[18]. As an effect that rarely occurs at concentrations lower than 100 nM, that does not appear to translate from in vitro to in vivo antiviral activity and that can result in dose-limiting toxicity[19], phospholipidosis may act as a confound to true antiviral drug discovery.

Here, we investigate the association between phospholipidosis and antiviral activity against SARS-CoV-2 in cell culture. This apparently general mechanism may be

responsible for many of the drug repurposing hits for SARS-CoV-2, and an extraordinary amount of effort and resources lavished on drug discovery against this disease. We explore the prevalence of this confound in SARS-CoV-2 repurposing studies, how phospholipidosis correlates with inhibition of viral infection, and how to eliminate such hits rapidly so as to focus on drugs with genuine potential against COVID-19, and against new pandemics yet to arise.

2.3 Results

Correlation of phospholipidosis and antiviral activity.

To investigate the role of phospholipidosis in antiviral activity in vitro, we tested 19 drugs for their induction of this effect in A549 cells using the well-established NBD-PE staining assay[20]. Here, the vesicular lipidic bodies characteristic of the effect may be quantified by high content imaging (**Fig. 2.2A**). Three classes of drugs and reagents were initially investigated: **A.** Sigma-binding antiviral CADs we had discovered, like amiodarone, sertraline, chlorpromazine, and clemastine (nine total); these molecules are predicted or known to induce phospholipidosis; **B.** Analogs of these CADs that no longer bound sigma receptors, but were still antiviral (four total); these molecules are predicted to induce phospholipidosis despite their lack of sigma binding; and **C.** Sigma-binding, ***non***-antiviral drugs, like melperone and DTG, that were much more polar than classic CADs (two total); these molecules are predicted not to induce phospholipidosis. Of the nine sigma-binding CADs that were antiviral (class **A**), six of which were also found in phenotypic screens from the literature as inhibitors of COVID-19, eight induced phospholipidosis, consistent with the hypothesis (**Fig. 2.2A-B, 2.S5-2.S6**). The only non-

phospholipidosis inducing antiviral from this set was elacridar, a promiscuous P-glycoprotein inhibitor; this investigational drug may therefore be active via another mechanism. Intriguingly, analogs of the potent sigma ligand PB28 that had lost their sigma-binding activity but remained CADs (ZZY-10-051 and ZZY-10-061, **Fig. 2.2B-F, 2.S5-2.S8**), did induce phospholipidosis, as did the antipsychotic olanzapine and the antihistamine diphenhydramine, which are weak sigma receptor ligands but are structurally related to potent sigma receptor tricyclics (e.g., chlorpromazine) and diarylethanolamines (e.g., clemastine; class **B**). Finally, melperone and DTG, which are potent cationic sigma receptor ligands but are not antiviral, did not induce phospholipidosis (**Fig. 2.2A-B, 2.S5-2.S6**; class **C**). These results do not prove phospholipidosis as the antiviral mechanism but are consistent with the phospholipidosis hypothesis.

If phospholipidosis is responsible for antiviral activity, then molecules known to induce phospholipidosis should also be antiviral. We tested three CADs for antiviral activity, including ebastine, ellipticine, and Bix 01294, all of which are reported to induce phospholipidosis[21] (Bix 01294 and ebastine have also been reported as drug repurposing hits against SARS-CoV-2[22]). We further tested azithromycin, also reported to induce phospholipidosis[23], but having different physical properties from typical CADs. We first confirmed phospholipidosis-inducing activity for these molecules, though it is difficult to separate cytotoxicity from phospholipidosis and antiviral activity for both ellipticine and ebastine (**Fig. 2.2B, 2.S5-2.S6**). All four molecules were next shown to be antiviral, here and elsewhere with live virus assays (e.g., SARS-CoV-2 strain

BetaCoV/France/IDF0372/2020; **Methods**), with IC_{50} values in the 400 nM to 3 μM range, overlapping with the activities of other CADs we and others have identified for SARS-CoV-2[22] (**Fig. 2.S6**). This too was consistent with the antiviral phospholipidosis hypothesis.

For phospholipidosis to explain antiviral activity, we might expect a correlation between concentration-response curves for phospholipidosis and for antiviral activity. We compared concentrations that induce phospholipidosis to those that inhibit SARS-CoV-2 for each drug individually. Typically, the correlations were high—not only did antiviral activity occur in the same concentration ranges as phospholipidosis, but the statistically significant R^2 values, ranging from 0.51 to 0.94, supported a quantitative relationship between the two effects (**Fig. 2.3A**). We then fit a sigmoidal model through all the 107 phospholipidosis versus antiviral activity observations (comprised of six concentration measurements each for 16 phospholipidosis-inducing drugs) and observed a strong negative correlation ($R^2 = 0.65$, 95%CI [.52, 0.76]) between induced phospholipidosis and SARS-CoV-2 viral load across all observations for all 16 drugs. Because phospholipidosis and antiviral effects are both saturable, the sigmoidal curve-fit plateaus at the extremes (**Fig. 2.3B**).

Concurrent measurement of viral infection and drug induced phospholipidosis.

In the previous experiments, drug-induced phospholipidosis and drug antiviral activity were measured separately. To measure the two effects in the same cells at the

same time, we dosed cells with either 1 or 10 µM of five characteristic CADs (amiodarone, sertraline, PB28, hydroxychloroquine (HCQ), and Bix 01294), followed by a mock or SARS-CoV-2 infection, and quantified phospholipidosis and the accumulation of viral spike protein (**Fig. 2.4A, 2.S9**). Compared to DMSO, drug treatments led to substantial increases in NBD-PE aggregates, indicating increased phospholipidosis (**Fig. 2.S9**). At 1 µM drug concentrations, SARS-CoV-2 spike protein was readily stained, and one could visualize both spike protein and phospholipidosis in the same cells (yellow puncta), suggesting at this low concentration of drug—often close to the antiviral IC_{50} value—both phospholipidosis and viral infection co-occur, though even here viral staining was reduced relative to the DMSO treated controls. As drug concentration rose to 10 µM, viral spike protein staining dropped while staining for phospholipidosis increased (**Fig. 2.S9**); there was nearly complete loss of spike protein signal with a concomitant increase in phospholipidosis (**Fig. 2.4A**) for all treatments. In seven-point concentration-response curves for amiodarone, sertraline, and PB28, viral staining monotonically decreased as phospholipidosis increased (**Fig. 2.4B-C**).

CADs are common among drug repurposing hits for SARS-CoV-2 and other viruses.

With the strong correlation between CAD phospholipidosis and antiviral efficacy (**Fig. 2.3**), including drugs that have been found in multiple SARS-CoV-2 repurposing studies, we investigated the prevalence of phospholipidosis-inducing CADs among 1,974 total reported repurposing hits identified in the literature. We focused on 12 drug repurposing efforts for SARS-CoV-2, including two screens of the ReFRAME library[24,25],

screens of the NCATS "approved drug" and "bioactive" libraries[15], among others[3,14,16,22,26–30]. Together, these 12 screens found 310 drugs, investigational drugs, and reagents that were antiviral in vitro against SARS-CoV-2. We used two physico-chemical features to identify likely CADs: drugs with calculated Log octanol:water coefficients above 3 (cLogP ≥ 3), and with pKa values ≥ 7.4[31,32]. We then further filtered for drugs that topologically resembled known phospholipidosis inducers[18,21] using an ECFP4-based Tanimoto coefficient ((Tc) ≥ 0.4) (**Supplementary File 2**). Sixty percent of the 310 drugs passed the cLogP and pKa threshold; 34% also resembled a known phospholipidosis inducer (**Fig. 2.1, 2.S4**, **2.S10**).

Although the two physical property filters do not capture atypical phospholipidosis inducers such as azithromycin, they do capture 16 of the other 18 CADs we had already tested (missing only the medium phospholipidosis inducers olanzapine and ellipticine); intriguingly, nine of these, including amiodarone, sertraline, chlorpromazine, Bix 01294, clemastine, and benztropine also appeared in at least one of the 12 other repurposing studies. To probe the reliability of this association, we tested another five drugs that passed our filters, and had been reported as antiviral against SARS-CoV-2, for their induction of phospholipidosis**.** Not only were all five were active in the NBD-PE assay, but we were able to confirm SARS-CoV-2 antiviral activity for these drugs (**Fig. 2.S10**). Additionally, these molecules fit into the sigmoidal model relating the extent of phospholipidosis to reduction in viral load (salmon points overlaid with sigmoidal model; **Fig. 2.3B**). Finally, we note a preliminary identification of 30 CADs, 19 of which overlap with the literature-derived SARS-CoV-2 list, active against other viruses including Middle

East Respiratory Syndrome and Severe Acute Respiratory Syndrome[33], Ebola[34–36], Marburg[36,37], Hepatitis C[38] (*38*), and Dengue[39] (**Table 2.S1**). It may be that most drugs repurposed against many viruses are CADs whose antiviral activities can be attributed to a phospholipidosis mechanism.

Animal efficacy for repurposed drugs.

Though phospholipidosis is considered a drug-induced side effect, it remains possible that it can be leveraged for antiviral efficacy. Accordingly, we tested four of the repurposed, phospholipidosis-inducing drugs most potent against SARS-CoV-2 in vitro, amiodarone, sertraline, PB28 and tamoxifen[5,18], for efficacy in a murine model of COVID-19[40]. In the same model, we also tested elacridar, which does not induce phospholipidosis (**Fig. 2.2B**) and remdesivir, which is unlikely to induce phospholipidosis at concentrations relevant to its antiviral activity. In pharmacokinetic studies, all molecules had relatively long half-lives, especially in the lung where tissue C_{max} values often exceeded 10 µM after a 10 mg/kg dose, or 10 to 1000 times higher than their in vitro antiviral IC_{50}s, suggesting that exposure would be high enough for plausible efficacy (**Table 2.S2-2.S6**). Guided by the pharmacokinetics of each drug, mice were dosed either once (amiodarone and elacridar) or twice per day (remdesivir, PB28, tamoxifen, and sertraline), for three days. Two hours following the first dose, mice were intranasally infected with 1×10^4 PFU of SARS-CoV-2 and lung viral titers were measured after a three-day infection period. Notwithstanding their high lung exposure, the four phospholipidosis-inducing drugs had no substantial effect on viral propagation in the mice. Conversely, remdesivir reduced viral load by two to three orders of magnitude. While the cationic non-phospholipidosis

drug elacridar had a modest antiviral effect, it did not rise to statistical significance (**Fig. 2.5**) and mice given elacridar doses higher than 3 mg/kg exhibited toxicities that limited further study.

Because phospholipidosis is typically an in vivo side effect that appears after chronic dosing, we then pre-treated mice with five-fold higher concentrations (50 mg/kg) of amiodarone over twelve days prior to a 3-day infection period. Even here, no diminution of viral titer was observed in mouse lungs after infection, and amiodarone offered no protection from infection-induced weight loss or from pulmonary inflammation and epithelial necrosis, as measured by histopathology scores (**Fig. 2.5, 2.S11**). We noted that foamy vacuolation and whorled vacuoles that are the hallmarks of phospholipidosis were not seen in lung and spleen by light or transmission electron microscopy. It is thus possible that this treatment was not long enough to induce a protective phospholipidosis phenomenon. Still, taken together, the *in vitro* activities of the phospholipidosis-inducing drugs did not translate *in vivo*, and drugs whose antiviral activity arises due to phospholipidosis seem non-viable for clinical progression.

2.4 Discussion

The emergence of COVID-19 has motivated intense effort to repurpose drugs as SARS-CoV-2 antivirals. An extraordinary number of diverse, apparently unrelated hits have emerged[1]. A key observation from this work is that many, perhaps most of these are active in antiviral assays via induction of phospholipidosis (**Fig. 2.1, 2.S4, 2.S10**). This disrupts lysosomal lipid catabolism and trafficking, which may in turn disrupt the double

membrane vesicles that the virus creates and on which it depends for propagation. Quantitatively, there is a close in vitro correlation between drug-induced phospholipidosis and antiviral activity, both drug-by-drug and over the set of drugs tested here (**Fig. 2.3**). The effect is predictive: molecules that induce phospholipidosis are antiviral over the same concentration range, irrespective of whether they are cationic amphiphilic drugs (CADs) or not (e.g., azithromycin), while molecules that are related by target activity to the CADs, but are more polar and do not induce phospholipidosis (e.g., melperone and DTG), are not antiviral. Unfortunately, CAD induction of phospholipidosis, at least at the potencies observed here, does not appear to translate in vivo (**Fig. 2.5**). More encouragingly, this study illuminates a method to rapidly identify confounds in cellular antiviral screens, allowing one to eliminate them from further study and to focus on those molecules with true potential.

Although the molecular mechanisms for the antiviral effects of phospholipidosis remain unclear, certain associations may be tentatively advanced. SARS-CoV-2, like many viruses, subverts the cell to produce double membrane vesicles in which it replicates[41–43]. Disruption of lipid homeostasis by the induction of phospholipidosis may disrupt these vesicles, reducing viral replication. The disruption of lysosomal[44] and endosomal[45] compartments and CAD-induced shifts in compartmental pH[46] may further affect viral entry and propagation[47]. For these reasons, targeting the endosomal-lysosomal pathway has been suggested as a viable strategy against SARS-CoV-2 infection[48], but developing potent and targeted inhibitors remains challenging. Of course, these mechanisms remain unproven, and currently are supported mostly by correlation,

but they suggest a route for further research.

The cost to the community of investments in what appears to be a confound merits consideration for future pandemics. According to the DrugBank[49] COVID-19 dashboard, which draws from U.S. and international clinical trials, putatively antiviral CADs have been promoted into an astonishing 316 Phase I to Phase III clinical trials against COVID-19. While 57% of these study the phospholipidosis-inducing CADs hydroxychloroquine (**Fig. 2.3A**, top row) or chloroquine, that still leaves 136 trials across 33 other predicted or known phospholipidosis-inducers. Using conservative estimates[50,51], the expense of the clinical trials component alone, over the last year, for phospholipidosis-inducing CADs may be over $6 billion US dollars (**Table 2.S7**).

Certain caveats merit airing. First, the correlation between antiviral activity and phospholipidosis, as strong as it is, does not illuminate the mechanism by which phospholipidosis is antiviral. Phospholipidosis is itself only partly understood, and there are no good genetic or chemical ways to either inhibit its induction by drugs nor to promote it by target-selective reagents. Second, predicting whether a molecule will induce phospholipidosis remains challenging, and even non-CAD molecules can induce it. Thus, we have chosen conservative criteria to predict phospholipidosis-inducers, which may miss many drugs. Third, phospholipidosis is a confound that only affects drugs repurposed for direct antiviral activity—it is irrelevant for drugs like dexamethasone[52] and fluvoxamine[53] that have been repurposed for immunomodulation in COVID-19, nor is it relevant for CADs whose antiviral activity is well-below the concentration range where

phospholipidosis occurs. Fourth, our estimates of the clinical trial costs of phospholipidosis-inducing CADs are obviously rough. Finally, we do not exclude exploiting phospholipidosis therapeutically, though we suspect that would have to go through a more target-directed mechanism than that of the CADs studied here.

These caveats should not obscure the central observation of this study. Many drugs repurposed for antiviral activity against SARS-CoV-2 are cationic amphiphiles, and despite their diverse structures and multiple targets, many likely have their antiviral effects via a single shared mechanism: phospholipidosis. Both because of the side effects with which it is associated, and the limited efficacy to which it leads—rarely better than 100 nM in vitro—drugs active due to phospholipidosis are unlikely to translate in vivo (**Fig. 2.5**). Many resources will be saved by counter-screening for phospholipidosis in even simple cellular assays[20], allowing investigators to focus on drugs with genuine promise as antivirals.

2.5 Acknowledgements

Author Contributions: The following authors designed and conceptualized the study: T.A.T, V.V.R, M.J.O., Z.Z., K.M.S., F.M., M.V., F.P., & B.K.S. The following authors performed experiments or data acquisition: T.A.T, V.V.R., B.F., A.F., B.M., T.V., K.M.W., Z.Z., A.A., H.R.O., R.R., B.L.M., R.R., S.J., M.S., & J.-R.G. The following authors conducted formal data analysis: T.A.T, V.V.R, B.F., A.F., M.J.O., B.M., T.V., K.M.W., Z.Z., A.A., H.S., J.L., & F.M. The following authors supervised or managed research: M.J.O., K.M.W., N.J.K., L.U., K.M.S., A.C.K., A.G.-S., O.S., F.M., M.V., F.P., & B.K.S. The following authors drafted the original manuscript: T.A.T., V.V.R., M.J.O, B.M., K.M.W., Z.Z., M.V., F.P., & B.K.S. All authors reviewed the manuscript prior to submission. **Competing interests:** A.C.K. is a founder and advisor to Tectonic Therapeutic, Inc., and the Institute for Protein Innovation. K.M.S. has consulting agreements for the following companies, which involve monetary and/or stock compensation: Black Diamond Therapeutics, BridGene Biosciences, Denali Therapeutics, Dice Molecules, eFFECTOR Therapeutics, Erasca, Genentech/Roche, Janssen Pharmaceuticals, Kumquat Biosciences, Kura Oncology, Merck, Mitokinin, Petra Pharma, Revolution Medicines, Type6 Therapeutics, Venthera, Wellspring Biosciences. The N.J.K. laboratory has received research support from Vir Biotechnology and F.

2.6 Figures

Figure 2.1. Representative examples of cationic amphiphilic drugs that are identified in SARS-CoV-2 drug repurposing screens.
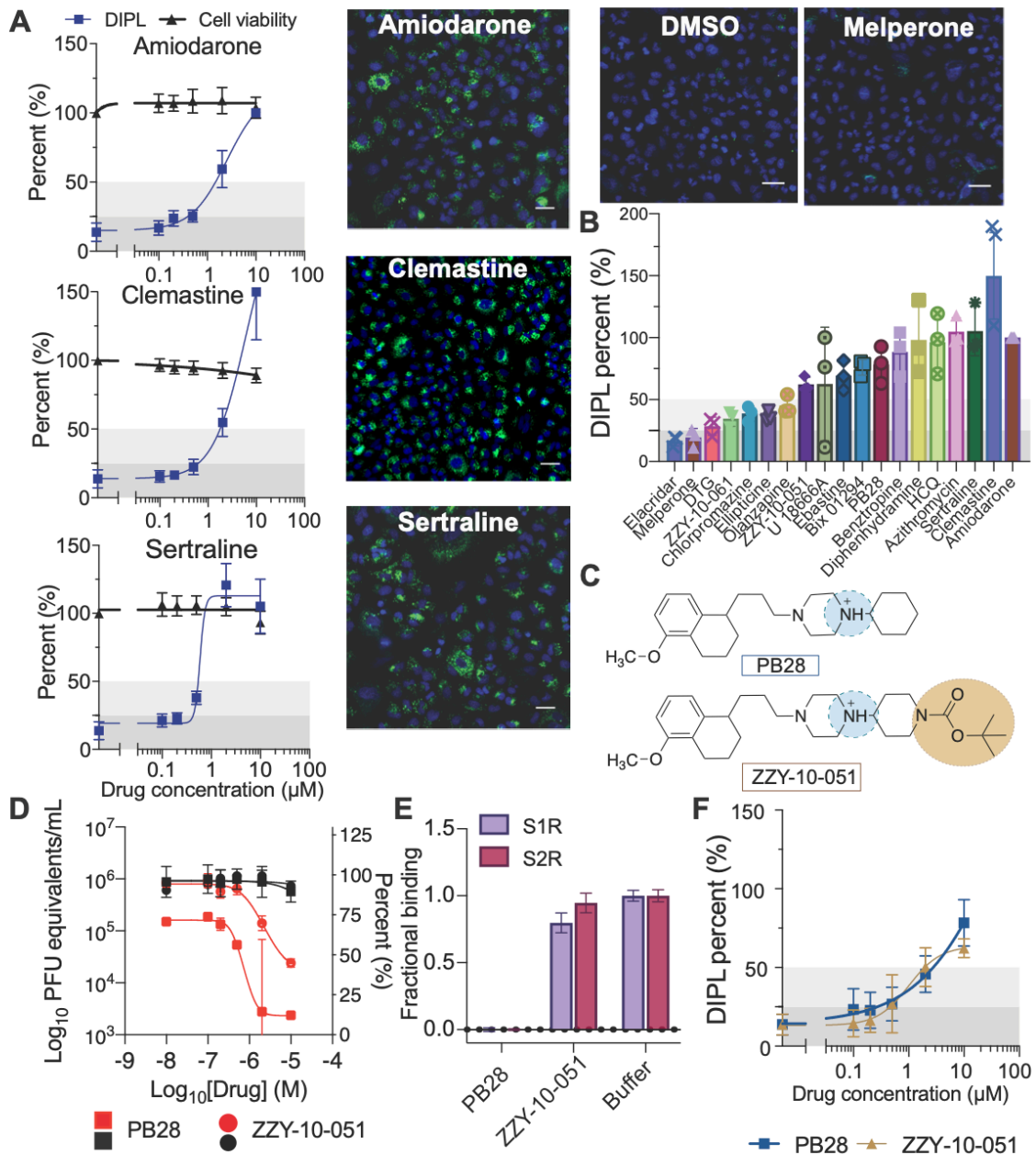
Figure 2.2. Cellular phospholipidosis may confound antiviral screening results.
A. Examples of NBD-PE quantification of phospholipidosis in A549 cells including dose response curves. Blue = Hoechst nuclei staining, Green = NBD-PE phospholipid staining, Red = EthD-2 staining for dead cells. Scale bars = 20 μm. Amiodarone is the positive control for assay normalization; sertraline and clemastine are two examples of high phospholipidosis inducing drugs (phospholipidosis (DIPL) > 50% of amiodarone). Images of DMSO and a non-phospholipidosis inducing molecule (melperone) are included for reference. (Continued on the next page.)

24

(Continued from the previous page.) Thresholds for determining phospholipidosis power are shaded in dark grey (low phospholipidosis), light gray (medium phospholipidosis) and no shading (high phospholipidosis). **B.** Pooled DIPL amounts (mean ± SD) at the highest non-toxic concentration tested for each drug. Results were pooled from three biological and three technical replicates and were normalized to amiodarone (100%) from the control wells in the same experimental batches. **C.** Structures of PB28 and its analog ZZY-10-051, the latter of which is inactive on the sigma receptors. **D**. Viral infectivity (red) and viability (black) data for PB28 (square) and ZZY-10-051 (circle) in A549-ACE2 cells. Data shown are mean ± SD from three technical replicates. **E.** Fractional binding of PB28 and ZZY-10-051 against Sigma-1 (purple; S1R) and Sigma-2 (maroon; S2R) normalized to a buffer control at 1.0 in a radioligand binding experiment. Data shown are mean ± SEM from three technical replicates. PB28 is a strong ligand of both Sigma-1 and Sigma-2 and has high displacement of the radioligands, whereas ZZY-10-051 is unable to displace the radioligands to a high degree at 1 µM. **F.** Dose response curves for PB28 (blue) and ZZY-10-051 (gold) show that these closely related analogs both induce phospholipidosis.

A

Amiodarone
$R^2 = 0.73$
P value $= 0.03$

Clemastine
$R^2 = 0.66$
P value $= 0.049$

Sertraline
$R^2 = 0.67$
P value $= 0.047$

Azithromycin
$R^2 = 0.72$
P value $= 0.03$

Hydroxychloroquine
$R^2 = 0.91$
P value $= 0.003$

Diphenhydramine
$R^2 = 0.78$
P value $= 0.02$

Benztropine
$R^2 = 0.82$
P value $= 0.01$

PB28
$R^2 = 0.65$
P value $= 0.05$

Bix 01294
$R^2 = 0.94$
P value $= 0.006$

Ebastine
$R^2 = 0.62$
P value $= 0.06$

U 18666A
$R^2 = 0.94$
P value $= 0.001$

ZZY-10-051
$R^2 = 0.76$
P value $= 0.03$

Olanzapine
$R^2 = 0.83$
P value $= 0.01$

Ellipticine
$R^2 = 0.51$
P value $= 0.10$

Chlorpromazine
$R^2 = 0.95$
P value $= 0.001$

ZZY-10-061
$R^2 = 0.57$
P value $= 0.08$

DTG
$R^2 = 0.09$
P value $= 0.57$

Melperone
$R^2 = 0.1$
P value $= 0.54$

Elacridar
$R^2 = 0.15$
P value $= 0.45$

SARS-CoV-2 percent (%)

DIPL percent (%)

B

$R^2 = 0.65 [0.52, 0.76]$

SARS-CoV-2 percent (%)

DIPL percent (%)

Amiodarone
Azithromycin
Benztropine
Bix 01294
Chlorpromazine
Clemastine
Diphenhydramine
Ebastine
Duloxetine
Fluspirilene
Methdilazine
Thiethylperazine
Toremifene

Ellipticine
HCQ
Olanzapine
PB28
Sertraline
U 18666A
ZZY-10-051
ZZY-10-061

26

Figure 2.3. Quantitative relationship between phospholipidosis and viral amounts.
A. Correlations between phospholipidosis (DIPL), normalized to amiodarone at 100%, and percent of SARS-CoV-2, normalized to DMSO at 100%, in the RT-qPCR assay in A549-ACE2 cells. Each dot represents the same concentration tested in both assays. A strong negative correlation emerges, with $R^2 \geq 0.65$ and $p \leq 0.05$ for all high and medium phospholipidosis-inducing drugs except ellipticine, which is confounded by its cytotoxicity in both experiments, ebastine, and ZZY-10-61. The latter two examples are marginally significant. **B.** The SARS-CoV-2 viral loads and induced phospholipidosis magnitude for each compound and dose in **A** are plotted as sqrt(viral_amount_mean) ~ 10*inv_logit(hill*4/10*(log(DIPL_mean)-logIC$_{50}$). Fitting a sigmoid Bayesian model with weakly informative priors yields parameters and 95% credible intervals of IC$_{50}$: 43 [38, 48]%, hill: -5.6 [-7.0, -4.5], and Sigma 2.0 [0.14, 1.78]. Forty draws from the fit model are shown as blue lines. Salmon points overlaid with the model represent predicted phospholipidosis inducers from the literature (**Fig. 2.S10**).
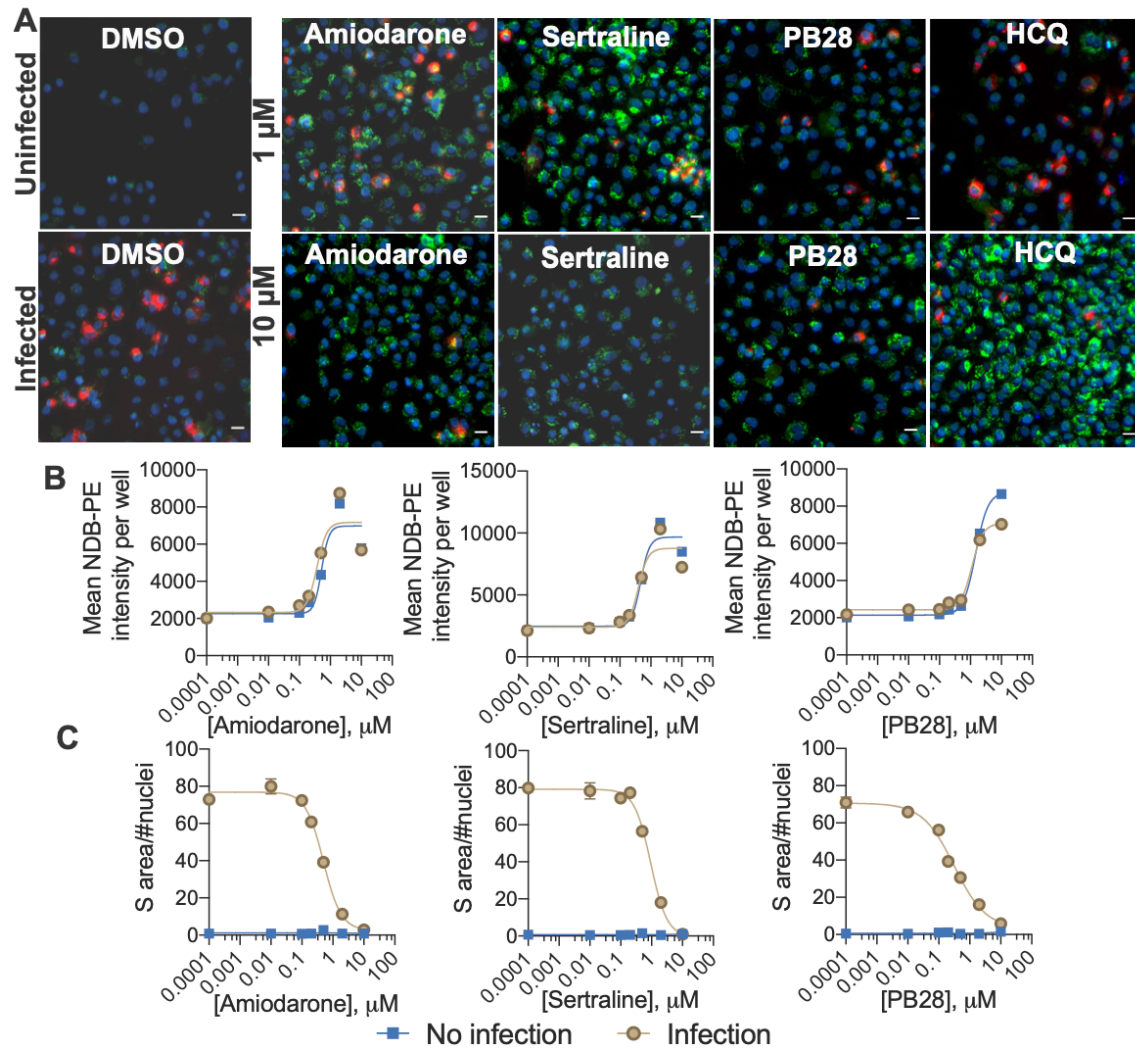
Figure 2.4. Phospholipidosis and spike protein measurements in the same cellular context.
A. Representative images from a co-staining experiment measuring phospholipidosis and SARS-CoV-2 spike protein in infected and uninfected A549-ACE2 cells. Five molecules (1 and 10 μM) and DMSO were measured; see **Fig. 2.S9** for Bix 01294. Blue = Hoechst nuclei staining, Green = NBD-PE phospholipid staining, Red = SARS-CoV-2 spike protein staining; Yellow = coexpression of spike protein and NBD-PE. Scale bar = 20 μm. **B.** Concentration-response curves for phospholipidosis induction measured by NBD-PE staining in infected cells for three characteristic CADs. **C.** Spike protein in infected cells decreases as phospholipidosis increases. For **B.** and **C.**, data are mean ± SEM from four biological replicates.
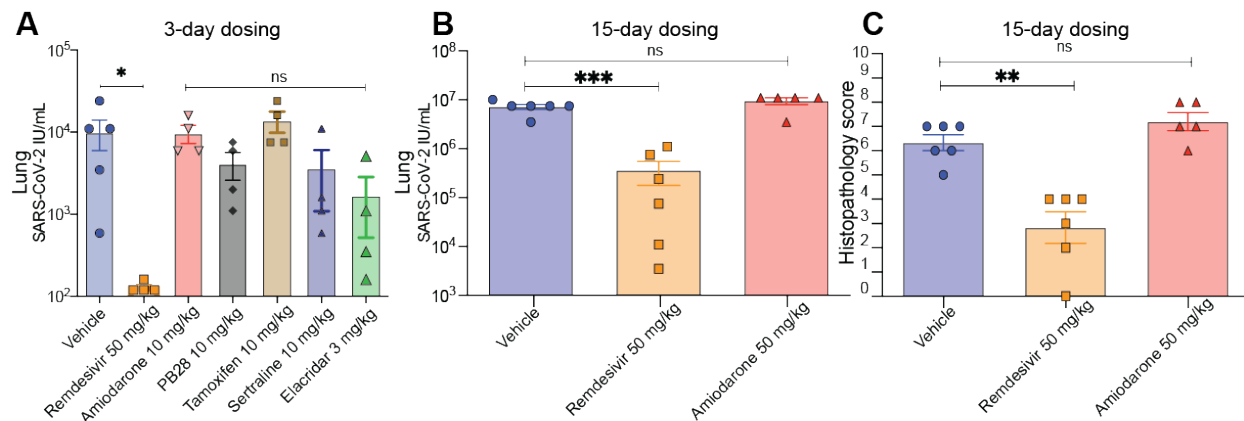
Figure 2.5. Phospholipidosis-inducing drugs are not efficacious in vivo.
A. Three-day dosing of six different drugs with a two-hour preincubation before SARS-CoV-2 treatment. Lung viral titers were quantified and groups were compared using the Kruskal-Wallis test ($H(7)$ = 22.76, P = 0.002) with Dunn's multiple comparison correction indicated (vehicle N = 5; remdesivir N = 4, *P = 0.02). All other groups N = 4, ns = not significant. **B.** Fifteen-day dosing of amiodarone (50 mg/kg) compared to 3-day remdesevir dosing. Lung viral titers were quantified and groups were compared with a two-way ANOVA (main effect of treatment $F(2,9)$ = 19.66, P = 0.0005; no main effect of mouse, $F(5,9)$ = 1.21, P = 0.38). Individual group comparisons determined using Dunnett's multiple comparison test are indicated (vehicle N = 6; remdesivir N = 6, ***P = 0.0008, amiodarone N = 5, ns = not significant). **C.** Histopathology scores after 15-day (amiodarone) or 3-day (remdesivir) treatments as in panel **B**. See **Materials and Methods** for scoring breakdown. Groups were compared with a two-way ANOVA (main effect of treatment $F(2,9)$ = 19.05, P = 0.0006; no main effect of mouse, $F(5,9)$ = 0.78, P = 0.59). Individual group comparisons determined using Dunnett's multiple comparison test are indicated (vehicle N = 6; remdesivir N = 6, **P = 0.0014, amiodarone N = 5, ns = not significant). All data are mean ± SEM.
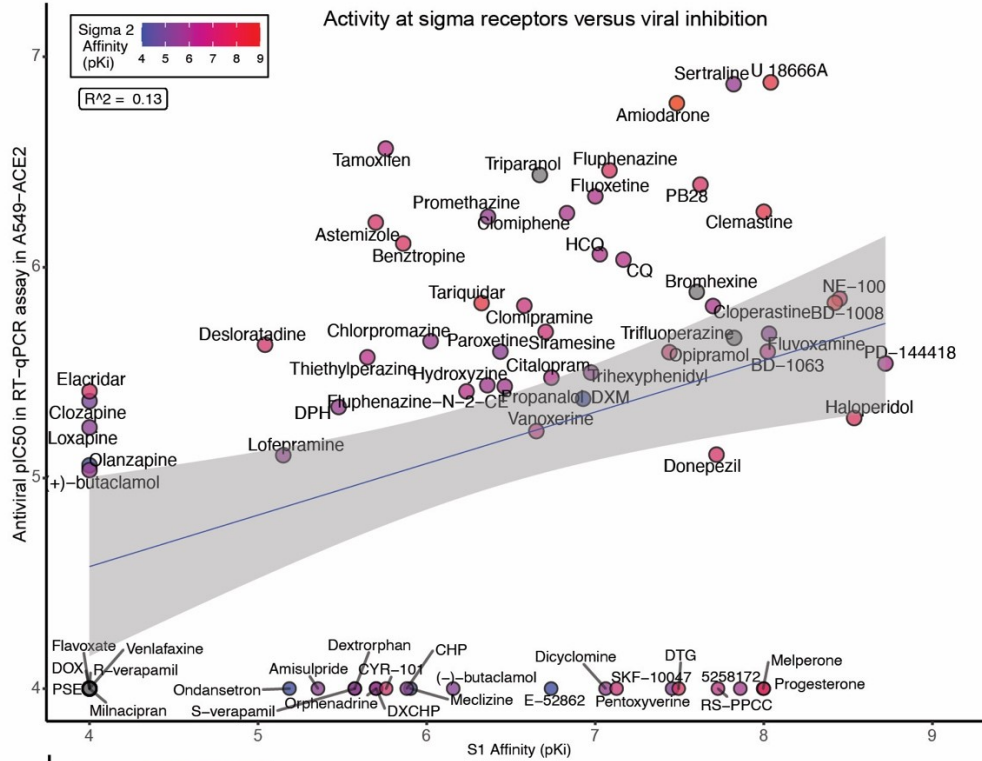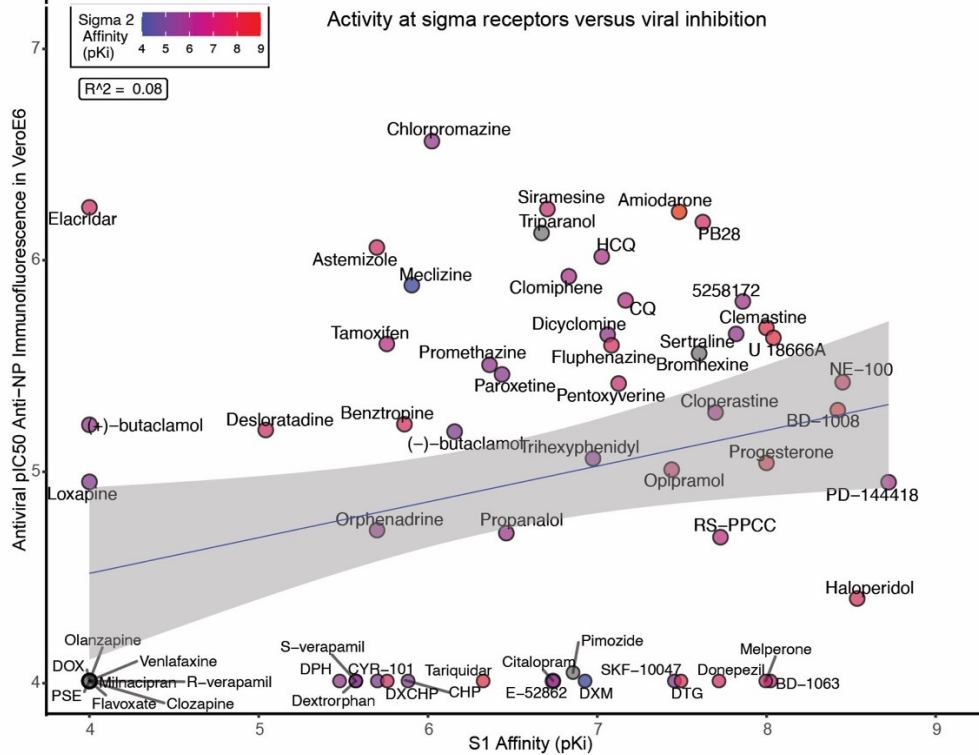
Figure 2.S1. Correlation analyses for sigma receptor affinity and antiviral activity.
A. pKi at Sigma-1 was correlated with pIC50 in the RT-qPCR assay in A549-ACE2 cells.
pKi at Sigma-2 is denoted by the colors. (Continued on the next page.)

(Continued from previous page.) **B**. pKi at Sigma-1 was correlated with pIC50 in the anti-NP immunofluorescence viral infectivity assay in VeroE6 cells. Abbreviations: CQ: chloroquine; HCQ: hydroxychloroquine; DOX: doxylamine; PSE: pseudephedrine; CHP: chlorpheniramine; DXCHP: dexchlorpheniramine; DXM: dextromethorphan; DPH: diphenhydramine; DTG: ditolylguanidine.

A

SARS-CoV-2 RT-qPCR   Cell Viability %

SARS-CoV-2 RT-qPCR   Cell Viability %
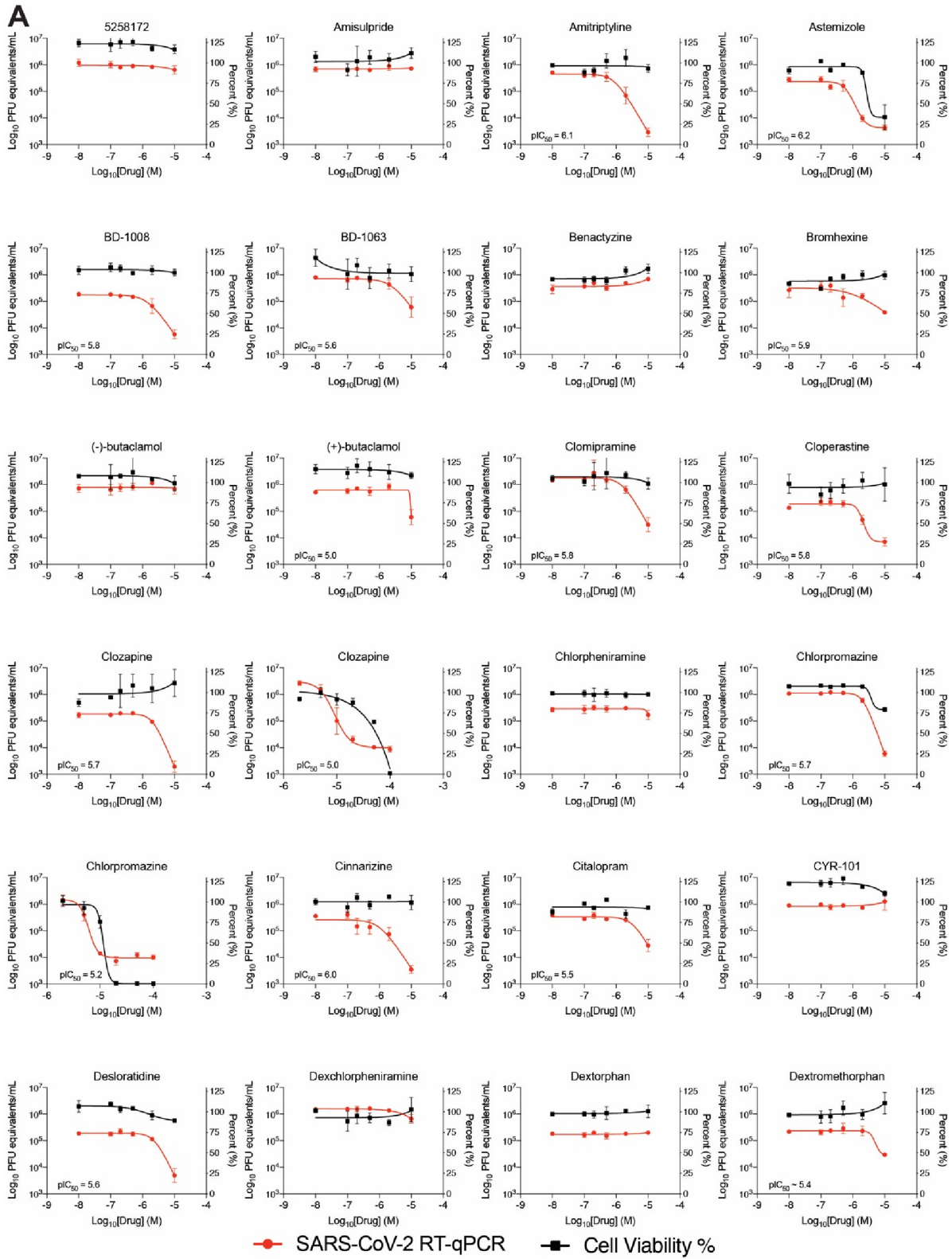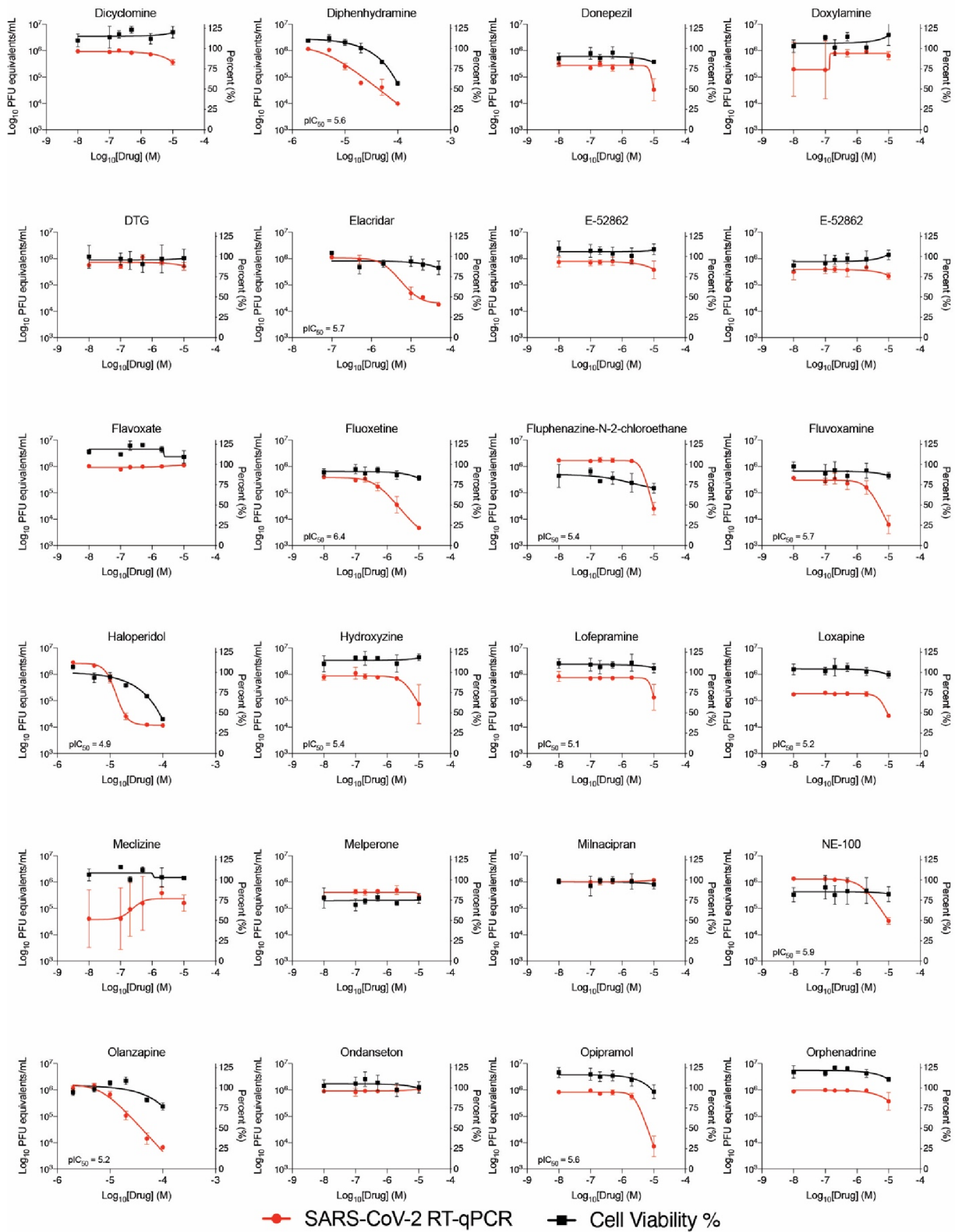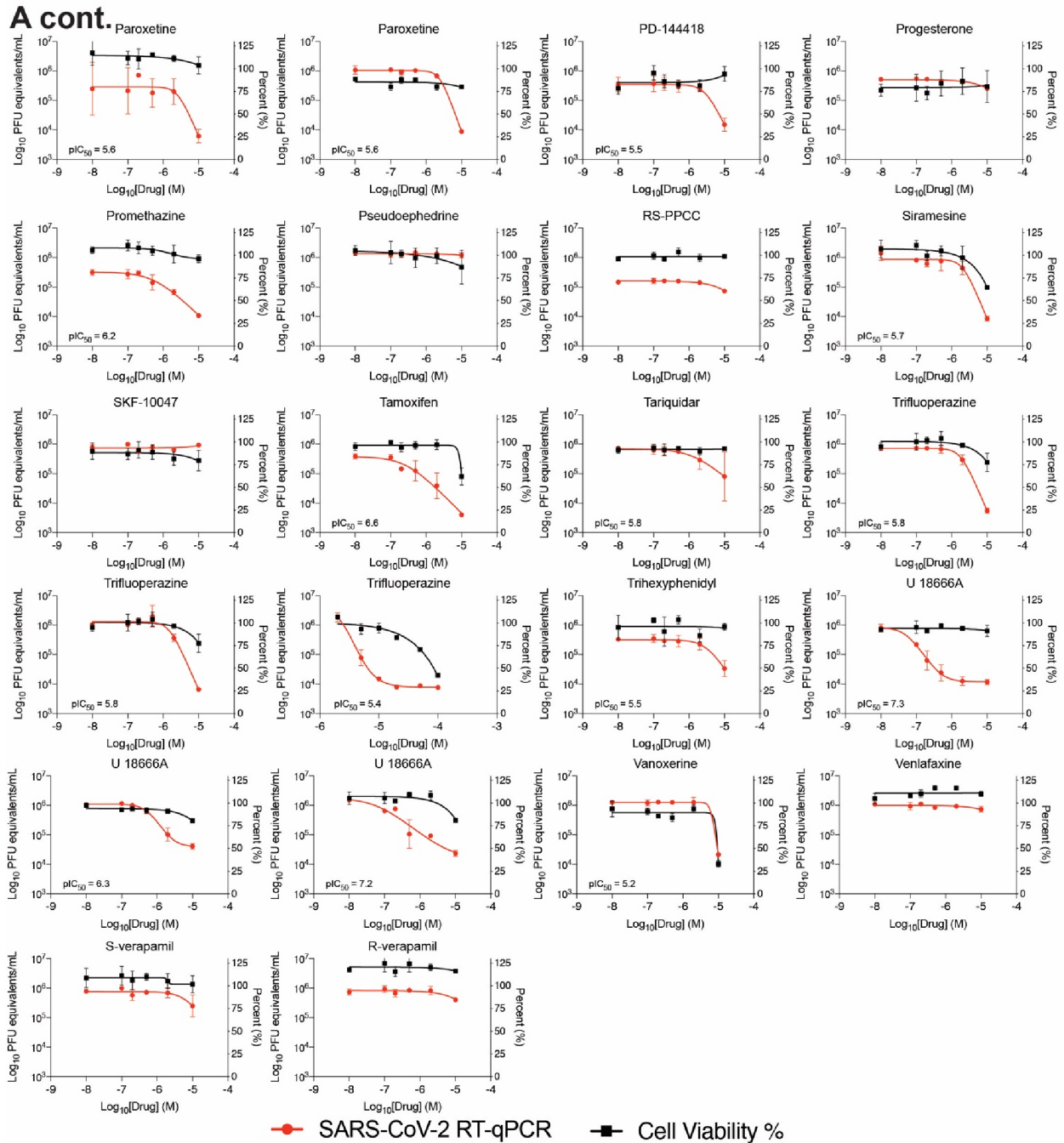
Figure 2.S2. Dose response curves for a set of cationic amphiphilic drugs in an RT-qPCR viral infectivity assay.
A. Viral infectivity and cell viability data for a subset of literature-identified cationic amphiphilic drugs in VeroE6 cells. Data shown are mean ± SD from three biological replicates. Independent experiments are shown as separate graphs when available.

A

5258172 — pIC$_{50}$ = 4.8
Amiodarone — pIC$_{50}$ = 6.4
Amiodarone — pIC$_{50}$ = 6.1
Amiodarone — pIC$_{50}$ = 6.2

Amitriptyline — pIC$_{50}$ = 5.6
Astemizole — pIC$_{50}$ = 6.3
Astemizole — pIC$_{50}$ = 5.7
Astemizole — pIC$_{50}$ = 6.2

BD-1063
BD-1008 — pIC$_{50}$ = 5.3
BD-1008 — pIC$_{50}$ = 5.5
Benactyzine

Benztropine — pIC$_{50}$ = 5.3
Benztropine — pIC$_{50}$ = 5.1
Bromhexine — pIC$_{50}$ = 5.6
+/- Butaclamol — pIC$_{50}$ = 5.2

-/- Butaclamol — pIC$_{50}$ = 5.2
Chloroquine — pIC$_{50}$ = 5.8
Chlorpheniramine
Chlorpromazine — pIC$_{50}$ = 6.5

Chlorpromazine
Cinnarizine — pIC$_{50}$ = 6.2
Citalopram
Clomiphene — pIC$_{50}$ = 6.2

● SARS-CoV-2 anti-NP %   ■ Cell Viability %

All plots: Y-axis = Percent (%), X-axis = Log$_{10}$[Drug] (M)

35

A cont.

Figure showing dose-response curves for various drugs plotted as Percent (%) versus Log₁₀[Drug] (M). Panels: Clomiphene (pIC₅₀ = 6.4), Clozapine (pIC₅₀ = 4.8), CYR-101, Desloratadine (pIC₅₀ = 5.2), Dexchlorpheniramine, Dextromethorphan, Dextrorphan, Dicyclomine (pIC₅₀ = 5.8), Dicyclomine (pIC₅₀ = 5.5), Diphenhydramine, Donepezil, Doxylamine, DTG, Elacridar (pIC₅₀ = 6.2), Elacridar (pIC₅₀ = 6.2), Elacridar (pIC₅₀ = 6.4), Elacridar (pIC₅₀ = 6.2), Flavoxate, Fluphenazine (pIC₅₀ = 5.7), Fluphenazine (pIC₅₀ = 5.5), Hydroxychloroquine (pIC₅₀ = 5.9), Loxapine (pIC₅₀ = 4.9), Loxapine, Meclizine (pIC₅₀ = 5.9).

Legend: ● SARS-CoV-2 anti-NP %   ■ Cell Viability %

A cont.

SARS-CoV-2 anti-NP %   Cell Viability %

A cont.

Figure 2.S3. Dose response curves for a set of cationic amphiphilic drugs (CADs) in an anti-NP immunofluorescence viral infectivity assay.
A.Viral infectivity and cell viability data for a subset of literature-identified CADs in VeroE6 cells. Data shown are mean ± SD from three biological replicates. Independent experiments are shown as separate graphs when available.
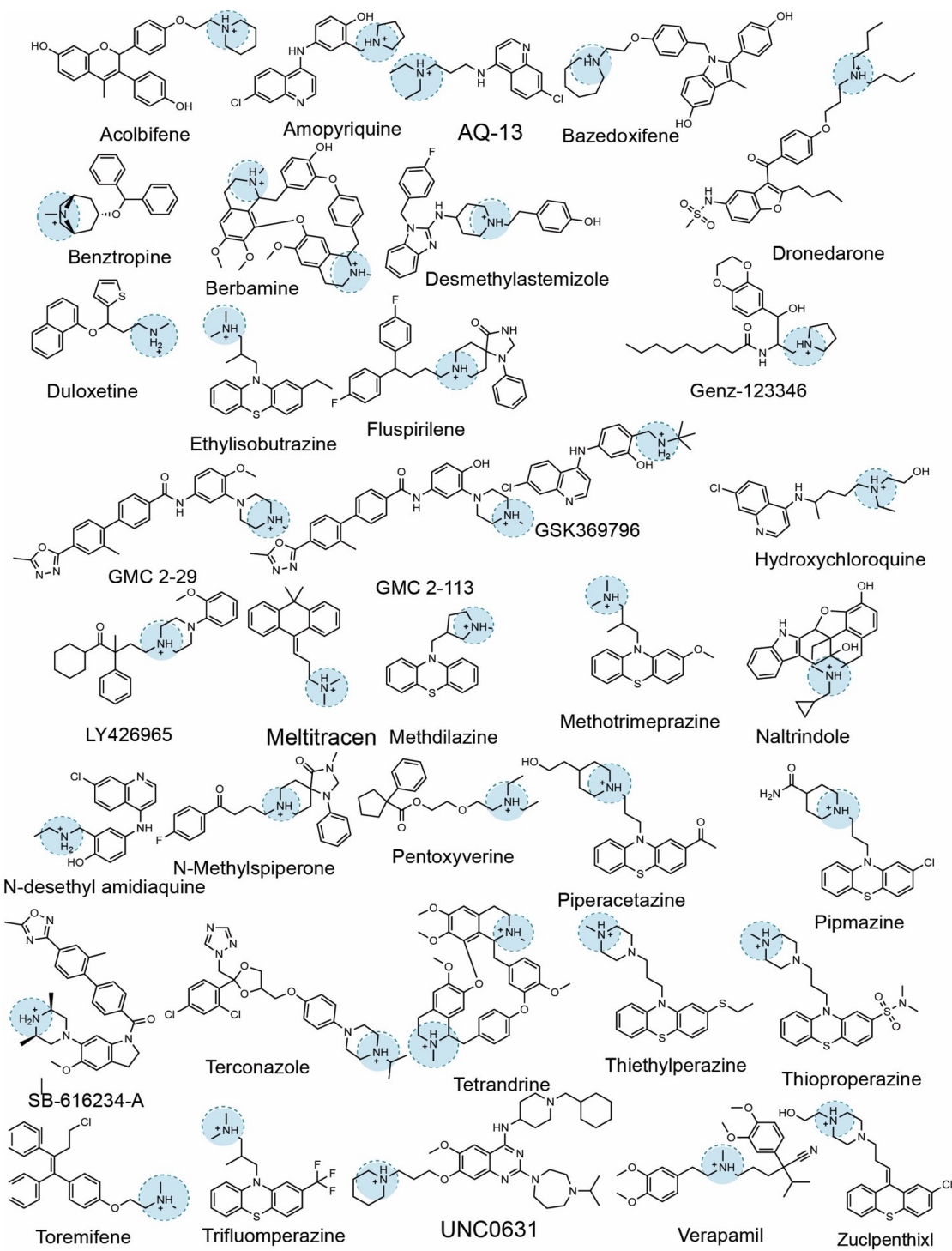
Figure 2.S4. Example cationic amphiphilic drugs identified from SARS-CoV-2 drug repurposing literature predicted to induce phospholipidosis.

A

Amiodarone

Clemastine

DMSO **Amiodarone** **Clemastine**

Sertraline **Benztropine** **PB28**

Sertraline **Benztropine** **PB28**

■ DIPL   ▲ Cell viability %
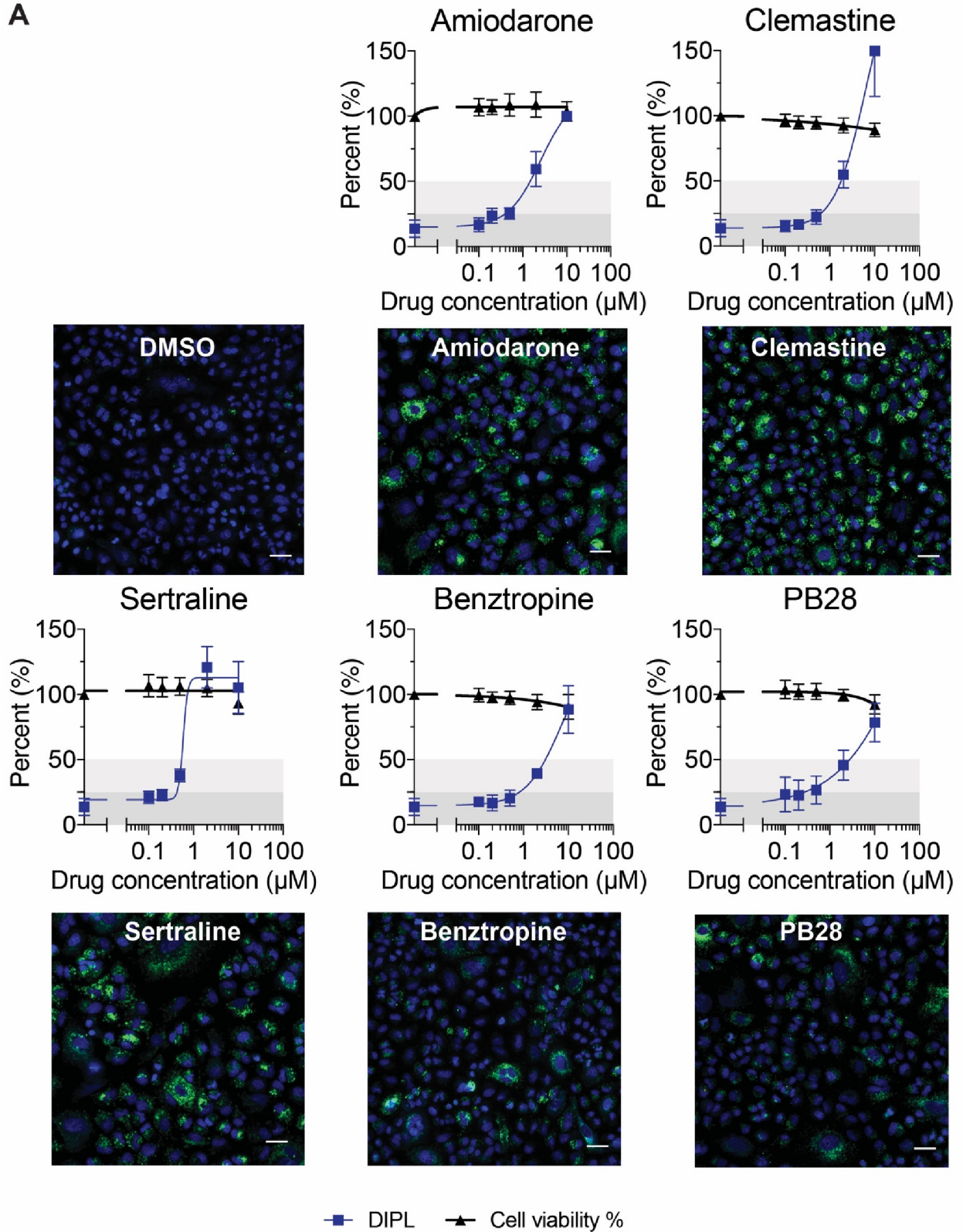
B
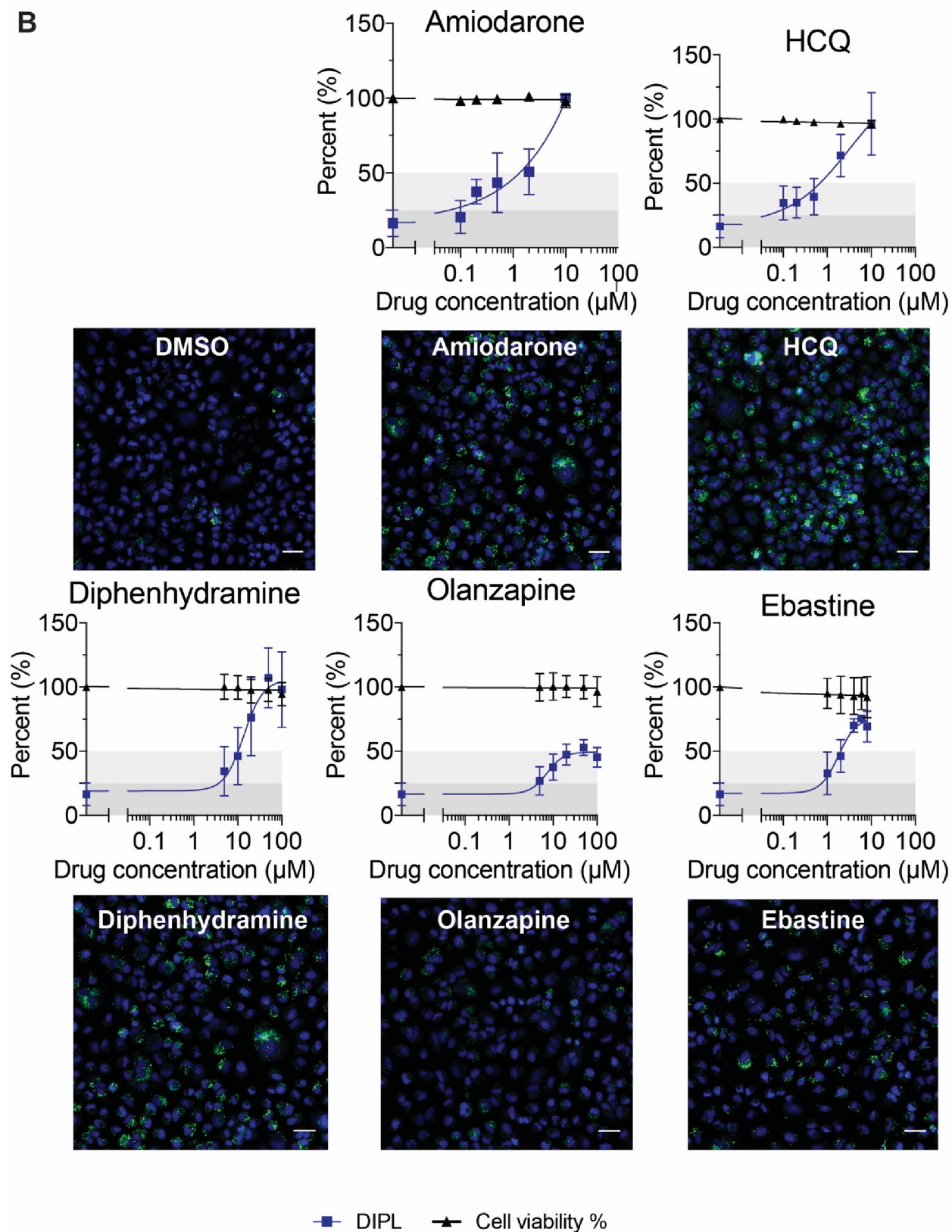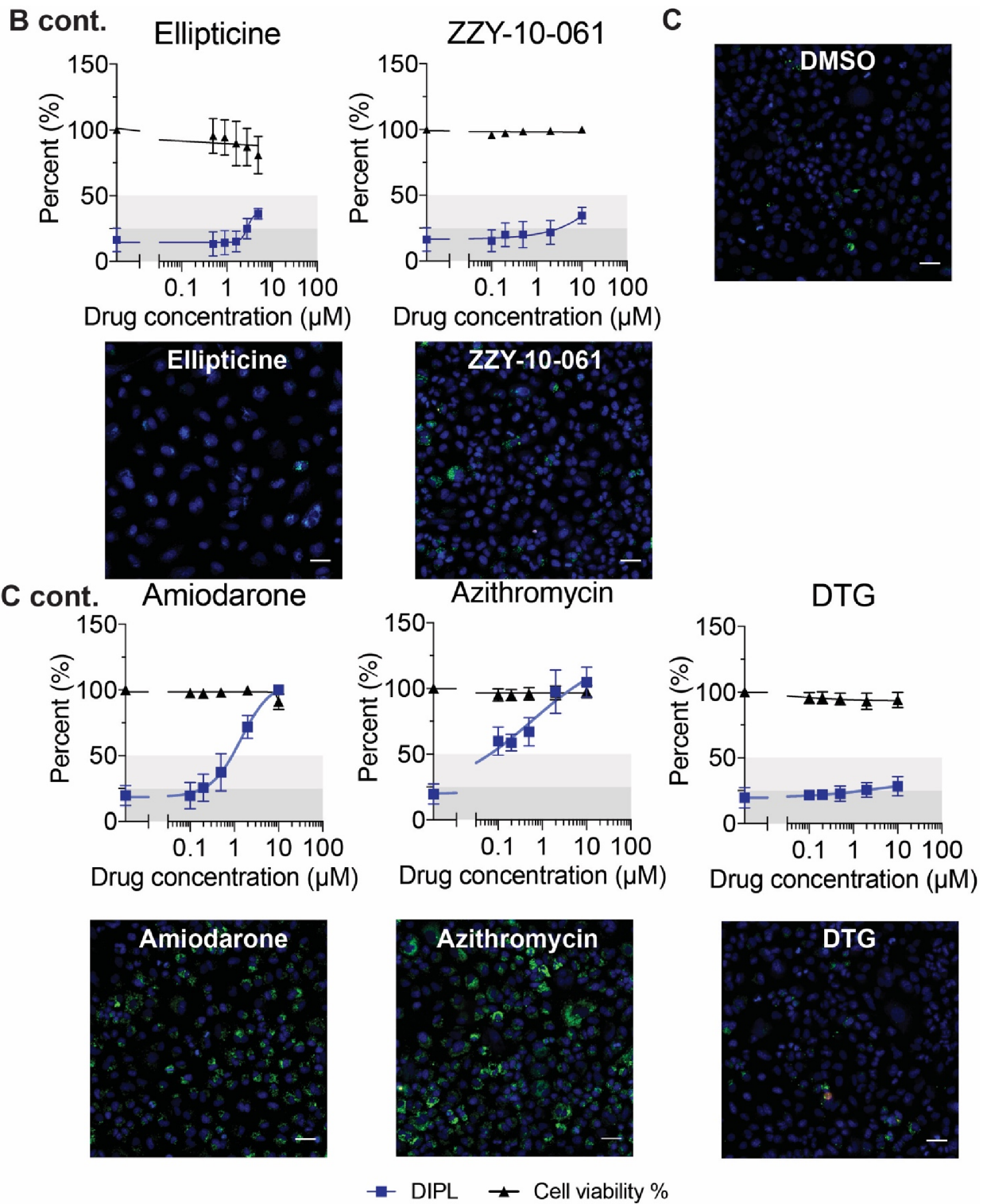
Figure 2.S5. Dose response curves for drugs measured in the phospholipidosis and cell viability assays and plate images at top tested concentrations. Batch 1- **A.**, batch 2- **B.,** and batch 3- **C.** Blue = Hoechst nuclei staining, Green = NBD-PE phospholipid staining, (Continued on the next page.)

(Continued from previous page.) Red = EthD-2 staining for dead cells. Dose response of NBD-PE aggregation was normalized to DMSO and 10 µM amiodarone from the same batch, and cell viability was normalized to DMSO. Data shown are pooled means ± SD from three independent experiments each with three biological replicates.
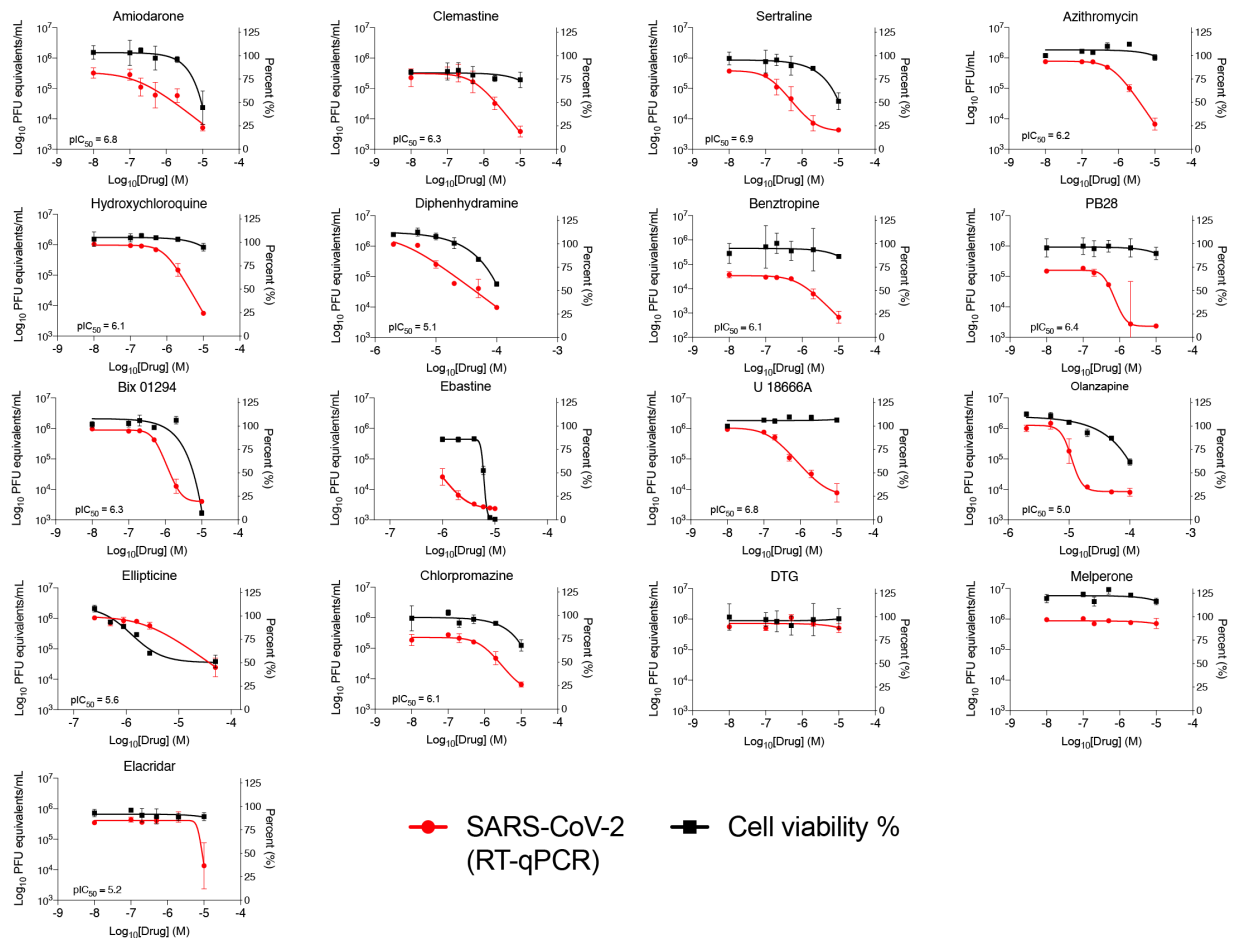
Figure 2.S6. Dose response curves for cationic amphiphilic drugs in the RT-qPCR viral infectivity assay that were measured for NBD-PE aggregation.
A. Viral infectivity and cell viability data for a subset of drugs that were selected for the DIPL correlation analysis. Data shown are mean ± SD from three biological replicates. The concentrations from these experiments match what was tested in the NBD-PE assay. Data for amiodarone, chlorpromazine, and hydroxychloroquine are reprinted from Gordon *et al.* with permission[14]. Data for ZZY-10-051 and ZZY-10-061 are in **Fig. 2.S8**.

A

B
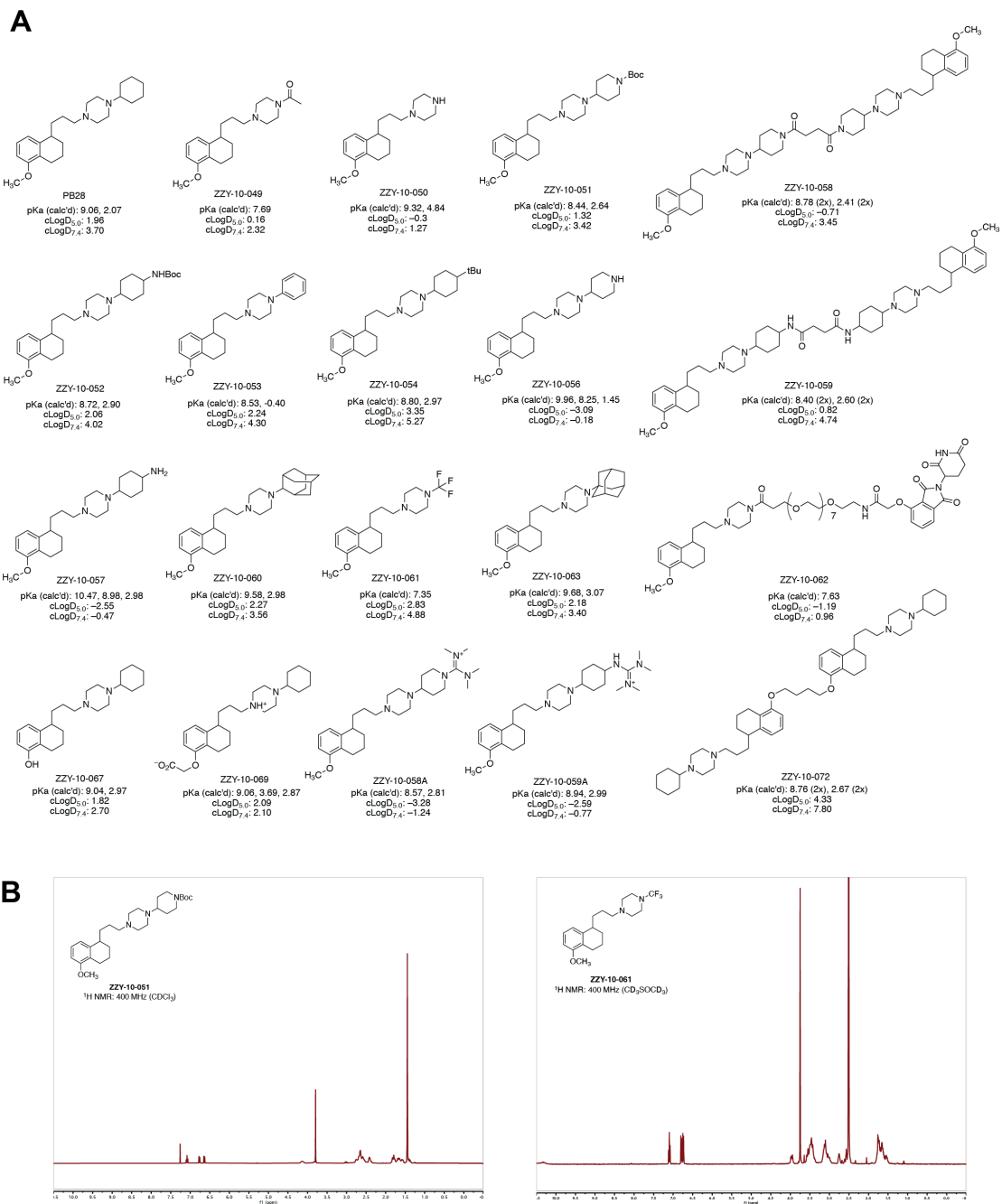
Figure 2.S7. PB28 analog structures.
A. Chemical structures of PB28 analogs tested, shown in their neutral form. All compounds were prepared as racemates. Compounds ZZY-10-061, ZZY-10-062, ZZY-10-064, ZZY-10-056, ZZY-10-057, ZZY-10-058, ZZY-10-059 and ZZY-10-072 contain mixtures of diastereomers that were not resolved or separated. With the exception of ZZY-10-061 and ZZY-10062, all compounds were prepared as HCl salts by acidification of their neutral forms with an ethereal solution of hydrogen chloride, or by lyophilization of their aqueous solutions in 50 mM HCl. **B.** ^{1}H NMR spectra for ZZY-10-051 and ZZY-10-061.

A

B

Sigma-1 Receptor

Sigma-2 Receptor

	S1R Ki	S2R Ki
ZZY-10-049	0.071 μM	>100 μM
ZZY-10-050	0.067 μM	0.26 μM
ZZY-10-056	0.65 μM	0.040 μM
ZZY-10-057	0.16 μM	0.011 μM

C

PB28 pIC$_{50}$ = 6.393

ZZY-10-049

ZZY-10-050 pIC$_{50}$ = 5.854

ZZY-10-051 pIC$_{50}$ = 6.066

ZZY-10-052 pIC$_{50}$ = 6.257

ZZY-10-053 pIC$_{50}$ = 5.892

ZZY-10-054 pIC$_{50}$ = 6.300

ZZY-10-056 pIC$_{50}$ = 7.204

ZZY-10-057 pIC$_{50}$ = 7.030

ZZY-10-063

ZZY-10-056 pIC$_{50}$ = 6.783

ZZY-10-060 pIC$_{50}$ = 5.804

ZZY-10-061 pIC$_{50}$ = 5.929

ZZY-10-062

ZZY-10-063

ZZY-10-067

ZZY-10-069

ZZY-10-072

ZZY-10-058A

ZZY-10-059A

SARS-CoV-2 RT-qPCR   Cell Viability %

46

Figure 2.S8. PB28 analog antiviral and sigma binding data.
A. Fractional binding of PB28 analogs against Sigma-1 (purple) and Sigma-2 (blue) normalized to a buffer control at 1.0 in a radioligand binding experiment. Data shown are mean ± SEM from three biological replicates. **B.** Dose-response curves for selected PB28 analogs in Sigma-1 and Sigma-2 radioligand competition binding assay. Data points shown are mean ± SEM from three biological replicates. **C.** Viral infectivity data for PB28 analogs A549-ACE2 cells. Data shown are mean ± SD from three biological replicates.
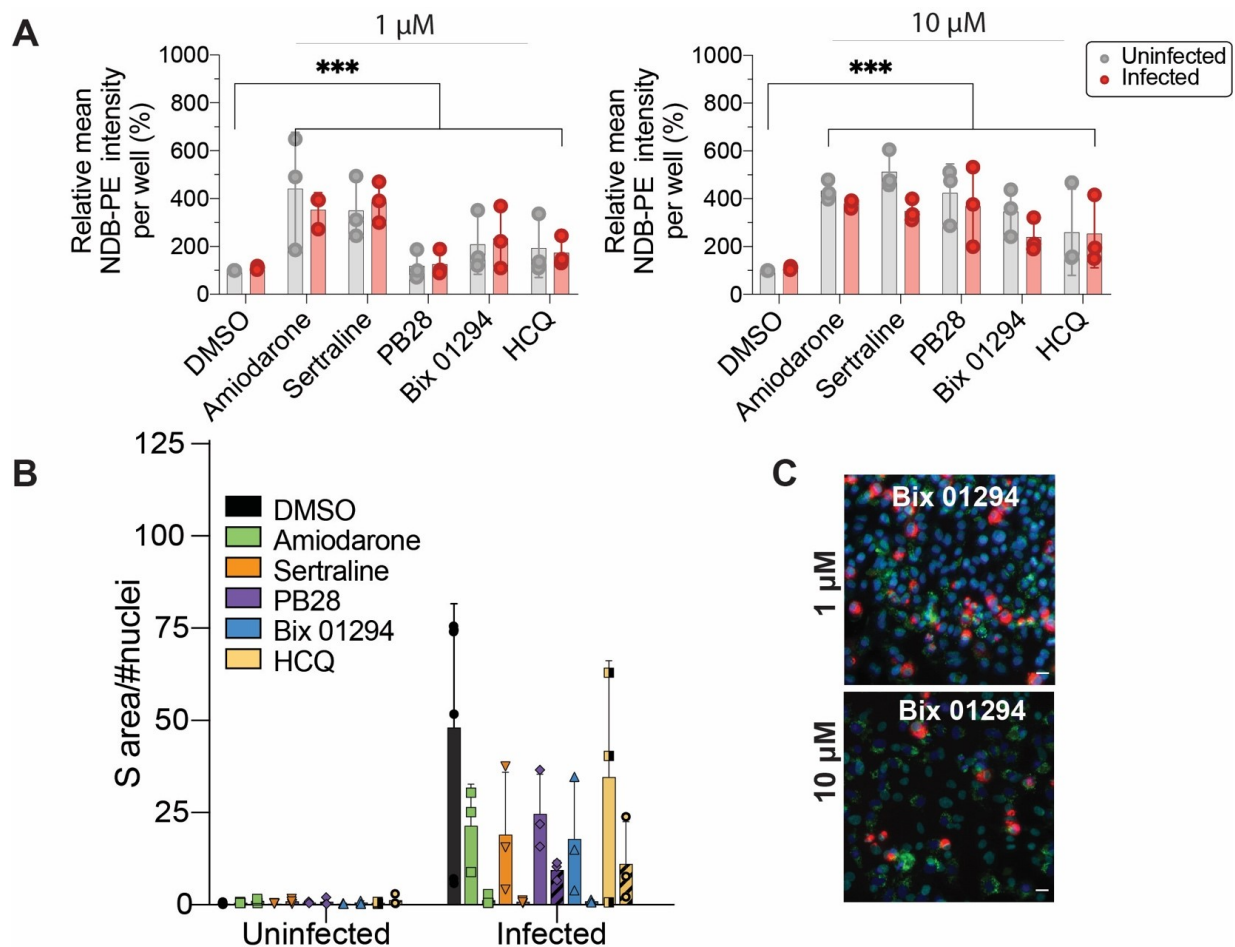
Figure 2.S9. Quantification of phospholipidosis and spike protein in the same cells. **A.** Relative mean ± SD NBD-PE intensity per well percent for 5 molecules and DMSO (1 and 10 µM) in uninfected and SARS-CoV-2 infected A549-ACE2 cells. Data shown are pooled from three independent experiments each in biological quadruplicate. Two-way ANOVA main effect of drug treatment at 1 µM, $F_{(5, 24)}$ = 7.7, ***$P < 0.001$, no main effect of infection state, $F_{(1,24)}$ = 0.02, P=0.90, and 10 µM, $F_{(5, 24)}$ = 9.1, ***$P < 0.001$, no main effect of infection state, $F_{(1,24)}$ = 3.48, P=0.07. **B.** Spike protein quantification in the same experiment as **A.** for both uninfected and SARSCoV-2 infected cells, and 1 (solid color bars) and 10 µM (hatched bars) drug treatments. Data represent mean ± SD from three independent experiments each in biological quadruplicate. Spike protein was quantified as S area / # nuclei per well. **C.** Example images from the costaining experiment measuring phospholipidosis and SARS-CoV-2 Spike protein in infected and uninfected A549-ACE2 cells. Bix 01294 (1 and 10 µM) is shown. Blue = Hoechst nuclei staining; Red = Spike protein staining; Green = NBD-PE phospholipid staining; Yellow = coexpression of spike protein and NBD-PE. Scale bar = 20 µm.
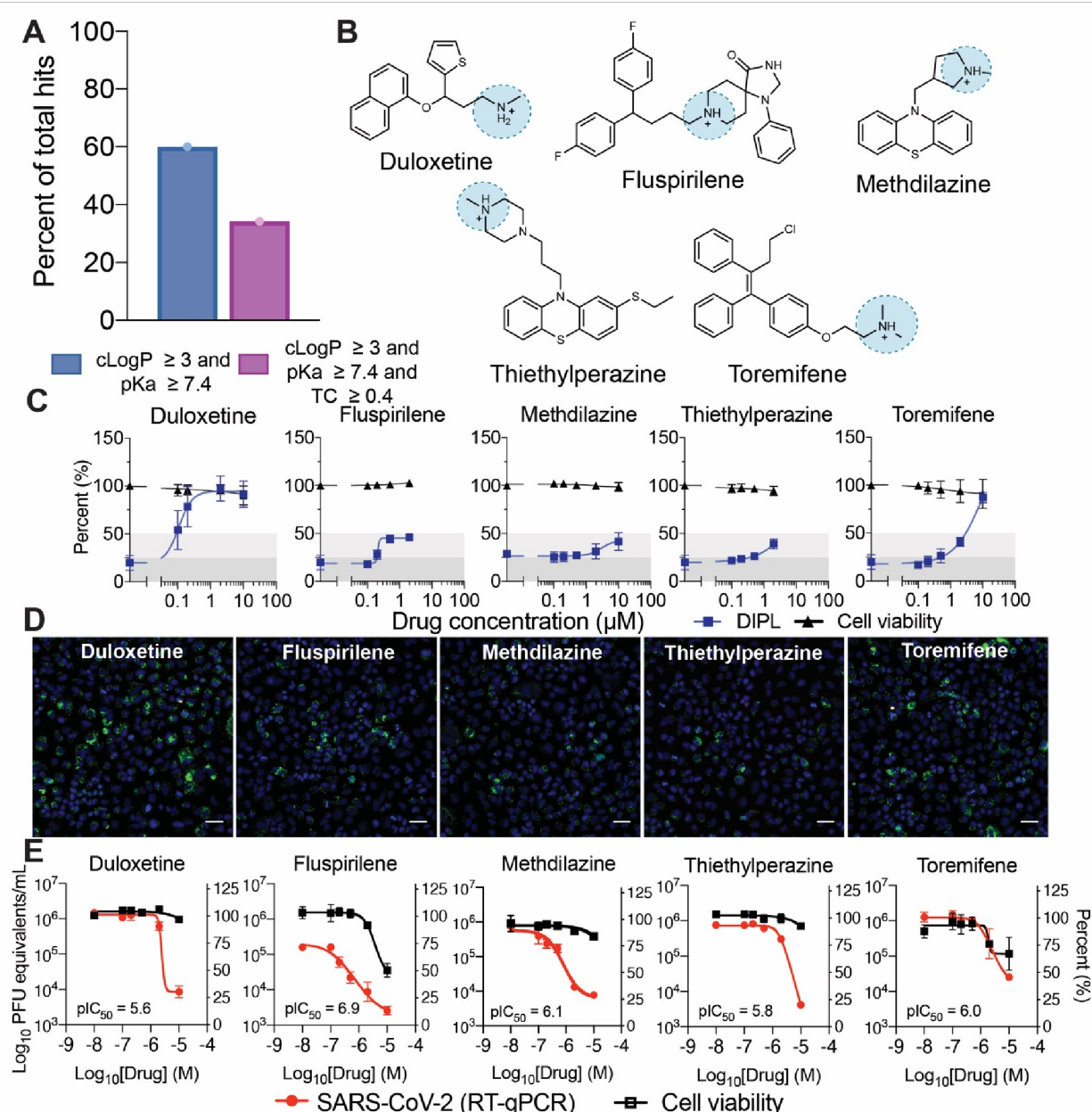
Figure 2.S10. Many drugs with activity against SARS-CoV-2 are CADs that induce phospholipidosis.

A. Percentage of total drug repurposing hits collected that pass CAD thresholds. **B.** Example repurposing hits from the literature that pass our CAD filters. **C.** Dose response curves for five predicted phospholipidosis inducers. All five induce measurable phospholipidosis (blue) with no impact on cell viability (black). **D.** Representative images of phospholipidosis quantification through NBD-PE staining in A549 cells. Blue = Hoechst nuclei staining, Green = NBD-PE phospholipid staining, Red = EthD-2 staining for dead cells. Scale bars = 20 μm. (Continued on the next page.)

(Continued from previous page.) **E.** Viral infectivity (red) and cytotoxicity (black) data for five example literature CADs tested in A549-ACE2 cells. Data shown are mean ± SD from three biological replicates.
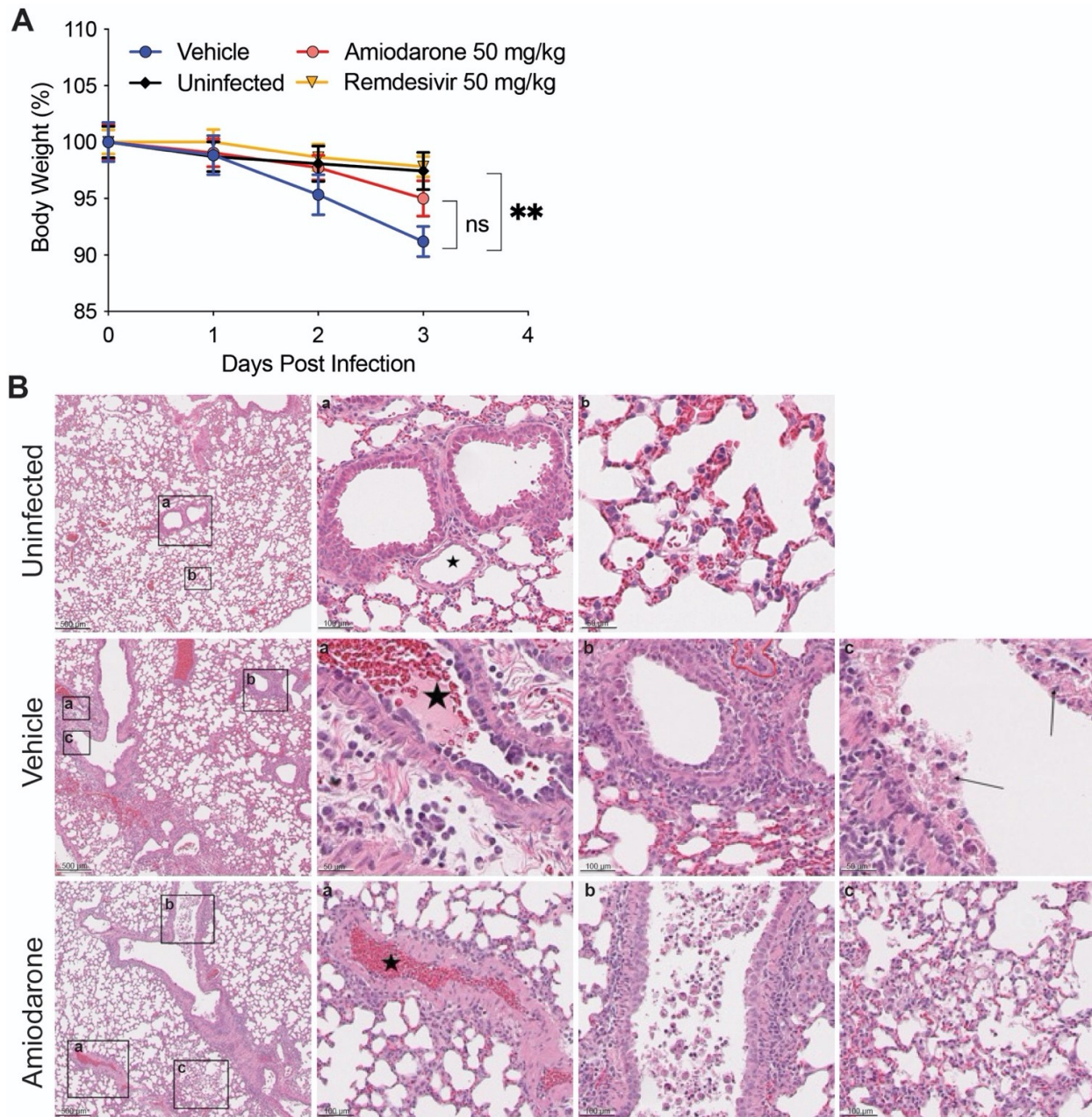
Figure 2.S11. Additional endpoints for 15-day dosing experiment.
Uninfected mice, vehicle-treated infected mice, and drug-treated mice were weighed over the course of three days after SARS-CoV-2 infection. Groups were compared with a two-way ANOVA (main effect of treatment $F_{(3,76)}$ = 2.98, P = 0.04; main effect of day post-infection, $F_{(3,76)}$ = 8.22, P < 0.0001; no interaction between variables, $F_{(9,76)}$ = 1.03, P = 0.42). Individual group comparisons determined using Dunnett's multiple comparison test are indicated (day 3: vehicle N = 6 vs. uninfected N = 6, **P = 0.006; vehicle vs. remdesivir N = 6, **P = 0.003; vehicle vs. amiodarone N = 5, ns = not significant, P = 0.17). Data are mean ± SEM. B. Lungs were harvested on day 3 post-infection and stained for hematoxylin and eosin. Regions of the lung anatomy where inflammation was assessed are highlighted by black boxes, (Continued on the next page.)

(Continued from previous page.) with the corresponding higher-magnification image indicated by matching letter. Regions where inflammation was detected are indicated by arrows. Quantification of histopathology scores are shown in **Fig. 2.5**.

2.7 Tables

Table 2.S1. Cationic amphiphilic drugs found active against other viruses in the literature.

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
Amiodarone	ZINC3830212	CCCCc1oc2ccccc2c1C(=O)c1cc(I)c(OCCN(CC)CC)c(I)c1	EBOV	VeroE6 EBOV-eGFP assay	EC50	7.6	27622822
Amodiaquine	ZINC608172	CCN(CC)Cc1cc(Nc2ccnc3cc(Cl)ccc23)ccc1O	MARV	EAhy MARV GP pseudoparticle assay	EC50	1.18 ug/mL	24710028
			SARS	VeroE6 SARS CPE assay	IC50	4.18	33060197
			HCV	Huh-7.5.1 luc-IRES assay	IC50	2.1	23659500
			EBOV	VeroE6 EBOV-eGFP assay	EC50	34	27622822
			MERS	VeroE6 MERS ELISA	EC50	6.212	24841273
			SARS	VeroE6 SARS CPE assay	EC50	1.274	24841273
Aripiprazole	ZINC1851149	O=C1CCc2ccc(OCCCN3CCN(c4cccc(Cl)c4Cl)CC3)cc2N1	EBOV	VeroE6 eGFP-EBOV assay	IC50	8.1	26041706
Astemizole	ZINC601274	COc1ccc(CCN2CCC(Nc3nc4ccccc4n3Cc3ccc(F	EBOV	VeroE6 eGFP-EBOV assay	IC50	6.17	26041706

53

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
		)cc3)CC2)cc1					
Benztropine	ZINC100036536	CN1[C@H]2CC[C@@H]1C[C@H](OC(c1ccccc1)c1ccccc1)C2	EBOV	VeroE6 eGFP-EBOV assay	IC50	8.07	26041706
			MERS	VeroE6 MERS ELISA	EC50	16.627	24841273
			SARS	VeroE6 SARS CPE assay	EC50	21.611	24841273
			EBOV	A549 HIV/EBOV pseudotyped virus	IC50	3.7	26202243
			MARV	A549 HIV/MARV pseudotyped virus + luciferase reporter gene assay	IC50	13.2	26202243
			EBOV	VeroE6 EBOV-eGFP assay	EC50	9.2	27622822
Chlorcyclizine		Clc1ccc(cc1)C(c2ccccc2)N3CCN(CC3)C	HCV	Huh7.5.1 HCV RT-qPCR assay	EC50	0.0331	25855495
Chloroquine	ZINC19144226	CCN(CC)CCC[C@@H](C)Nc1ccnc2cc(Cl)ccc12	MERS	VeroE6 MERS ELISA	EC50	6.275	24841273

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
			SARS	VeroE6 SARS CPE assay	EC50	6.538	24841273
			EBOV	VeroE6 EBOV-eGFP assay	EC50	16	27622822
Chlorpromazine	ZINC44027	CN(C)CCCN1c2ccccc2Sc2ccc(Cl)cc21	MERS	VeroE6 MERS ELISA	EC50	9.514	24841273
			SARS	VeroE6 SARS CPE assay	EC50	12.971	24841273
Clemastine	ZINC402830	CN1CCC[C@@H]1CCO[C@](C)(c1ccccc1)c1ccc(Cl)cc1	EBOV	VeroE6 eGFP-EBOV assay	IC50	5.44	26041706
			EBOV	VeroE6 EBOV-eGFP assay	EC50	5.2	27622822
Clomiphene	ZINC1530601	CCN(CC)CCOc1ccc(/C(=C(/Cl)c2ccccc2)c2ccccc2)cc1	EBOV	VeroE6 eGFP-EBOV assay	IC50	2.42	26041706
			EBOV	VeroE6 EBOV-eGFP assay	EC50	11	27622822
Clomipramine	ZINC20248	CN(C)CCCN1c2ccccc2CCc2ccc(Cl)cc21	EBOV	VeroE6 eGFP-EBOV assay	IC50	11.4	26041706
			MERS	VeroE6 MERS ELISA	EC50	9.332	24841273
			SARS	VeroE6 SARS CPE assay	EC50	13.238	24841273
Fluphenazine	ZINC19203912	OCCN1CCN(CCC	EBOV	VeroE6 eGFP-EBOV assay	IC50	5.54	26041706

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
		N2c3cccc c3Sc3ccc (C(F)(F)F)cc32)CC 1					
			MERS	VeroE6 MERS ELISA	EC50	5.868	24841273
			SARS	VeroE6 SARS CPE assay	EC50	21.431	24841273
			EBOV	VeroE6 EBOV-eGFP assay	EC50	12	27622822
Fluspirilene	ZINC53 7755	O=C1NC N(c2cccc c2)C12C CN(CCC C(c1ccc(F)cc1)c1c cc(F)cc1) CC2	MERS	VeroE6 MERS ELISA	EC50	7.477	24841273
			SARS	VeroE6 SARS CPE assay	EC50	5.963	24841273
Hydroxychloroq uine	ZINC15 30652	CCN(CC O)CCC[C @@H](C) Nc1ccnc2 cc(Cl)ccc 12	MERS	VeroE6 MERS ELISA	EC50	8.279	24841273
			SARS	VeroE6 SARS CPE assay	EC50	7.966	24841273
			EBOV	VeroE6 EBOV-eGFP assay	EC50	22	27622822
Hydroxyzine	ZINC19 364222	OCCOCC N1CCN([C@@H](c2ccccc2) c2ccc(Cl) cc2)CC1	HCV	Huh7.5.1 HCV RT-qPCR assay	EC50	0.0503	25855495

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
Maprotiline	ZINC15 30688	CNCCCC 12CCC(c 3ccccc31) c1ccccc1 2	EBOV	VeroE6 eGFP-EBOV assay	IC50	9.63	26041706
Paroxetine	ZINC52 7386	Fc1ccc([C @@H]2C CNC[C@ H]2COc2 ccc3c(c2) OCO3)cc 1	EBOV	VeroE6 eGFP-EBOV assay	IC50	7.45	26041706
			EBOV	VeroE6 EBOV-eGFP assay	EC50	27	27622822
Pimozide	ZINC41 75630	Oc1nc2cc ccc2n1C1 CCN(CC CC(c2ccc (F)cc2)c2 ccc(F)cc2)CC1	EBOV	VeroE6 eGFP-EBOV assay	IC50	3.12	26041706
Prochlorperazin e	ZINC19 796018	CN1CCN(CCCN2c3 ccccc3Sc 3ccc(Cl)c c32)CC1	EBOV	VeroE6 eGFP-EBOV assay	IC50	5.96	26041706
			EBOV	VeroE6 EBOV-eGFP assay	EC50	11	27622822
Promazine	ZINC10 402	CN(C)CC CN1c2cc ccc2Sc2c cccc21	EBOV	VeroE6 EBOV-eGFP assay	EC50	21	27622822
Promethazine	ZINC20 250	C[C@@H](CN1c2c cccc2Sc2 ccccc21) N(C)C	MERS	VeroE6 MERS ELISA	EC50	11.802	24841273

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
			SARS	VeroE6 SARS CPE assay	EC50	7.545	24841273
			EBOV	A549 HIV/EBOV pseudotyped virus + luciferase reporter gene assay	IC50	19.4	26202243
			MARV	A549 HIV/MARV pseudotyped virus + luciferase reporter gene assay	IC50	19.1	26202243
Sertraline	ZINC18 53550	CN[C@H] 1CC[C@ @H](c2cc c(Cl)c(Cl) c2)c2cccc c21	EBOV	VeroE6 eGFP-EBOV assay	IC50	3.13	26041706
Tamoxifen	ZINC15 30689	CC/C(=C(\c1ccccc1)c1ccc(O CCN(C)C)cc1)c1cc ccc1	MERS	VeroE6 MERS ELISA	EC50	10.117	24841273
			SARS	VeroE6 SARS CPE assay	EC50	92.886	24841273
			EBOV	VeroE6 EBOV-eGFP assay	EC50	3	27622822
Thi-ethylperazine	ZINC22 446674	CCSc1cc c2c(c1)N(CCCN1C CN(C)CC 1)c1ccccc	MERS	VeroE6 MERS ELISA	EC50	7.865	24841273

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
Thioridazine	ZINC15 30695	CSc1ccc2 c(c1)N(C C[C@@H]1CCCCN 1C)c1ccc cc1S2	EBOV	VeroE6 eGFP-EBOV assay	IC50	6.24	26041706
Toremifene	ZINC12 404516	CN(C)CC Oc1ccc(/ C(=C(/CC Cl)c2cccc c2)c2cccc c2)cc1	EBOV	VeroE6 eGFP-EBOV assay	IC50	0.162	26041706
			MERS	VeroE6 MERS ELISA	EC50	12.915	24841273
			SARS	VeroE6 SARS CPE assay	EC50	11.969	24841273
Triflupromazine	ZINC53 8507	CN(C)CC CN1c2cc ccc2Sc2c cc(C(F)(F)F)cc21	MERS	VeroE6 MERS ELISA	EC50	5.758	24841273
			SARS	VeroE6 SARS CPE assay	EC50	6.398	24841273
Trimipramine	ZINC96 8275	C[C@H](CN(C)C) CN1c2cc ccc2CCc 2ccccc21	EBOV	A549 HIV/EBOV pseudotyped virus + luciferase reporter gene assay	IC50	11.1	26202243
			MARV	A549 HIV/MARV pseudotyped virus + luciferase reporter gene assay	IC50	10.9	26202243

Compound	ZINC ID	SMILES	Virus	Antiviral Assay	Activity Type	Activity Value (uM)	Antiviral Reference
Triparanol	ZINC16 92389	CCN(CC) CCOc1cc c([C@](O)(Cc2ccc(Cl)cc2)c2 ccc(C)cc2)cc1	MERS	VeroE6 MERS ELISA	EC50	5.283	24841273
			EBOV	VeroE6 EBOV-eGFP assay	IC50	1.92	23441171
U 18666A	ZINC11 8915627	[H][C@@]23CC=C 1C[C@@ H](OCCN (CC)CC) CC[C@@](C)1[C@] ([H])2CC[C@@]4(C)[C@]([H])3CCC 4=O	EBOV	VeroE6 EBOV-eGFP assay	IC50	8.05	23441171
			DENV	A549 dengue replicon assay	EC50	6.2	22146564

Table 2.S2. Measured pharmacokinetic parameters for Amiodarone.

Amiodarone		Dose Route	Dose (mg/kg)	T_{max} (min)	Cmax, ng/mL (g)	$AUC_{0 \to tlast}$ ng*min/mL (g)	$AUC_{0 \to \infty}$ ng*min/ mL (g)	$T_{1/2}$ (min)	K_{el} (min^{-1})	Cmax (nM)
MWT (Da) 681.8	Plasma	i.p[a]	0.3	30	37.4	18800	34200	1270	0.000548	55
		i.p.	1	30	73.5	42000	52500	583	0.00119	108
		i.p.	3	30	178	96700	144000	769	0.000901	261
		i.p.	10	15	519	205000	278000	585	0.0118	761
		i.p.	30	60	1070	476000	577000	505	0.00137	1569
	Lung	i.p.	0.3	360	120	121000	162000	698	0.000993	176
		i.p.	1	120	328	310000	375000	599	0.00116	481
		i.p.	3	120	1240	1110000	1230000	501	0.00138	1819
		i.p.	10	120	8020	3650000	4870000	645	0.00107	11763
		i.p.	30	120	13900	8930000	17900000	1250	0.000554	20387

[a]i.p. = intraperitoneally

Table 2.S3. Measured pharmacokinetic parameters for Sertraline.

Sertraline		Dose Route	Dose (mg/kg)	T_{max} (min)	Cmax, ng/mL (g)	$AUC_{0 \to tlast}$ ng*min/mL (g)	$AUC_{0 \to \infty}$ ng*min/ mL (g)	$T_{1/2}$ (min)	K_{el} (min^{-1})	Cmax (nM)
MWT (Da) 306.2	Plasma	i.p[a]	0.3	15	11.8	979	10400	626	0.00111	39
		i.p.	1	30	18.4	1750	3730	123	0.00565	60
		i.p.	3	30	79	11500	14600	147	0.00472	258
		i.p.	10	30	332	41100	44700	103	0.00672	1084
		i.p.	30	60,0	902	193000	198000	272	255	2945
	Lung	i.p.	0.3	60	949	163000	197000	134	0.00518	3099
		i.p.	1	60	3290	576000	845000	200	0.00347	10743
		i.p.	3	30	11500	3270000	3320000	257	0.0027	37553
		i.p.	10	30	29100	6700000	6740000	198	0.0035	95025
		i.p.	30	15	43500	14500000	14500000	188	0.00369	142047

[a]i.p. = intraperitoneally

Table 2.S4. Measured pharmacokinetic parameters for Tamoxifen.

PB28		Dose Route	Dose (mg/kg)	T_{max} (min)	Cmax, ng/mL (g)	$AUC_{0 \to tlast}$ ng*min/mL (g)	$AUC_{0 \to \infty}$ ng*min/ mL (g)	$T_{1/2}$ (min)	K_{el} (min^{-1})	Cmax (nM)
MWT (Da) 370.6	Plasma	i.p.[a]	0.3	30	15.8	269	ND[b]	ND	ND	43
		i.p.	1	30	16.4	3260	4690	200	0.00347	44
		i.p.	3	15	25.1	5760	7120	169	0.00411	68
		i.p.	10	15	89.6	13700	29700	431	0.00161	242
		i.p.	30	5	336	41900	56300	191	0.00362	907
	Lung	i.p.	0.3	15	206	37600	50800	188	0.00368	556
		i.p.	1	30	735	116000	144000	150	0.00463	1983
		i.p.	3	15	2500	316000	367000	130	0.00535	6746
		i.p.	10	15	6720	2120000	2130000	218	0.00318	18134
		i.p.	30	5	20600	7350000	7380000	186	0.00373	55590

[a]i.p. = intraperitoneally
[b]ND = not determined

Table 2.S5. Measured pharmacokinetic parameters for PB28.

Tamoxifen		Dose Route	Dose (mg/kg)	T_{max} (min)	Cmax, ng/mL (g)	$AUC_{0 \to tlast}$ ng*min/mL (g)	$AUC_{0 \to \infty}$ ng*min/ mL (g)	$T_{1/2}$ (min)	K_{el} (min^{-1})	Cmax (nM)
MWT (Da) 371.5	Plasma	i.p.[a]	0.3	ND[b]	ND	ND	ND	ND	ND	NA
		i.p.	1	120	13.7	4030	23100	1120	0.00062	37
		i.p.	3	120	49.4	13600	36700	464	0.00149	133
		i.p.	10	120	105	31400	42100	669	0.00104	283
		i.p.	30	360	436	409000	493000	538	0.00129	1174
	Lung	i.p.	0.3	360	174	147000	152000	278	0.00249	468
		i.p.	1	360	639	572000	609000	348	0.00199	1720
		i.p.	3	120	1970	1640000	1770000	398	0.00174	5303
		i.p.	10	360	4950	4390000	4650000	337	0.00205	13324
		i.p.	30	360	33400	28200000	31800000	430	0.00161	89902

[a]i.p. = intraperitoneally
[b]ND = not determined

Table 2.S6. Measured pharmacokinetic parameters for Elacridar.

Elacridar		Dose Route	Dose (mg/kg)	T_{max} (min)	Cmax, ng/mL (g)	$AUC_{0 \to tlast}$ ng*min/mL (g)	$AUC_{0 \to \infty}$ ng*min/ mL (g)	$T_{1/2}$ (min)	K_{el} min^{-1}	Cmax (nM)
MWT (Da) 563.7	Plasma	i.p[a]	0.3	120	42.4	26900	29900	437	0.00159	75
		i.p.	1	60	78.9	57900	80300	790	0.000878	140
		i.p.	3	120	182	141000	155000	427	0.00162	323
		i.p.	10	360	489	511000	752000	836	0.000829	868
		i.p.	30	360	717	802000	1400000	1120	0.000618	1272
	Lung	i.p.	0.3	60	507	286000	350000	604	0.00115	899
		i.p.	1	60	1560	764000	864000	478	0.00145	2768
		i.p.	3	60	4090	2550000	2650000	330	0.0021	7256
		i.p.	10	60	13300	10600000	12600000	571	0.00121	23596
		i.p.	30	120	18700	17900000	37500000	1470	0.000471	33176

[a]i.p. = intraperitoneally

Table 2.S7. Estimates of expenditures of COVID-19 cationic amphiphilic drug clinical trials.

Phase	Mean cost[a]	Number of CAD trials[b]	Cost of CAD trials	Number of non-CAD HCQ/ CQ trials	Cost of non-CAD HCQ/ CQ trials	Number of non HCQ/ CQ CAD trials[c]	Cost of non-HCQ/ CQ CAD trials	Number of small molecule trials	Cost of small molecule trials	Percentage CAD trials versus all trials
0		9		4		5		58		
1	$4.2	13	$55	7	$29	6	$25	56	$235	23%
2	$14.2	99	$1406	48	$682	51	$724	409	$5808	24%
3	$22.8	162	$3694	98	$2234	64	$1459	428	$9758	38%
1/2[e]	$14.2	4	$57	2	$28	2	$28	40	$568	10%
2/3[e]	$22.8	38	$866	25	$570	13	$296	125	$2850	30%
3/4[e]	$11	0						3	$33	
4	$11	68	$748	45	$495	23	$253	154	$1694	44%
Not available		46		29		17		217		21%
Phase I to III Totals		316	$6078	180	$3544	136	$2533	1058	$19219	30%
Phase I to IV Totals		442	$6826	258	$4039	181	$2786	1490	$20946	29%

[a]All costs are in US millions of dollars, not accounting for inflation since the time of the study (*50, 51*)
[b]Cationic amphiphilic drugs (CADs) estimated using cLogP ≥ 3 and pKa ≥ 7.4 criteria; trials using azithromycin, a non-CAD PLD inducer, is included in these numbers; trials are only counted one time if more than one CAD was administered in a trial
[c]Trials that administered a CAD in addition to HCQ or CQ are included in these numbers
[e] Costs for mixed-phase trials were estimated to be the cost of only the higher phase trial

2.8 Supplemental Files

Supplementary File 1. Summary information for compounds tested in antiviral and binding studies.
Tab 1 includes Sigma-1 and Sigma-2 binding affinities, antiviral pIC50 values in the anti-NP (New York) and RT-qPCR (Paris) antiviral assays, vendor codes, and physico-chemical property information. Tab 2 includes raw binding data expressed as percent radioligand binding for each compound at Sigma-1 and Sigma-2.

Supplementary File 2. Summary information for compounds identified as antiviral hits in SARS-CoV-2 drug repurposing studies.
Tab 1 includes a list of hit compounds identified in each SARS-CoV-2 repurposing study explored in this paper, their physico-chemical properties, and the Tanimoto coefficients and name of each compounds' most similar known phospholipidosis inducer.

2.9 Materials and Methods

Competition Binding Assays. Competition curves were measured using membranes from Expi293F cells (Thermo, A14527) with a stably integrated tetracycline repressor[54] (provided by the lab of Dr. Robert Lefkowitz, lefko001@receptor-biol.duke.edu) transiently overexpressing either the human Sigma-1 or the human Sigma-2 receptors. For binding assays, ^{3}H-(+)-pentazocine (PerkinElmer, net1056250uc) and ^{3}HDTG (PerkinElmer, net986250uc) were used as the radioactive probes for Sigma-1 and Sigma-2, respectively. Membranes were incubated in a 100 µL reaction with 50 mM Tris (pH 8.0), 0.1% (w/v) bovine serum albumin (Rockland, BSA-50), 10 nM radioligand, and eight concentrations of the competing cold ligand. Reactions were incubated for two hours at 37 °C and then were terminated by filtration through a glass fiber filter using a Brandel harvester. Glass fiber filters were pre-soaked in 0.3% (v/v) polyethylenimine for 30 minutes at room temperature before harvesting. All reactions were performed in triplicate using a 96-well block format. After the membranes were transferred to the filters and washed, the filters were soaked in 5 mL Cytoscint scintillation fluid (MP Biomedicals, 0188245304) overnight, and radioactivity was measured using a Beckman Coulter LS 6500 scintillation counter. K_D values for each receptor were calculated in GraphPad Prism version 8.0.0 (San Diego, CA) using a saturation binding assay model with eight concentrations of the radioactive ligand. Non-specific binding was measured in the presence of 10 µM haloperidol (Tocris, 0931). The K_D of the sigma-1 radioligand probe was measured to be 21 nM and K_D of the sigma-2 radioligand probe was measured to be 15 nM, and these values were used to calculate K_i values for cold ligands in GraphPad Prism using the Nonlinear fit, one site-fit K_i method after normalizing to solvent baseline

at 100 (Percent = 100*Value/Baseline).

Antiviral drug and cell viability assays at Institut Pasteur. A549-ACE2 cells were kindly provided by the lab of Dr Olivier Schwartz[55]. A549-ACE2 cells were propagated at 37°C, 5% CO_2 in DMEM supplemented with 10% FBS and 20 µg/mL blasticidin S. 6250 of these cells per well were seeded into 384-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO_2. Two hours prior to infection, the media was replaced with 50 µL of DMEM (2% FBS) containing the compound of interest at the indicated concentration. At the time of infection, the media was replaced with virus inoculum (multiplicity of infection, MOI = 0.1 PFU/cell) and incubated for one hour at 37°C, 5% CO_2. The SARS-CoV-2 strain used (BetaCoV/France/IDF0372/2020) was propagated once in Vero-E6 cells and is a kind gift from the National Reference Centre for Respiratory Viruses at Institut Pasteur, Paris, originally supplied through the European Virus Archive goes Global platform. Following the adsorption period, the inoculum was removed, replaced with 50 µL of drug-containing media, and cells were incubated for an additional 72 hours at 37°C, 5% CO_2. At this point, the cell culture supernatant was harvested and viral load was assessed by RT-qPCR. Briefly, the cell culture supernatant was collected, heat inactivated at 95°C for 5 minutes and used for RT-qPCR analysis. SARS-CoV-2 specific primers targeting the N gene region: 5′-TAATCAGACAAGGAACTGATTA-3′ (Forward) and 5′-CGAAGGTGTGACTTCCATG-3′ (Reverse) were used with the Luna Universal One-Step RT-qPCR Kit (New England Biolabs, #E3005) in an Applied Biosystems QuantStudio 6 thermocycler, with the following cycling conditions: 55°C for 10 min, 95°C for 1 minute, and 40 cycles of 95°C

for 10 seconds, followed by 60°C for 1 minute. The number of viral genomes is expressed as PFU equivalents/m, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titer. Data was fit using nonlinear regression and IC_{50}s for each experiment were determined using GraphPad Prism version 8.0.0 (San Diego, CA).

Cell viability was assayed in uninfected, drug-treated cells using the CellTiter-Glo assay following the manufacturer's instructions (Promega, G7570). Luminescence was measured in a Tecan Infinity 2000 plate reader, and percentage viability calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol (0% viability), included in each experiment.

Cells and viruses for Anti-NP Immunofluorescence. Vero E6 (ATCC, CRL-1586) cells were maintained in DMEM (Corning, 10-013) supplemented with 10% FB (Peak Serum, PS-FB) and Penicillin/Streptomycin (Corning, 30-002) at 37°C and 5% CO_2. This cell line was regularly screened for mycoplasma contamination using the Universal Detection Kit (ATCC, 30-1012K). Cells were infected with SARS-CoV-2, isolate USA-WA1/2020 (BEI Resources NR-52281) under biosafety level 3 (BSL3) containment in accordance with the biosafety protocols developed by the Icahn School of Medicine at Mount Sinai. Viral stocks were grown in Vero E6 cells as previously described[56] and were validated by genome sequencing.

Antiviral drug and cell viability assays at Mt. Sinai. Two thousand Vero E6 cells were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO_2. Two hours before infection, the medium was replaced with 100 µL of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. Plates were then transferred into the BSL3 facility and 100 PFU (MOI = 0.025) was added in 50 µL of DMEM (2% FBS), bringing the final compound concentration to those indicated. Plates were then incubated for 48 hours at 37°C. After infection, supernatants were removed and cells were fixed with 4% formaldehyde for 24 hours prior to being removed from the BSL3 facility. The cells were then immunostained for the viral NP protein (an inhouse mAb 1C7, provided by Dr. Thomas Moran, Thomas.Moran@mssm.edu) with a DAPI counterstain. Infected cells (488 nm) and total cells (DAPI) were quantified using the Celigo (Nexcelcom) imaging cytometer. Infectivity was measured by the accumulation of viral NP protein in the nucleus of the Vero E6 cells (fluorescence accumulation). Percent infection was quantified as ((Infected cells/Total cells) - Background) *100 and the DMSO control was then set to 100% infection for analysis. Data was fit using nonlinear regression and IC_{50}s for each experiment were determined using GraphPad Prism version 8.0.0 (San Diego, CA). Cytotoxicity was also performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected VeroE6 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates. The Vero E6 cell line used in this study is a kidney cell line; therefore, we cannot exclude that lung cells yield different results for some inhibitors (see also 'Antiviral drug and cell viability assays at Institut Pasteur' for

studies carried out at Institut Pasteur).

Phospholipidosis quantification in uninfected cells. Phospholipidosis was assessed as previously described (*20*). Briefly, A549 cells (ATCC, ref. CCL-185) were cultivated in Ham's F12-K Medium (ThermoFisher, ref. 21127-022) containing 10% FCS and seeded in a black 96-well plate with clear bottom at a density of 15000 cells per well. The day after seeding, the cells were treated for 24 hours with a dose-range of different drugs in presence of 7.5 µM NBD-PE (ThermoFisher, ref. N360). The final DMSO concentration was 0.2%. Amiodarone (Hydrochloride salt, internal Novartis supply) was used as a positive control for phospholipidosis.

Before imaging, the cells were stained for 20 minutes at 37°C, 5% CO_2 with a solution containing Hoechst (ThermoFisher, ref. H3570) (10 µg/ml) and Ethidium homodimer-2 (EthD-2; ThermoFisher, E3599) (2 µM) in complete culture medium for visualizing the total and dead cell populations respectively. Cells were washed once with pre-warmed HBSS +/+ and images were taken on an Arrayscan XTI (ThermoFisher) equipped with a 20x objective and the LED/filter combinations BGRFR_386_23, BGRFR_485_20 and BGRFR_549_15 for acquisition of Hoechst, NBD-PE and EthD-2 dyes respectively. The images were analyzed using the HCS Studio software (ThermoFisher). Briefly, the nuclei were detected using the Hoechst dye and the dead cell nuclei showing a costaining with EthD-2 were excluded from the analysis. Then, the dots of NBD-PE were detected in the cytoplasm of each cell corresponding to a dilation of the nuclei to a maximum of 50 µm width and the total intensity of NBD-PE was

measured in each cell for quantification of DIPL. After imaging, cytotoxicity was assessed doing an ATP quantification with the CellTiter-Glo® 2.0 Cell Viability assay kit (Promega, ref. G9241) following manufacturer's instructions.

NBD-PE and Spike protein staining in infected cells. 15000 A549-ACE2 cells per well were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO_2. Two hours prior to infection, the media was replaced with 100 µL of DMEM (2% FBS) containing the compound of interest at the indicated concentration and 7.5 µM NBD-PE dye (Invitrogen, N360). At the time of infection, the media was replaced with virus inoculum (MOI 2 PFU/cell) and incubated for one hour at 37°C, 5% CO_2. Following the adsorption period, the inoculum was removed, replaced with 100 µL of drug and NBD-PE-containing media, and cells incubated for an additional 24 hours at 37°C, 5% CO_2. At this point, the supernatant was replaced with DMEM containing 10 µg/mL of Hoechst (Invitrogen, H3570) and 2 µM Ethidium homodimer-2 (Invitrogen, E3599) and incubated for 20 min at 37°C, 5% CO_2. The cells were then washed once with HBSS+/+ and fixed with 4% (v/v) formalin in PBS. The plates were then imaged using the Opera Phenix High Content Screening System, taking 12 images per well with a 20x objective. Subsequently, the cells were washed and stained for Spike antigen, to identify infected cells. Briefly, the cells were permeabilized with Triton 0.1% for 10 minutes at room temperature and nonspecific staining was blocked with PBS 5% Bovine Serum Albumin (BSA) for two hours at room temperature. The cells were then stained with 1ug/mL of the anti-Spike human monoclonal antibody mAb48 recognizing the RBD[57] (kindly provided by the lab of Hugo Mouquet, hugo.mouquet@pasteur.fr) overnight at 4°C and with a goat

anti-human secondary antibody AlexaFluor647 for one hour at room temperature. Upon staining, the cells were imaged once more in the same fields of view using the Opera Phenix screening system.

For image analysis the Columbus image analysis (PerkinElmer) system was used. Nuclei touching the border of the image were rejected. Living cells (not stained with Ethidium homodimer 2) were identified and the total intensity of NBD-PE dots in a 50 μm radius circle centered on the nucleus of living cells was measured, using the spot detection algorithm. To quantify SARS-CoV-2 infection, the area of Spike+ staining was quantified and normalized by the number of nuclei using the same software. Two-way ANOVAs were performed on pooled data from three independent experiments each in biological triplicate (**Fig. 2.S9**) using GraphPad Prism version 8.0.0 (San Diego, CA).

General Chemical Synthesis Procedure. Anhydrous solvents were purchased from Acros Organics. Unless specified below, all chemical reagents were purchased from Sigma-Aldrich and AK Scientific. Commercial solvents and reagents were used as received. All reactions were performed in oven-dried glassware fitted with rubber septa under a positive pressure of argon, unless otherwise noted. Air- and moisture-sensitive liquids were transferred via syringe. Solutions were concentrated by rotary evaporation at or below 40 °C. Analytical thin-layer chromatography (TLC) was performed using glass plates pre-coated with silica gel (0.25-mm, 60-Å pore size, 230−400 mesh, Merck KGA) impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by exposure to ultraviolet light (UV), then were stained by submersion in a 10% solution of

phosphomolybdic acid (PMA) in ethanol or an acidic ethanolic solution of p-anisaldehyde followed by brief heating on a hot plate. The latter solution was prepared by sequential additions of concentrated sulfuric acid (5.0 mL), glacial acetic acid (1.5 mL) and panisaldehyde (3.7 mL) to absolute ethanol (135 mL) at 23 °C with efficient stirring. Flash column chromatography was performed with Teledyne ISCO CombiFlash EZ Prep chromatography system, employing pre-packed silica gel cartridges (Teledyne ISCO RediSep).

Instrumentation. Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on Bruker AvanceIII HD 2-channel instrument (400 MHz/100 MHz) at 23 °C. Proton chemical shifts are expressed in parts per million (ppm, δ scale) and are referenced to residual protium in the NMR solvent (CHCl$_3$: δ 7.26, D$_2$HSOCD$_3$: δ 2.50). Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, dt = doublet of triplets, m = multiplet, br = broad, app = apparent), integration, and coupling constant (J) in Hertz (Hz). High-resolution mass spectra were obtained using a Waters Xevo G2-XS time-of-flight mass spectrometer.

72
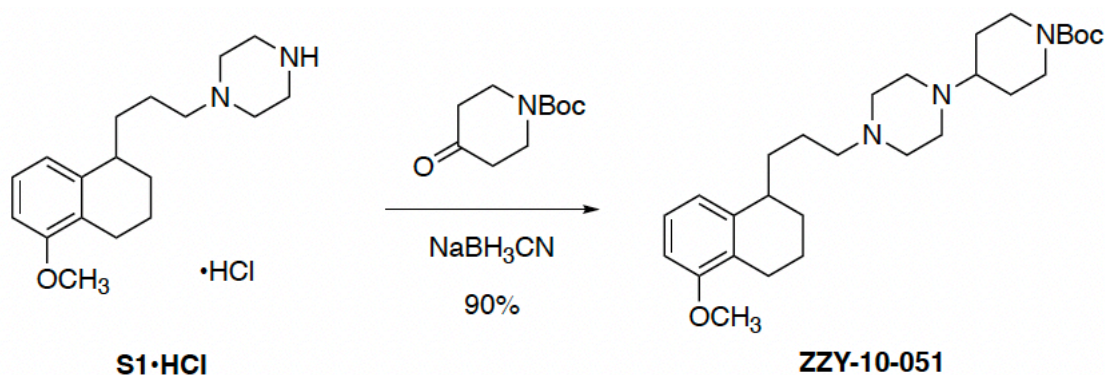
ZZY-10-051. N-Boc-4-piperidone (36.8 mg, 0.185 mmol) and sodium cyanoborohydride (11.6 mg, 0.185 mmol) were added sequentially to a stirred suspension of 1-[3-(5-methoxytetralin-1yl)propyl]piperazine hydrochloride[58] (S1•HCl, 20.0 mg, 0.0616 mmol) in methanol (0.50 mL). The mixture was stirred at 23 ºC and the reaction progress was monitored by LC-MS. After 24 h, the reaction mixture was partitioned between saturated aqueous sodium bicarbonate solution (5 mL) and dichloromethane (5 mL). The layers were separated, and the aqueous layer was extracted with dichloromethane (2 x 5 mL). The combined organic layers were dried over sodium sulfate. The dried solution was filtered, and the filtrate was concentrated. The residue was purified by column chromatography (0–10% methanol–dichloromethane, 4-g RediSep(R) Rf column, Teledyne ISCO, Lincoln, NE) to afford the product as a white foam (26.0 mg, 90%).

^{1}H NMR (400 MHz, CDCl3) δ 7.08 (t, J = 7.9 Hz, 1H), 6.77 (d, J = 7.7 Hz, 1H), 6.64 (dd, J = 8.2, 1.1 Hz, 1H), 4.13 (s, 2H), 3.79 (s, 3H), 2.81 – 2.48 (m, 12H), 2.48 – 2.32 (m, 4H), 1.89 – 1.51 (m, 10H), 1.44 (s, 9H), 1.42 – 1.33 (m, 2H).

HRMS (ESI$^+$): Calculated for $[C_{28}H_{45}N_3O_3+H]^+$: 472.3539. Found: 472.3518.
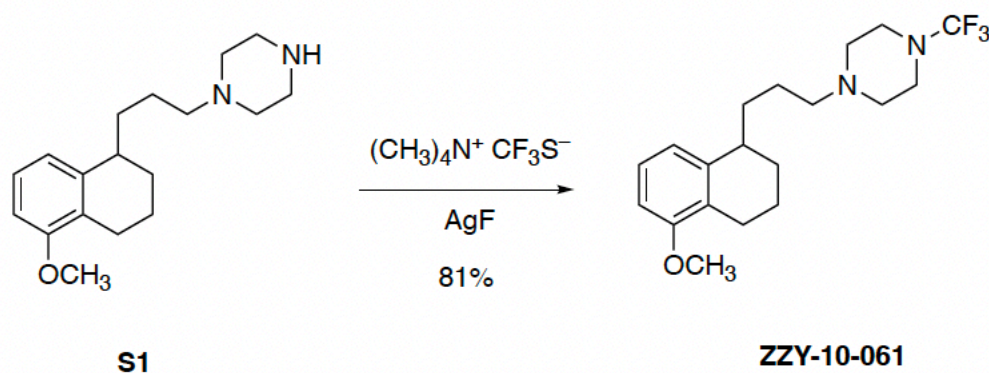
S1   ZZY-10-061

ZZY-10-061. An oven-dried one-dram vial was charged with 1-[3-(5-methoxytetralin-1-yl)propyl]piperazine (*58*) (S1, 40.0 mg, 0.139 mmol), acetonitrile (0.35 mL) and a magnetic stir bar. The mixture was sonicated until an even suspension was

formed. The solids did not fully dissolve. Tetramethylammonium trifluoromethanethiolate (31.6 mg, 0.180 mmol) was added in one portion, and the mixture was stirred for 15 min at 23 ºC. Silver(I) fluoride (52.8 mg, 0.416 mmol) was then added in a single portion, and the mixture was warmed to 50 ºC and stirred at that temperature for 4 h. At this point, TLC analysis showed complete conversion. The reaction mixture was diluted with ether (10 mL) and filtered through a cotton plug. The filtrate was directly concentrated to afford the product as pale white powder (40.0 mg, 81%).

^{1}H NMR (400 MHz, DMSO-d6) δ 7.10 (t, J = 7.9 Hz, 1H), 6.79 (d, J = 7.7 Hz, 1H), 6.74 (d, J = 8.1 Hz, 1H), 3.97 (d, J = 14.2 Hz, 1H), 3.62 – 3.34 (m, 7H), 3.23 – 2.93 (m, 4H), 2.75 (s, 1H), 1.84 – 1.46 (m, 9H).

HRMS (ESI+): Calculated for $[C_{19}H_{27}F_3N_2O+H]+$: 357.2154. Found: 357.2162.

Correlation analyses. Compound-by-compound phospholipidosis versus SARS-CoV-2 infection correlation analyses were performed using the "correlation" function in GraphPad Prism using raw RT-qPCR values normalized to DMSO from the same experiment at 100%. To model the pooled correlation we used Bayesian inference through the brms R package v2.14.4[59] the log PLD. With weakly informative priors, fitting 107 drug/concentration pairs yielded posterior parameter means and 95% credible interval (95%CI) estimates of IC$_{50}$: 43 [38, 48]%, hill: -5.6 [-7.0, -4.5], and Sigma 2.0 [0.14, 1.78]. Bayesian leave-one-out R^2 values were computed by using the loo package v2.4.1[60]. Analysis and accompanying raw data can be found at https://github.com/momeara/DIPL_SARS-CoV-2.

Literature search for SARS-CoV-2 repurposed antivirals. Twelve major phenotypic antiviral repurposing papers were sourced from the literature [3,8,14–16,24–29]. Drug names were selected from each paper based on the author-reported number of in vitro hits at the most strict reported threshold. If an explicit mention of the number of hits was not mentioned, all molecules with demonstrated dose-response inhibition of SARS-CoV-2 were selected. SMILES for each compound were retrieved from PubChem[61] using the PubChemPy API (https://pubchempy.readthedocs.io). Using the PubChem canonical SMILES, each molecule's cLogP was calculated using RDkit-2019.09.3.0 (http://www.rdkit.org), and JChem's cxcalc command line tool was used to calculate each molecule's most basic pKa, JChem-15.11.23.0, ChemAxon (https://www.chemaxon.com). Molecules with cLogP ≥ 3 and pKa ≥ 7.4 were considered CADs. ECFP-4-based Tanimoto coefficients (Tcs) were calculated for the list of CADs to all molecules from a list of known phospholipidosis inducers [Citation error], and the maximum Tc for each CAD to a known phospholipidosis inducer was used for filtering the CAD list for known, and highly likely, phospholipidosis inducers (Tc ≥ 0.4).

Pharmacokinetics. Pharmacokinetic experiments were performed by Bienta (Enamine Biology Services) in accordance with Enamine pharmacokinetic study protocols and Institutional Animal Care and Use Guidelines (protocol number 1-2/2020). Plasma pharmacokinetics and lung distribution for amiodarone, sertraline, PB28, tamoxifen, and elacridar were investigated following five intraperitoneal (i.p.) doses of each drug in 21 male mice per drug condition plus 1 mouse per vehicle condition (106 mice per drug experiment total). CD-1 mice were used for PB28, amiodarone, elacridar,

and sertraline studies and C57BL/6 mice were used for tamoxifen studies. The formulations for each compound was as follows: amiodarone- PG-PEG400 (80%:20%); sertraline- DMA-PEG400physiological saline (20%:20%:60%); PB28- PG-PEG400 (80%:20%); tamoxifen- corn oil (100%); Elacridar- DMA-PEG400-2HPβCD-water (25%:25%:25%:25%). Mice were injected i.p. with 2,2,2tribromoethanol at the dose of 150 mg/kg prior to drawing blood. Blood collection was performed from the orbital sinus in microtainers containing K2EDTA at 0, 5, 15, 30, 60, 120, 360, and 1440 minutes after drug injection. Animals were sacrificed by cervical dislocation after the blood samples collection. After this, lung samples were collected and weighted. All samples were immediately processed, flashfrozen and stored at -70°C until subsequent analysis.

Plasma samples (50 μL) were mixed with 200 μL of internal standard (IS) solution. After mixing by pipetting and centrifuging for 4 min at 6000 rpm, 0.5 μL of each supernatant was injected into the LC-MS/MS system. Solution of each compound (400 ng/ml in acetonitrile-methanol mixture, 1:1, v/v) was used as IS for drug quantification in plasma samples. Lung samples were dispersed in 3.5 volumes of IS400(90) using stainless steel beads (115 mg ± 5 mg) in The Bullet Blender® homogenizer for 90 sec at speed 10. After this, the samples were centrifuged for 4 min at 14,000 rpm, and 0.5 μL of each supernatant was injected into the LC-MS/MS system. Solution of a reference compound (sertraline or amiodarone; 400 ng/ml in water-methanol mixture, 1:9, v/v) was used as IS for quantification of the drugs in lung samples.

Analyses of plasma and lung samples were conducted at Enamine/Bienta. The

concentrations of drugs in plasma and lung samples were determined using high performance liquid chromatography/tandem mass spectrometry (HPLC-MS/MS). The Shimadzu HPLC system used consists of two isocratic pumps LC-10ADvp, an autosampler SIL-30AC MP, a sub-controller FCV20AHs and a degasser DGU-14A. Mass spectrometric analysis was performed using a 4000 Q TRAP instrument from MDS Sciex (Canada) with an electro-spray (ESI) interface. The data acquisition and system control was performed using Analyst 1.6.3 software from AB Sciex. The concentrations of the test compound below the lower limit of quantitation (LLOQ, amiodarone: 5 ng/mL for plasma and 7 ng/g for lung; sertraline: 10 ng/mL for plasma and 17.5 ng/g for lung; PB28: 5 ng/mL for plasma and 20 ng/g for lung; tamoxifen: 10 ng/mL for plasma and 8 ng/g for lung; elacridar: 5 ng/mL for plasma and 35 ng/g for lung) were designated as zero. The pharmacokinetic data analysis was performed using noncompartmental, bolus injection or extravascular input analysis models in WinNonlin 5.2 (PharSight). Data below LLOQ were presented as missing to improve validity of T½ calculations.

3-day dosing animal models of SARS-CoV-2 infection experiments. All the antiviral animal studies were performed in an animal biosafety level 3 (BSL3) facility at the Icahn school of Medicine in Mount Sinai Hospital, New York City. All work was conducted under protocols approved by the Mt. Sinai Institutional Animal Care and Use Committee (IACUC). We used a model of BALB/c mice transduced intranasally with an adenovirus expressing human ACE2 (hACE2) for all acute dosing experiments as described previously[62]. We used 29 female 12-week old specific pathogen–free BALB/c mice (the Jackson laboratory strain 000651). Five days prior to infection with SARS-CoV-

2, BALB/c mice were infected intranasally with 2.5×10^8 PFU of an adenovirus carrying the gene for hACE2. Viral seed stocks for non-replicating E1/E3 deleted viral vectors based on human adenovirus type-5 expressing the human angiotensin-converting enzyme 2 (Ad-ACE2) receptor under the control of a CMV promoter, were obtained from the Iowa Viral Vector Core Facility. Viral stocks were amplified to high titers following infection of T-Rex TM-293 cells and purification using two sequential rounds of cesium chloride (CsCl) ultracentrifugation, as described previously[63,64]. The infectious titer was determined using a tissue culture infectious dose-50 ($TCID_{50}$) end-point dilution assay, and physical particle titer quantified by micro-bicinchoninic acid (microBCA) protein assay, both described previously[63]. Remdesivir was administered subcutaneously (s.c.) while the other five remaining drugs were injected intraperitoneally (i.p.), with amiodarone and elacridar dosed once per day and remdesivir, PB28, tamoxifen, and sertraline being dosed twice per day, for a total of 3 days, consistent with their pharmacokinetic profiles. We administered vehicle (PBS, i.p., twice-daily) or drug treatments, two hours before intranasal infection with 1×10^4 PFU of SARS-CoV-2 in 50 µL of PBS. Mice were anesthetized with a mixture of ketamine/xylazine before each intranasal infection. Three days post infection (dpi) animals were humanely euthanized. Whole left lungs were harvested and homogenized in PBS with silica glass beads then frozen at −80°C for viral titration via $TCID_{50}$. Briefly, infectious supernatants were collected at 48 hours post infection and frozen at −80°C until later use. Infectious titers were quantified by limiting dilution titration using Vero E6 cells. Briefly, Vero E6 cells were seeded in 96well plates at 20,000 cells/well. The next day, SARS-CoV-2-containing supernatant was applied at serial 10-fold dilutions ranging from 10^{-1} to 10^{-6} and, after 5 days, viral cytopathic effect

(CPE) was detected by staining cell monolayers with crystal violet. $TCID_{50}$/mL were calculated using the method of Reed and Muench. GraphPad Prism version 8.0.0 (San Diego, CA) was used to determine differences in lung titers between treatments and vehicle using the Kruskal-Wallis test with Dunn's multiple comparison correction.

15-day dosing animal models of SARS-CoV-2 infection experiments. All the antiviral animal studies were performed in animal biosafety level 3 (BSL3) facility at the Icahn school of Medicine in Mount Sinai Hospital, New York City. All work was conducted under protocols approved by the Institutional Animal Care and Use Committee (IACUC). SARS-CoV-2 isolate USA-WA1/2020 (BEI resources; NR-52281), referred in this report as WT-SARS-CoV-2, was used to challenge mice intranasally. A variant of virus (termed MA- SARS-CoV-2) was obtained after series of passaging in different backgrounds of laboratory mice as well as mACE-2 expressing VeroE6 cells[65]. Briefly, the virus was serially passaged every 2 days via intranasal inoculation of the virus in 50 μL volume derived from the spun-down supernatants of lung homogenates. The mouse adaptation of the SARS-CoV-2 variant was studied in C57Bl6, BALB/c and 129S1/SVMJ (termed 129 for simplicity) mice models. Viral stocks were sequenced after propagation to verify the integrity of the viral genome with the mutations associated with mouse adaptation.

We utilized female 8-week-old specific pathogen–free 129 mice (the Jackson laboratory strain 002448). The amiodarone, vehicle, and remdesivir groups each had 6 mice. Dosing schemes were as follows; Group 1: 50 mg/kg intraperitoneal (i.p.) amiodarone, once per day (QD); Group 2: i.p. vehicle QD, Group 3: 50 mg/kg

subcutaneous (s.c.) remdesevir, twice per day (BID). Amiodarone and vehicle groups were dosed QD for 12 days prior to infection in order to induce phospholipidosis in the animals and dosing was continued through the course of the experiment. Remdesivir was administered starting 1 hour before intranasal infection with 2.5×10^4 PFU of MA-SARS-CoV-2 in 50 µL of PBS. Mice were anesthetized with a mixture of ketamine/xylazine before each intranasal infection. Mice were weighed daily for signs of pathogenesis. 3 days post infection (dpi) animals were humanely euthanized. The right whole lung of each mouse was harvested and preserved in 10% formalin for later histopathology analysis. The left whole lung of each mouse was harvested and homogenized in PBS with silica glass beads then frozen at −80°C for viral titration via TCID50. Infectious titers were quantified by limiting dilution titration using Vero E6 cells. Briefly, Vero E6 cells were seeded in 96-well plates at 20,000 cells/well. The next day, SARS-CoV-2-containing supernatant was applied at serial 10-fold dilutions ranging from 10^{-1} to 10^{-6} and, after 5 days, viral cytopathic effect (CPE) was detected by staining cell monolayers with crystal violet. TCID50/ml were calculated using the method of Reed and Muench. The Prism software (GraphPad) was used to determine differences in lung titers using 2-way ANOVA on log-transformed data. One mouse from the amiodarone group died prior to SARS-CoV-2 infection, leaving only 5 mice in Group 1 for the statistical analyses.

129 mice infected with 2.5×10^4 MA-SARS-CoV-2/WA, treated one daily (QD) or twice-daily (BID) with the indicated drugs, had lungs harvested on day 3 post-infection for histopathology analysis. Paraffin-embedded lung tissue blocks for mouse lungs were cut into 5 µm sections. Sections were stained with hematoxylin and eosin (H&E) and

analyzed by HistoWiz Inc. (histowiz.com, Brooklyn, NY). Digital light microscopic scans of whole lung processed in toto were examined by an experienced veterinary pathologist. H&E stained sections of lung from 129 mice were examined by implementing a semi quantitative, 5-point grading scheme (0 - within normal limits, 1 - mild, 2 - moderate, 3 - marked, 4 - severe) that took into account four different histopathological parameters: 1) perivascular inflammation 2) bronchial or bronchiolar epithelial degeneration or necrosis 3) bronchial or bronchiolar inflammation and 4) alveolar inflammation. These changes were absent (grade 0) in lungs from uninfected mice from groups that were utilized for this assessment. All mice from amiodarone, remdesivir, and vehicle treated infected groups exhibited multifocal pulmonary lesions.

COVID-19 clinical trial expenditure analysis. Data for COVID-19 clinical trials were downloaded from the DrugBank[49] COVID-19 dashboard (accessed on 2021-02-16). Supplemental information for each compound, including its SMILES and DrugBank ID was downloaded and unpacked from the DrugBank "All drugs" 2021-01-03 release[66]. A total of 3395 treatments were recorded in the COVID-19 dashboard at the time of download, and 2244 of those annotations were for small molecules from a total of 1490 unique clinical trials worldwide. Using the DrugBank annotated SMILES, each molecule's cLogP was calculated using RDkit-2019.09.3.0 (http://www.rdkit.org), and JChem's cxcalc command line tool was used to calculate each molecule's most basic pKa, JChem-15.11.23.0, ChemAxon (https://www.chemaxon.com). Molecules with cLogP ≥ 3 and pKa ≥ 7.4 were considered CADs. Azithromycin, a non-CAD phospholipidosis inducer, was also included in the CAD dataset due to its shared mechanism of action with CAD

antivirals whereas fluvoxamine, a CAD known to act through immune-mediated processes was excluded from the CAD dataset. Additional subsets of these data, including filtering out chloroquine and hydroxychloroquine trials from the CADs, were also analyzed. To quantify the estimated cost of clinical trials, the molecules from a given subset were first filtered for unique clinical trial IDs, then were grouped based on clinical trial phase and multiplied by the average cost of an anti-infective clinical trial in that phase[50,51]. Mixed-phase trials were multiplied by the cost of the more advanced phase only.

2.10 References

1. Kuleshov, M. V. *et al.* The COVID-19 Drug and Gene Set Library. *Patterns* **1**, 100090 (2020).

2. Edwards, A. What Are the Odds of Finding a COVID-19 Drug from a Lab Repurposing Screen? *J. Chem. Inf. Model* **60**, 5727–5729 (2020).

3. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

4. Dahlin, J. L. *et al.* Nuisance compounds in cellular assays. *Cell Chem. Biol.* **28**, 356–370 (2021).

5. Breiden, B. & Sandhoff, K. Emerging mechanisms of drug-induced phospholipidosis. *Biol. Chem.* **401**, 31–46 (2019).

6. Shayman, J. A. & Abe, A. Drug induced phospholipidosis: An acquired lysosomal storage disorder. *Biochimica Et Biophysica Acta Bba - Mol. Cell Biology Lipids* **1831**, 602–611 (2013).

7. Abu-Farha, M. *et al.* The Role of Lipid Metabolism in COVID-19 Virus Infection and as a Drug Target. *Int. J. Mol. Sci.* **21**, 3544 (2020).

8. Müller, C. *et al.* Inhibition of Cytosolic Phospholipase A 2 α Impairs an Early Step of Coronavirus Replication in Cell Culture. *J. Virol.* **92**, e01463-17 (2018).

9. Salata, C., Calistri, A., Parolin, C., Baritussio, A. & Palù, G. Antiviral activity of cationic amphiphilic drugs. *Expert Rev. Anti-infe.* **15**, 483–492 (2017).

10. Gunesch, A. P. *et al.* Filovirus Antiviral Activity of Cationic Amphiphilic Drugs Is Associated with Lipophilicity and Ability To Induce Phospholipidosis. *Antimicrob. Agents Ch.* **64**, (2020).

11. Guigui, B. *et al.* Amiodarone-induced hepatic phospholipidosis: A morphological alteration independent of pseudoalcoholic liver disease. *Hepatology* **8**, 1063–1068 (1988).

12. Linthorst, G. E. & Hollak, C. E. M. Chloroquine-induced phospholipidosis of the kidney mimicking Fabry's disease. *Hum. Pathol.* **34**, 1358–1359 (2003).

13. FITZHUGH, O. G., NELSON, A. A. & HOLLAND, O. L. The chronic oral toxicity of chloroquine. *J. Pharmacol. Exp. Ther.* **93**, 147–52 (1948).

14. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020).

15. Chen, C. Z. *et al.* Drug Repurposing Screen for Compounds Inhibiting the Cytopathic Effect of SARS-CoV-2. *Front. Pharmacol.* **11**, 592737 (2021).

16. Weston, S. *et al.* Broad Anti-coronavirus Activity of Food and Drug Administration-Approved Drugs against SARS-CoV-2 In Vitro and SARS-CoV In Vivo. *J. Virol.* **94**, e01218-20 (2020).

17. Plaze, M. *et al.* Inhibition of the replication of SARS-CoV-2 in human cells by the FDA-approved drug chlorpromazine. *Int. J. Antimicrob. Ag.* **57**, 106274 (2021).

18. Muehlbacher, M., Tripal, P., Roas, F. & Kornhuber, J. Identification of Drugs Inducing Phospholipidosis by Novel in vitro Data. *ChemMedChem* **7**, 1925–1934 (2012).

19. Obeidat, M., Isaacson, A. L., Chen, S. J., Ivanovic, M. & Holanda, D. Zebra-like bodies in COVID-19: is phospholipidosis evidence of hydroxychloroquine induced acute kidney injury? *Ultrastruct. Pathol.* **44**, 519–523 (2020).

20. Morelli, J. K. *et al.* Validation of an in vitro screen for phospholipidosis using a high-content biology platform. *Cell Biol. Toxicol.* **22**, 15–27 (2006).

21. Shahane, S. A. *et al.* Detection of Phospholipidosis Induction. *J. Biomol. Screen* **19**, 66–76 (2013).

22. Dittmar, M. *et al.* Drug repurposing screens reveal cell-type-specific entry pathways and FDA-approved drugs active against SARS-Cov-2. *Cell Reports* **35**, 108959 (2021).

23. Baronas, E. T., Lee, J.-W., Alden, C. & Hsieh, F. Y. Biomarkers to monitor drug-induced phospholipidosis. *Toxicol. Appl. Pharm.* **218**, 72–78 (2007).

24. Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).

25. Bakowski, M. A. *et al.* Oral drug repositioning candidates and synergistic remdesivir combinations for the prophylaxis and treatment of COVID-19. *Biorxiv* 2020.06.16.153403 (2020) doi:10.1101/2020.06.16.153403.

26. Drayman, N. *et al.* Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* **373**, 931–936 (2021).

27. Mirabelli, C. *et al.* Morphological cell profiling of SARS-CoV-2 infection identifies drug repurposing candidates for COVID-19. *Proc. National Acad. Sci.* **118**, e2105815118 (2021).

28. Touret, F. *et al.* In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci. Rep.* **10**, 13093 (2020).

29. Xiao, X. *et al.* Identification of Potent and Safe Antiviral Therapeutic Candidates Against SARS-CoV-2. *Front. Immunol.* **11**, 586572 (2020).

30. Jeon, S. *et al.* Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob. Agents Ch.* **64**, e00819-20 (2020).

31. Lowe, R., Glen, R. C. & Mitchell, J. B. O. Predicting Phospholipidosis Using Machine Learning. *Mol. Pharmaceut.* **7**, 1708–1714 (2010).

32. Ploemen, J.-P. H. T. M. *et al.* Use of physicochemical calculation of pKa and CLogP to predict phospholipidosis-inducing potential A case study with structurally related piperazines. *Exp. Toxicol. Pathol.* **55**, 347–355 (2004).

33. Dyall, J. *et al.* Repurposing of Clinically Developed Drugs for Treatment of Middle East Respiratory Syndrome Coronavirus Infection. *Antimicrob. Agents Ch.* **58**, 4885–4893 (2014).

34. Madrid, P. B. *et al.* Evaluation of Ebola Virus Inhibitors for Drug Repurposing. *Acs Infect. Dis.* **1**, 317–326 (2015).

35. Johansen, L. M. *et al.* A screen of approved drugs and molecular probes identifies therapeutics with anti–Ebola virus activity. *Sci. Transl. Med.* **7**, 290ra89 (2015).

36. Cheng, H. *et al.* Inhibition of Ebola and Marburg Virus Entry by G Protein-Coupled Receptor Antagonists. *J. Virol.* **89**, 9932–9938 (2015).

37. Gehring, G. *et al.* The clinically approved drugs amiodarone, dronedarone and verapamil inhibit filovirus cell entry. *J. Antimicrob. Chemoth.* **69**, 2123–2131 (2014).

38. He, S. *et al.* Repurposing of the antihistamine chlorcyclizine and related compounds for treatment of hepatitis C virus infection. *Sci. Transl. Med.* **7**, 282ra49 (2015).

39. Poh, M. K. *et al.* U18666A, an intra-cellular cholesterol transport inhibitor, inhibits dengue virus entry and replication. *Antivir. Res.* **93**, 191–198 (2012).

40. White, K. M. *et al.* Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A. *Science* **371**, 926–931 (2021).

41. Wolff, G., Melia, C. E., Snijder, E. J. & Bárcena, M. Double-Membrane Vesicles as Platforms for Viral Replication. *Trends Microbiol.* **28**, 1022–1033 (2020).

42. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).

43. Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395–1398 (2020).

44. Holland, L. K. K., Nielsen, I. Ø., Maeda, K. & Jäättelä, M. SnapShot: Lysosomal Functions. *Cell* **181**, 748-748.e1 (2020).

45. Glebov, O. O. Understanding SARS-CoV-2 endocytosis for COVID-19 drug repurposing. *Febs J* **287**, 3664–3671 (2020).

46. Hamaguchi, R., Haginaka, J., Tanimoto, T. & Kuroda, Y. Maintenance of luminal pH and protease activity in lysosomes/late endosomes by vacuolar ATPase in chlorpromazine-treated RAW264 cells accumulating phospholipids. *Cell Biol. Toxicol.* **30**, 67–77 (2014).

47. Norinder, U., Tuck, A., Norgren, K. & Kos, V. M. Existing highly accumulating lysosomotropic drugs with potential for repurposing to target COVID-19. *Biomed. Pharmacother.* **130**, 110582 (2020).

48. Yang, N. & Shen, H.-M. Targeting the Endocytic Pathway and Autophagy Process as a Novel Therapeutic Strategy in COVID-19. *Int. J. Biol. Sci.* **16**, 1724–1731 (2020).

49. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).

50. Sertkaya, A., Birkenbach, A., Berlind, A., Eyraud, J. & Group, E. R. *Examination of clinical trial costs and barriers for drug development*. https://aspe.hhs.gov/reports/examination-clinical-trial-costs-barriers-drug-development-0 (2014).

51. Sertkaya, A., Wong, H.-H., Jessup, A. & Beleche, T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin. Trials* **13**, 117–126 (2016).

52. Group, R. C. *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *New Engl. J. Med.* **384**, 693–704 (2020).

53. Lenze, E. J. *et al.* Fluvoxamine vs Placebo and Clinical Deterioration in Outpatients With Symptomatic COVID-19. *Jama* **324**, 2292–2300 (2020).

54. Wingler, L. M., McMahon, C., Staus, D. P., Lefkowitz, R. J. & Kruse, A. C. Distinctive Activation Mechanism for Angiotensin Receptor Revealed by a Synthetic Nanobody. *Cell* **176**, 479-490.e12 (2019).

55. Buchrieser, J. *et al.* Syncytia formation by SARS-CoV-2-infected cells. *Embo J.* **39**, e106267 (2020).

56. Amanat, F. *et al.* An In Vitro Microneutralization Assay for SARS-CoV-2 Serology and Drug Screening. *Curr Protoc Microbiol* **58**, e108 (2020).

57. Planas, D. *et al.* Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat Med* **27**, 917–924 (2021).

58. Berardi, F. *et al.* Exploring the Importance of Piperazine N-Atoms for σ2 Receptor Affinity and Activity in a Series of Analogs of 1-Cyclohexyl-4-[3-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)propyl]piperazine (PB28). *J. Med. Chem.* **52**, 7817–7828 (2009).

59. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Soft.* **80**, 1–28 (2017).

60. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).

61. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, gkaa971- (2020).

62. Rathnasinghe, R. *et al.* Comparison of transgenic and adenovirus hACE2 mouse models for SARS-CoV-2 infection. *Emerg. Microbes Infec.* **9**, 2433–2445 (2020).

63. Coughlan, L. *et al.* Ad5:Ad48 Hexon Hypervariable Region Substitutions Lead to Toxicity and Increased Inflammatory Responses Following Intravenous Delivery. *Mol. Ther.* **20**, 2268–2281 (2012).

64. Coughlan, L. *et al.* In Vivo Retargeting of Adenovirus Type 5 to αvβ6 Integrin Results in Reduced Hepatotoxicity and Improved Tumor Uptake following Systemic Delivery. *J. Virol.* **83**, 6416–6428 (2009).

65. Rathnasinghe, R. *et al.* The N501Y mutation in SARS-CoV-2 spike leads to morbidity in obese and aged mice and is neutralized by convalescent and post-vaccination human sera. *Medrxiv* 2021.01.19.21249592 (2021) doi:10.1101/2021.01.19.21249592.

66. Himmelstein, D. S. *User-friendly extensions of the DrugBank database v1.0*. (2016).

Gloss to Chapter 3

While the phospholipidosis project focused on ways cell-based antiviral repurposing screens were confounded by certain types of drugs, we began to wonder how common it was for there to be confounding compounds coming out of other types of drug repurposing COVID-19 projects, namely biochemical screens. In this chapter, I focus on another artifact common in early drug discovery— colloidal aggregation— and how drugs that induce colloidal aggregation led to false-positive hits in biochemical repurposing screens. Further, we demonstrate that by using physicochemical property filters on screening libraries, a large percentage of molecules have the potential to confound biochemical screens due to their inherent aggregation phenotypes, highlighting the importance of controlling for colloidal aggregation in these projects.

Although this project ultimately became a stand-alone paper, there was a time when it was combined with the phospholipidosis project as a more general paper about screening confounds. Ultimately, the individual stories benefitted from being told separately, but the lessons that each paper teaches remain similar. First, when screening drugs at high concentrations regardless of which assay you use, you need to be careful that the compounds are not working through confounding mechanisms. Second, just because something is an FDA-approved drug does not mean it is free from unfavorable properties or behaviors. Lastly, do not bet against Brian's sniff test. If he thinks something quacks like a duck, it probably is a duck, and you should be careful to not call it a swan until you have corroborating evidence.

Chapter 3: Colloidal aggregators in biochemical SARS-CoV-2

repurposing screens

Colloidal aggregators in biochemical SARS-CoV-2 repurposing screens

Henry R. O'Donnell[1], Tia A. Tummino[1,2,3], Conner Bardine[4], Charles S. Craik[1,3], Brian K. Shoichet[1,3]*

* Corresponding author: bshoichet@gmail.com

[1]Department of Pharmaceutical Chemistry, University of California San Francisco (UCSF), San Francisco, CA, USA.

[2]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA.

[3]QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA.

[4]Graduate Program in Chemistry & Chemical Biology, UCSF, San Francisco, CA, USA.

3.1 Abstract

To fight COVID-19, much effort has been directed toward in vitro drug repurposing. Here, we investigate the impact of colloidal aggregation, a common screening artifact, in these repurposing campaigns. We tested 56 drugs reported as active in biochemical assays for aggregation by dynamic light scattering and by detergent-based enzyme counter screening; 19 formed colloids at concentrations similar to their literature IC_{50}'s, and another 13 were problematic. From a common repurposing library, we further selected another 15 drugs that had physical properties resembling known aggregators, finding that six aggregated at micromolar concentrations. This study suggests not only that many of the drugs repurposed for SARS-CoV-2 in biochemical assays are artifacts but that, more generally, at screening-relevant concentrations, even drugs can act artifactually via colloidal aggregation. Rapid detection of these artifacts will allow the community to focus on those molecules that genuinely have potential for treating COVID-19.

3.2 Introduction

Drug repurposing is an attractive idea in the face of a global pandemic, when rapid antiviral drug development is crucial. Although the historical pragmatism of this approach has drawn scrutiny[1,2], drug repurposing has the potential to dramatically cut both the time and cost needed to develop a new therapeutic.[3] Repurposing campaigns typically screen curated libraries of thousands of approved drugs and investigational new drugs (INDs), and several assays have been developed to test these libraries for activity against SARS-CoV-2.[4–6] Most high throughput, biochemical screens were developed to detect activity against two proteins that are used in viral infection and maturation: the human ACE-2 (angiotensin converting enzyme 2) and 3CL-Pro[7], the major polypeptide processing protease of SARS-2-CoV-2.

When testing molecules for biochemical activity at micromolar concentrations, it is important to control for artifacts[8–12] including colloidal aggregation, which is perhaps the single most common artifact in early drug discovery.[13,14] Drugs, though in many ways de-risked, are not immune to aggregation and artifactual behavior when screened at relevant concentrations[15,16] (though they are not expected to aggregate at on-target relevant concentrations). Knowing this, we wondered if colloidal aggregation was causing false positives in some COVID-19 drug repurposing studies, especially since several known aggregators, such as manidipine and methylene blue, were reported as apparently potent hits for COVID-19 targets.[17,18]

Aggregation is a common source of false positives in early drug discovery,[19] arising

from spontaneous formation of colloidal particles when organic, drug-like molecules are introduced into aqueous media.[15,16,19,20] The resulting liquid particles are densely packed spheres[21] that promiscuously inhibit proteins by sequestering them on the colloid surface[22], where they suffer partial unfolding.[23] The resulting inhibition is reversible by disruption of the colloid and is characterized by an incubation effect on an order of several minutes due to enzyme crowding on the surface of the particle.[24] Colloids often can be disrupted by the addition of small amounts, often sub-critical micelle concentrations, of non-ionic detergent such as Triton-X–100.[25] Accordingly, addition of detergent is a common perturbation to rapidly detect aggregates in counter screens against model enzymes such as AmpC β-lactamase or malate dehydrogenase (MDH). Aggregation can be physically detected by biophysical techniques such as nuclear magnetic resonance (NMR)[26] and by dynamic light scattering (DLS), as the colloids typically form particles in the 50 to 500 nm radius size range, which is well suited to measurement by this latter technique.

Here, we investigate the role of colloidal aggregation as a source of false positives in drug repurposing studies for SARS-CoV-2 targets. We focused on *in vitro*, ACE2 and 3CL-Pro screens since these are relevant for aggregation. We searched the literature and compiled hits from 12 sources[18,27–37] where drug activities were in the micromolar and sub-micromolar range typical of colloidal aggregation. Drugs with cLogP values over 3.0 (most of those selected) or with conjugated ring systems conducive to stacking, such as methylene blue, chiniofon, and theaflavin (most of the remaining), were prioritized for testing. How the results of this study may impact the design of future

repurposing screens both for SARS-2 and for other indicators, will be considered.

3.3 Results

Colloidal aggregators are common hits in drug repurposing screens for SARS-CoV-2. We tested 56 drugs for colloidal aggregation, which had been reported to be active in biochemical repurposing screens against SARS-CoV-2[18,27–30,32,38] (**Table 3.S1** and **Methods** for a description of the literature search). In short, the 2D structures of compounds with reported activities in the micromolar range typical of colloidal aggregation were visually inspected for molecular features in known aggregators (e.g., multiple conjugated ring systems or calculated LogP (cLogP) >3). Five criteria were used to investigate whether reported hits formed colloidal aggregates: (a) particle formation indicated by scattering intensity, (b) clear autocorrelation curves, (c) an MDH IC_{50} value in the micromolar—high nanomolar range, (d) restoration of MDH activity with the addition of detergent, and less stringently (e) high Hill slopes in the inhibition concentration response curves (**Fig. 3.1**).

Using the literature reported IC_{50} for the repurposed drugs as a starting point, we tested each drug for MDH inhibition and calculated the IC_{50} and Hill slope. We used IC_{50} values from the MDH concentration response curves and tested for detergent sensitivity at three-fold the MDH IC_{50} (**Fig. 3.2**). Next, we calculated the critical aggregation concentration (CAC) by measuring normalized scattering intensity on the DLS; any point above 1×10^6 was considered from the aggregated form. By plotting a best fit line for aggregating concentrations and nonaggregating concentrations, the CAC was given by

the point of intersection (**Fig. 3.3**). We also measured the DLS auto-correlation curve as a criterion: if this was well formed, it gave further confidence (**Fig. 3.S1**).

Nineteen molecules formed well-behaved particles by DLS with clean autocorrelation curves, and inhibited MDH in the absence of, but not the presence of, 0.01% Triton X-100; these seem to be clear colloidal aggregators (**Table 3.1** and **Fig. 3.2 and 3.3**). Both DLS-based critical aggregation concentrations and MDH IC_{50} values were in the range of the IC_{50}'s reported in the literature against the two SARS-CoV-2 enzymes; indeed, molecules like gossypol, manidipine, and TTNPB inhibited MDH even more potently than they did either ACE2 or 3CLPro. For most of the 19 drugs, the Hill slopes were high, though for several clear aggregators, such as Hemin and Shikonin, they were only in the 1.3–1.4 range. The Hill slope depends on the ratio of enzyme concentration to true K_D and can vary from assay to assay[39] and from aggregator to aggregator[13]; while many consider it as a harbinger of aggregation, we take it as a soft criterion.[13] Finally, two molecules, Evans blue and TBB, did not show particles by DLS, perhaps for spectral reasons, but did pass the other four criteria. To investigate them further, we asked whether they could be precipitated by gentle centrifugation. We tested these molecules for MDH inhibition before and after centrifugation (**Fig. 3.S2**) and found that enzyme activity was restored after centrifugation. This suggests that these molecules are forming colloidal aggregates, which can be spun down unlike small molecules that are genuinely in solution.[22,23]

A characteristic example of a reported hit that is likely acting artifactually through colloidal aggregation is the calcium channel blocker lercanidipine, which has been reported to inhibit 3CL-Pro with an IC_{50} of 16.2 µM[18]. Lercanidipine satisfies our five criteria for aggregation: in aqueous buffer it forms particles that can be detected by a 10-fold increase in DLS scattering intensity (Cnts/sec), by a clearly defined autocorrelation curve in the DLS; it inhibits the counter-screening enzyme MDH with an IC_{50} of 2.2 µM, while MDH activity is restored upon addition of 0.01% Triton-X 100 detergent (**Fig. 3.1**). In the absence of detergent, lercanidipine inhibits MDH with a Hill slope of 2.9.

In addition to the 17 molecules that passed all five criteria for aggregation, another 19 molecules were more ambiguous, either forming particles by DLS but not inhibiting MDH or inhibiting MDH in a detergent-dependent manner but not forming particles detectable by DLS (**Table 3.S1**). These 19 drugs may also be acting artifactually; however, further investigation is needed to determine their exact mechanisms. For this study, we focused only on clear colloidal aggregators.

Molecules repurposed for 3CL-Pro show little activity against that enzyme in the presence of detergent. In addition to testing the repurposed molecules against a counter-screening enzyme like MDH, we also tested the 10 that had been repurposed against 3CL-Pro against that enzyme itself. Because 3CL-Pro is unstable in buffer without either the presence of detergent or substantial amounts of serum albumin—both of which disrupt colloids[22,40,41]—we could not investigate the impact of detergent with 3CL-Pro as we could do with MDH. Still, we could ask whether the drugs repurposed for 3CL-Pro inhibited the enzyme in the presence of 0.05% Tween-20 used to keep the

enzyme stable. Of the 12 drugs tested, only two had detectable potency below 200 μM in the presence of detergent, and for one of these two, 4E1RCat, their inhibition was reduced five-fold over its literature values (18.28 to 100µM) (**Table 3.2**, **Fig. 3.S3**). Only hemin continued to inhibit 3CL-Pro substantially, with an IC50 of 25 μM (but even this was 2.6-fold less potent than its literature value). As hemin's inhibition of MDH was disrupted by detergent (**Table 3.3**) and it formed clear particles by DLS (**Fig. 3.3**, **Fig. 3.S3**), we further tested it against the model counter-screening enzyme AmpC β-lactamase. Hemin inhibited AmpC with an IC_{50} of 23 μM; at 25 μM hemin, addition of 0.01% (v/v) Triton X-100 fully restored enzyme activity—inhibition was abolished. Taken together, these observations further support the aggregation-based activity of these 12 repurposed drugs.

Colloidal aggregators in repurposing libraries Target-based drug repurposing screens are common not only for SARS-CoV-2 but for many other viruses and indeed other indications. We thought it interesting to explore, if only preliminarily, the occurrence of colloidal aggregators in drug repurposing libraries. We prioritized drugs in the widely used SelleckChem FDA-approved library as potential aggregators, using a simple chemoinformatics approach.[42] Library molecules were compared to a database of known aggregators using the Aggregator Advisor[42] command line tool, which calculates molecular similarity (Tanimoto coefficients; Tc) between the two sets of molecules (**Table 3.S2**). Molecules similar to a known aggregator (1>Tc's>0.65) that were also hydrophobic (cLogP > 4) were drawn, inspected for diversity from one another and for the presence of features in known aggregators such as conjugated ring systems, and were prioritized for

testing. Of the 2336 unique drugs in the library, 73 are already known aggregators, and another 356 (16%) closely resemble known aggregators. We selected 15 of the latter for aggregation: six of these drugs satisfied our five criteria for aggregation; they inhibited MDH in the absence of, but not in the presence of, 0.01% Triton X-100 (**Fig. 3.4**) and formed well-behaved particles detectable by DLS (**Fig. 3.5**) with clean autocorrelation curves (**Fig. 3.S4**), often with steep Hill slopes. In aggregates, these data suggest that these six drugs are prone to colloidal aggregation at screening-relevant concentrations (**Table 3.2**).

3.4 Discussion

Two broad observations from this study merit emphasis. First, many drugs repurposed for COVID-19 aggregate and inhibit counter-screening enzymes promiscuously at concentrations relevant to their reported IC_{50}'s against the COVID-19 targets (ACE2 and 3CL-Pro). Of the 56 drugs tested, 17 fulfilled all five of our criteria for acting via colloidal aggregation: (i) they formed particles that were scattered strongly by DLS with (ii) well-behaved autocorrelation curves, (iii) they inhibited the counter-screening enzyme malate dehydrogenase—unrelated to either ACE2 or 3CL-Pro—at relevant concentrations in the absence, but (iv) not the presence, of detergent, and (v) they typically inhibited with steep Hill slopes. Each of these criteria individually is a harbinger of colloidal aggregation; when combined, they strongly support its occurrence. The other 19 of the 56 drugs fulfilled only some of these criteria, for instance, forming particles at relevant concentrations but not inhibiting MDH in a detergent-dependent manner. Some of these 19 drugs may also be aggregators, while others, like those that inhibit MDH but

cannot be reversed by detergent, like tannic acid, may be acting as pan assay interference compounds (PAINS). A second observation from this study is that these artifacts are not so much a feature of SARS-CoV-2 repurposing, but rather reflect the behavior of drugs at screening relevant concentrations. Thus, 6 of 15 drugs investigated from a general purposing library were also aggregators at micromolar concentrations. An attraction of drug repurposing is that the molecules are thought to be de-risked from the pathologies of early discovery. However, at micromolar concentrations, drugs, which are often larger and more hydrophobic than the lead-like molecules found in most high-throughput screening and virtual libraries, are if anything more likely to aggregate, something that earlier studies also support.[15,16]

For 4 of the 19 aggregators found in this study, Triton X-100 detergent was already used in the reaction buffer used in the original publication (**Table 3.S1**), reflecting the care of those studies. However, while it is commonly thought that detergent addition protects against aggregation from the outset, in fact, detergent often only right-shifts the onset of aggregation-based inhibition. Thus, even screens that control for aggregation by including detergent in the reaction buffer may consider +/− detergent controls during hit confirmation. On the other hand, several of the aggregators, including emodin, hemin, and hypericin (**Table 3.1**), notwithstanding their provenance from a drug repurposing library, have features that would ordinarily give medicinal chemists pause. Sometimes the "drugs" in drug repurposing libraries are not actually drugs, and despite their origins as phytochemical natural products, as with these molecules, they can have features, e.g., multiple phenolic groups in conjugated ring systems, that might prejudice them against

further study.

Certain caveats should be mentioned. We do not pretend to have undertaken a comprehensive study of the increasingly large literature around drug repurposing for COVID-19. The molecules tested here represent only a subset of those investigated, drawn from an analysis of some of the literature then available. Also, we have not demonstrated that aggregation is actually occurring in the ACE-2 assay itself, though the lack of inhibition of 3CLPro in the presence of detergent fortifies our conclusions for the 12 molecules that inhibited this enzyme. Finally, it is important to note that just because some repurposed drugs aggregate at micromolar concentrations, the repurposing enterprise is not sunk. There are, after all, examples of drugs successfully repurposed, even for COVID-19, and some have even begun from screening hits (though typically they are subsequently modified chemically[43]).

These caveats should not obscure the main observations from this study. Many drugs repurposed for COVID-19 in biochemical assays are aggregators—still, others may be inhibiting through other artifactual mechanisms—and their promise as leads for treating the disease merits reconsideration. Indeed, while some repurposed drugs have advanced further into development,[44] the aggregators described here do not seem have been further progressed. More broadly, drugs in repurposing libraries, though de-risked for whole body toxicity, pharmacokinetic exposure, and metabolism, are not de-risked for artifactual activity at screening relevant concentrations. More encouragingly, what this study illuminates is a series of facile assays that can rapidly distinguish drugs acting

artifactually via colloidal aggregation from those drugs with true promise for treating

SARS-CoV-2, and from pandemics yet to be faced.

3.5 Acknowledgements

3.6 Figures
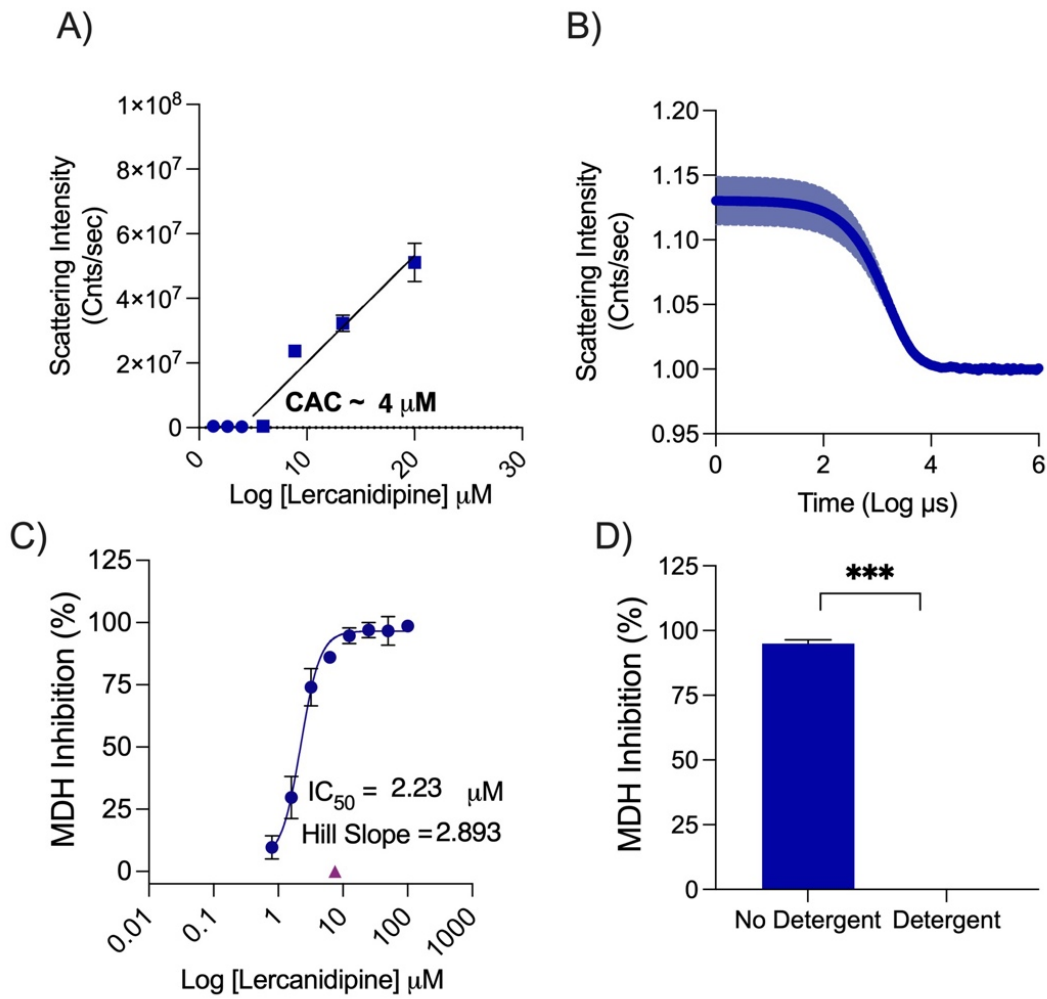
Figure 3.1. Lercanidipine's behavior as an aggregator.
(A) Critical aggregation concentration determined using scattering intensity measured on DLS. (B) Autocorrelation curve from DLS at 100 µM. (C) Dose response measured against MDH and showing the Hill slope. (D) MDH inhibition measured with or without 0.01% Triton-X-100 at 7.5 µM.
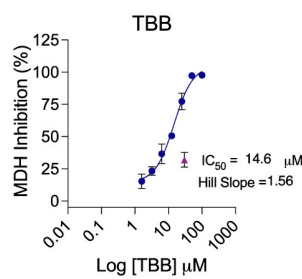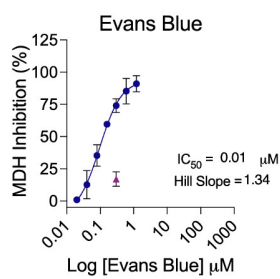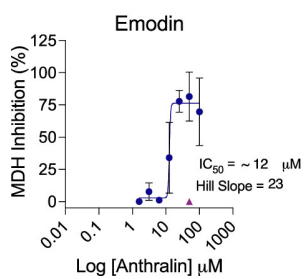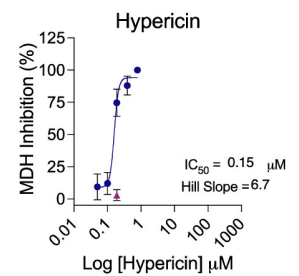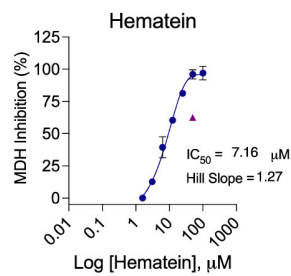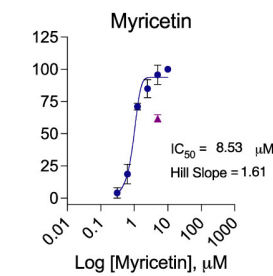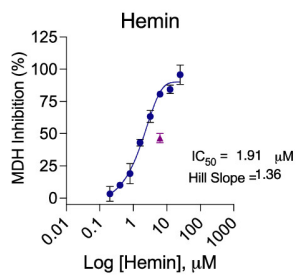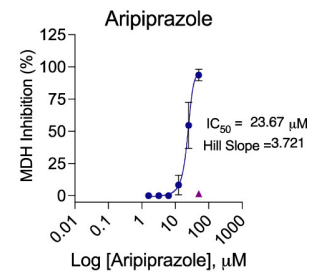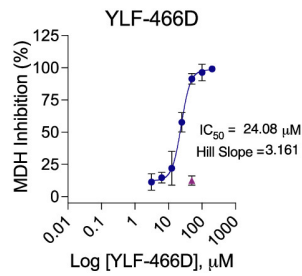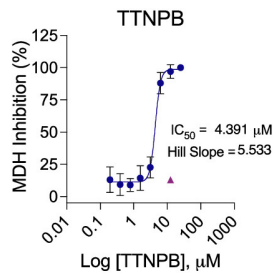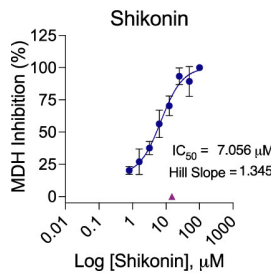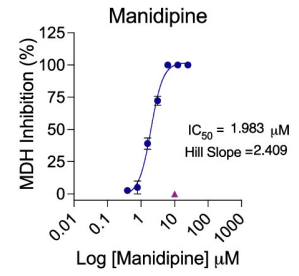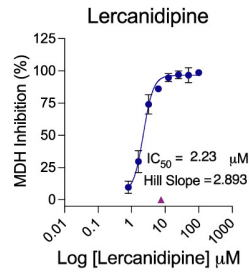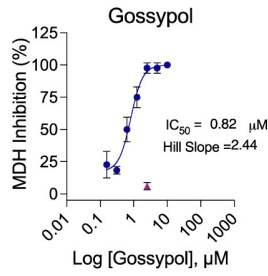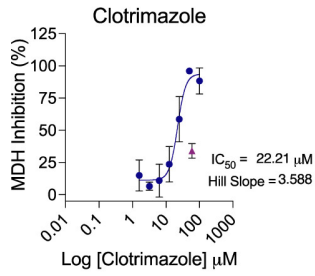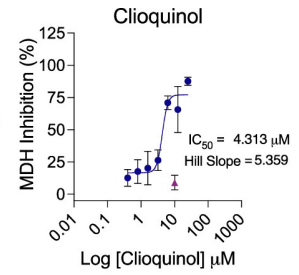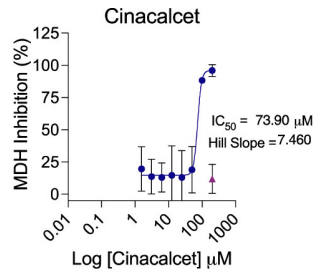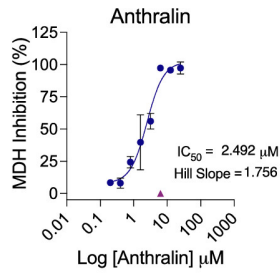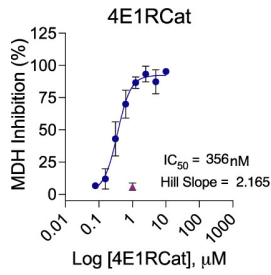
Figure 3.2. MDH inhibition concentration-response curves for literature active compounds.

IC_{50} and Hill slopes are shown. Purple triangles indicate single point MDH inhibition with the addition of 0.01% TritonX-100, tested at 3 times IC_{50}. All measurements are in triplicate.

4E1RCat
CAC = 6.609 μM

Anthralin
CAC ~ 4 μM

Cinacalcet
CAC = 53.23 μM

Clioquinol
CAC ~ 15 μM

Clotrimazole
CAC = 21.12 μM

Gossypol
CAC = 33.89 μM

Lercanidipine
CAC ~ 4 μM

Manidipine
CAC = 2.822 μM

Shikonin
CAC = 7.665 μM

TTNPB
CAC = 5.230 μM

YLF-466D
CAC = 49.24 μM

Aripiprazole
CAC ~18.76 μM

Hemin
CAC ~ 12.97 μM

Myricetin
CAC ~7.39 μM

Hematein
CAC ~123 μM

Hypericin
CAC = 0.3 μM

Emodin
CAC = 6 μM

110
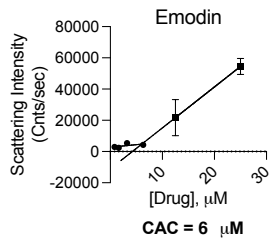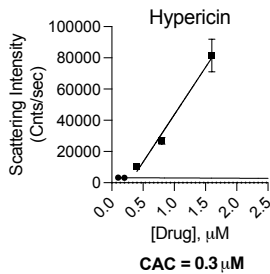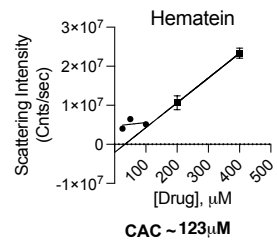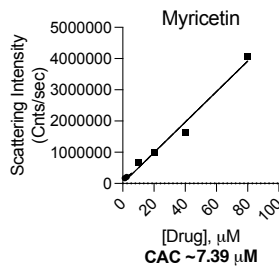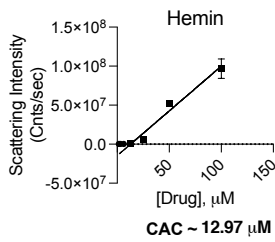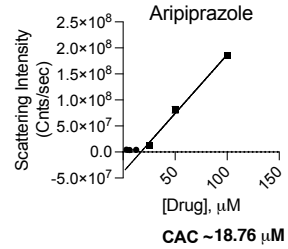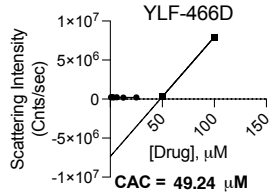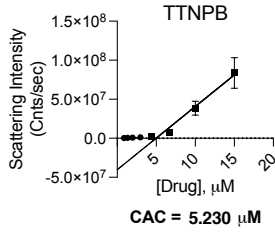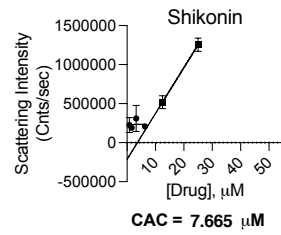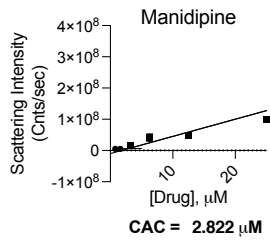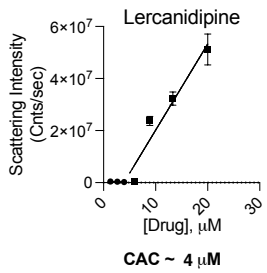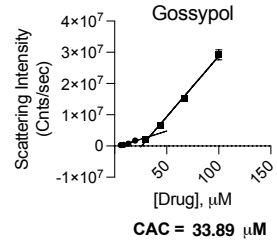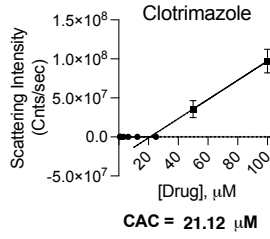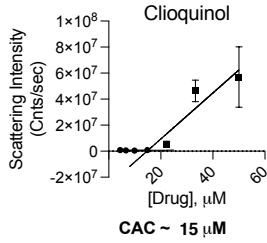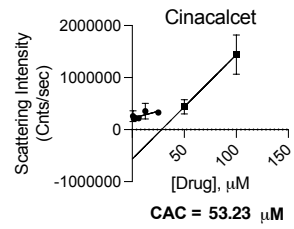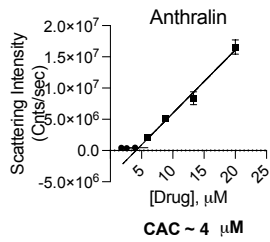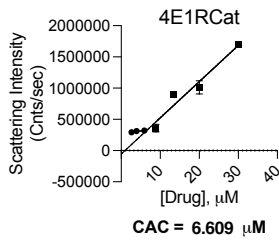
Figure 3.3. Critical aggregation concentrations for literature active compounds.
The CAC is determined by finding the intersection of two best fit lines for points with scattering intensity above or below 1×10^6. All measurements are in triplicate.
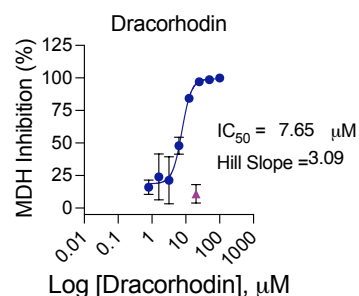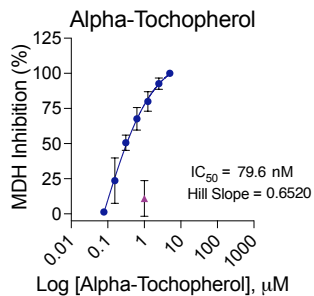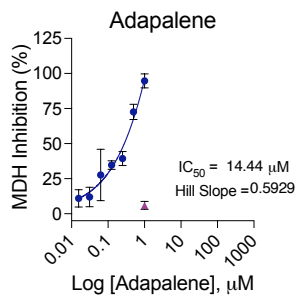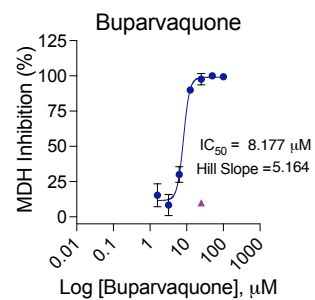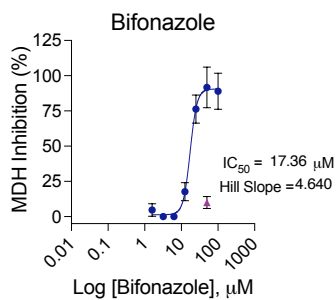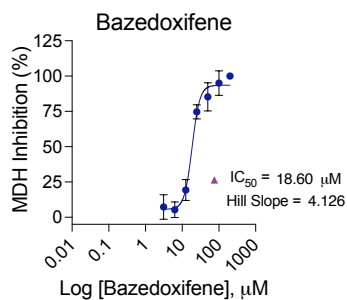
Figure 3.4. MDH inhibition dose-response curves for drugs drawn from a repurposing library.
All measurements were in triplicate.
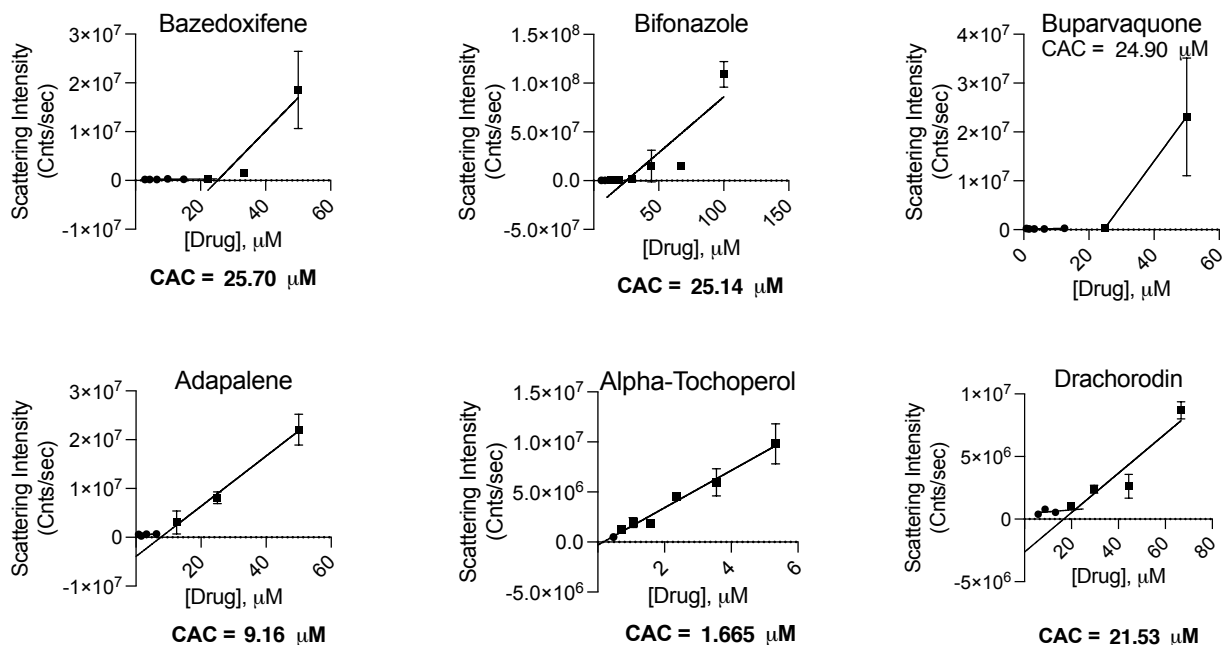
Figure 3.5. Critical aggregation concentrations for drugs drawn from a repurposing library.
The CAC is determined by the intersection of two best fit lines, for points with scattering intensity above or below 1×10^6. All measurements were in triplicate.
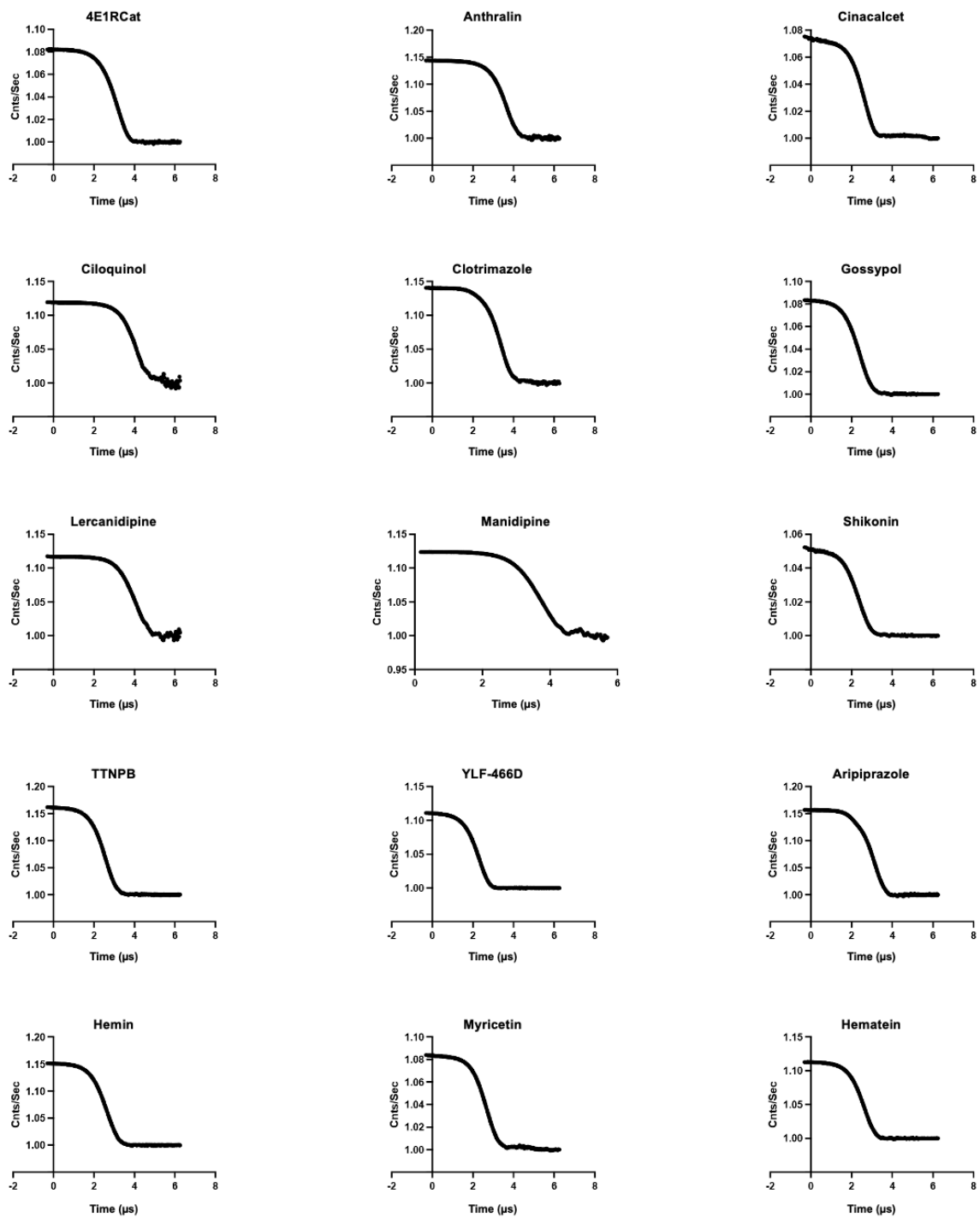
Figure 3.S1. DLS autocorrelation curves for literature reported hits.
Drug concentrations were at 3x IC_{50} measured for MDH.
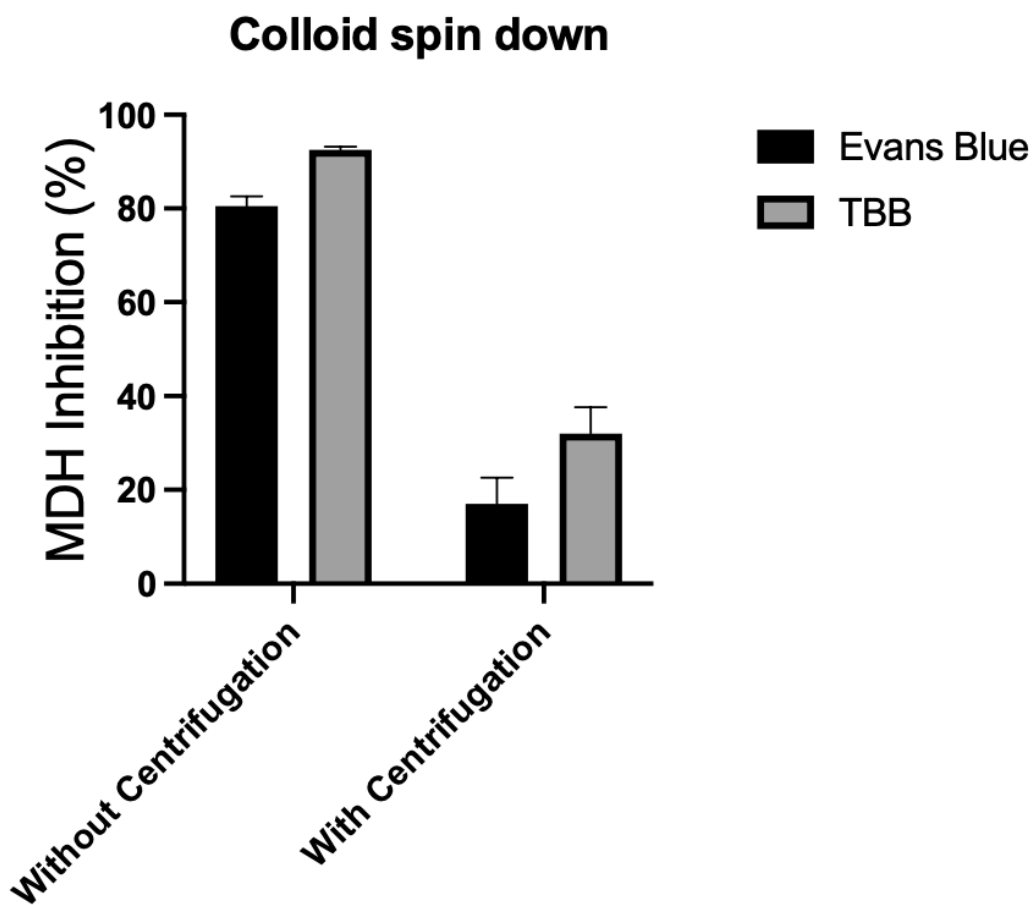
Figure 3.S2. MDH activity is restored when colloidal solution is centrifuged.
Evans blue was tested at 0.02 μM and TBB was tested at 25 μM.
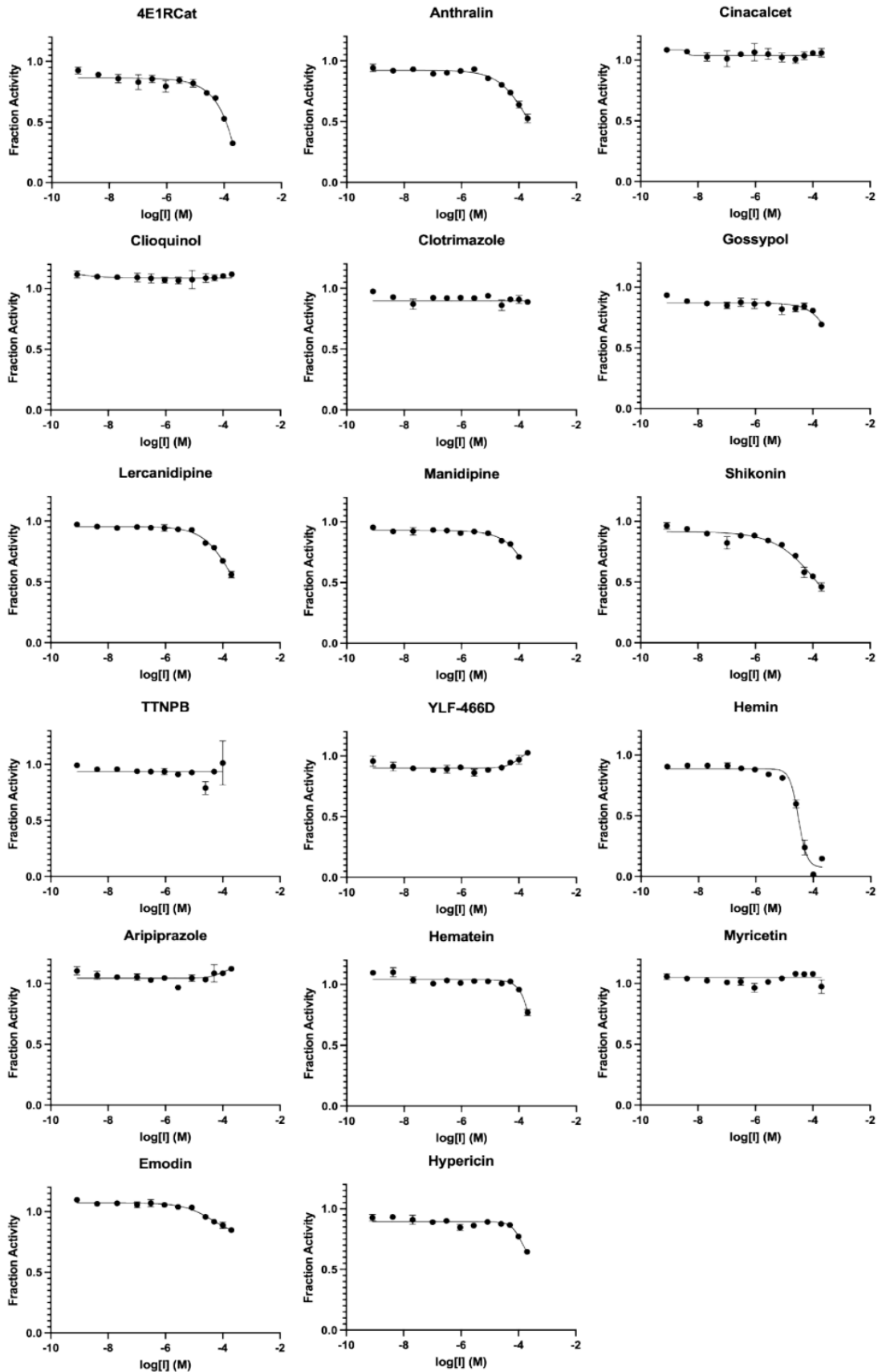
Figure 3.S3. Concentration response curves for literature compounds with 3CL-Pro in the presence of detergent.

With the exception of hemin, drugs at concentrations up to 200 µM showed little potency on the enzyme in the presence of 0.05% Tween-20.
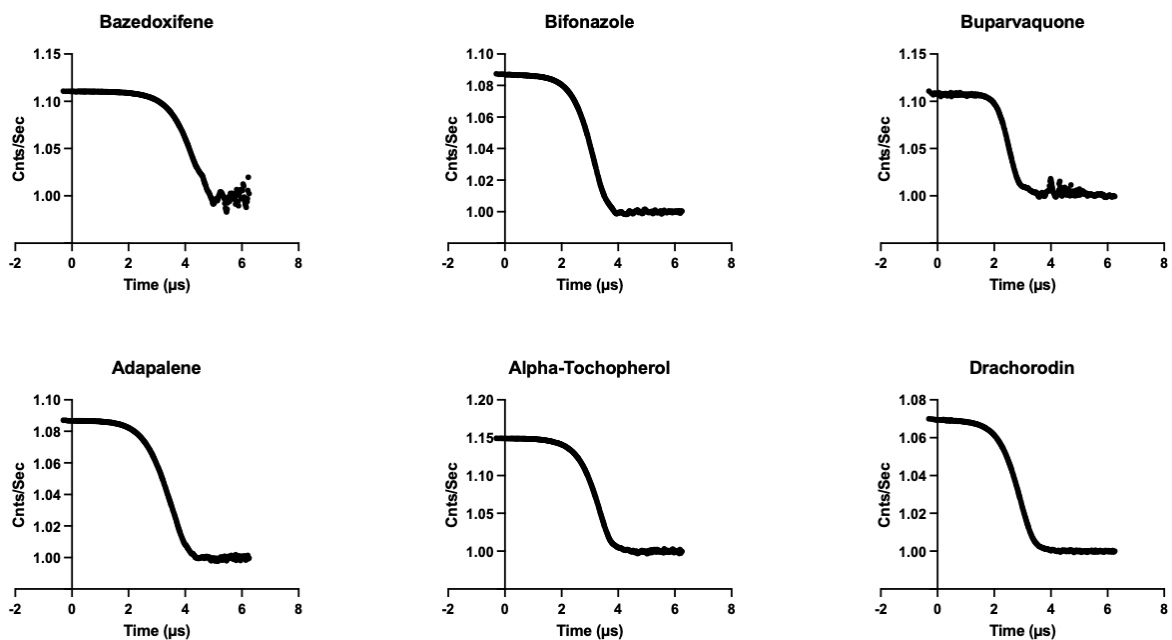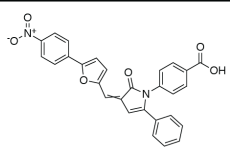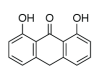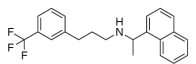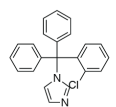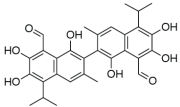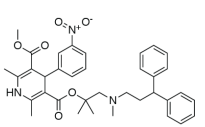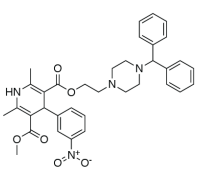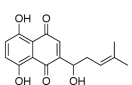
Figure 3.S4. DLS autocorrelation curves for drugs drawn from the repurposing library.
Drug concentrations were at 3x IC_{50} measured for MDH.
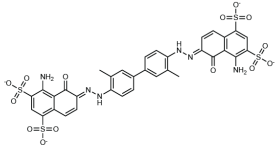
3.7 Tables

Table 3.1. Literature SARS-CoV-2 repurposing hits shown to cause colloidal aggregation.

Compound	Literature IC_{50}[a] (µM)	MDH IC_{50} (µM)	% Reduction in MDH inhibition on detergent addition[b]	CAC[c] (µM)	Colloid radius ±SD (nm)	Structure
4E1RCat	18.3	0.36	75%	6.6	102 ± 8	
Anthralin	Z ≤ 2	2.5	93%	4.5	1014 ± 53	
Cinacalcet	28.2	74	81%	53	77 ± 4	
Clioquinol	5.6, 2.8	4.3	57%	15	1818 ± 177	
Clotrimazole	39.8	22	53%	21	283 ± 15	
Gossypol	39.8	0.82	92%	34	70 ± 6	
Lercanidipine	16.2	2.2	95%	4	853 ± 147	
Manidipine	4.8	2.0	92%	2.8	929 ± 75	
Shikonin	15.8	7.0	92%	7.7	169 ± 9	

Compound	Literature IC$_{50}$[a] (µM)	MDH IC$_{50}$ (µM)	% Reduction in MDH inhibition on detergent addition[b]	CAC[c] (µM)	Colloid radius ±SD (nm)	Structure
TTNPB	35.5	4.4	83%	5.2	72 ± 3	
YLF-466D	35.5	24	22%	49	40 ± 0.4	
Hemin	9.7	1.9	39%	13	84 ± 0.9	
Aripiprazole	26(min)*	24	91%	19	222 ± 5.	
Hematein	10*	7.6	24%	123	64 ± 1	
Myricetin	10*	8.5	32%	7.4	100 ± 4	
Emodin	51.23	12.7	59%	6	58 ± 17	
Hypericin	19.34	0.15	83%	0.3	143 ± 6	

Compound	Literature IC$_{50}$[a] (μM)	MDH IC$_{50}$ (μM)	% Reduction in MDH inhibition on detergent addition[b]	CAC[c] (μM)	Colloid radius ±SD (nm)	Structure
Evans Blue	4.8	0.01	64%	ND[e]	ND[e]	
TBB	4.0	14.6	61%	ND[e]	ND[e]	

[a]IC$_{50}$s measured against mPro or ACE2 in a variety of assays; see citations in **Table 3.S1**
[b]Single-point Triton-X 0.01% reversal assay performed at approximately 3x MDH IC$_{50}$
[c]Critical Aggregation Concentration
*No IC$_{50}$ available, single point or retention time

Table 3.2. Literature repurposing hits do not potently inhibit 3CL-Pro in the presence of detergent.

Compound	Literature IC$_{50}$^{a}(µM)	3CL-Pro IC$_{50}$ with 0.05% Tween-20 IC$_{50}$ (µM)
4E1RCat	18.3	~100
Anthralin	Z ≤ 2	>200
Clotrimazole	39.8	>200
Gossypol	39.8	>200
Lercanidipine	16.2	>200
Manidipine	4.8	>100*
Shikonin	15.8	~200
TTNPB	35.5	>200
YLF-466D	35.5	>200
Hemin	9.7	25
Hematein	10*	>200
Emodin	51.2	>200

*100µM was the highest concentration used for Manidipine, instead of 200µM

Table 3.3. Six drugs from a repurposing library aggregate at screening-relevant concentrations.

Compound	MDH IC$_{50}$ (µM)	% Change in MDH Inhibition in presence of Triton-X[a]	CAC[c] (µM)	Colloid Radius ± SD (nm)	Structures
Adapalene	14	77%	9.2	422 ± 65	
Buparvaquone	8.2	90%	32	186 ± 15	
Bifonazole	17	87%	25	184 ± 80	
Alpha-Tochopherol	0.079	69%	1.5	466 ± 5.6	
Bazedoxifene	18	70%	26	2822 ± 814	
Dracorhodin	7.7	86%	22	101 ± 22	

[a.] Single-point Triton-X 0.01% reversal assay performed at approximately 3X MDH IC$_{50}$
[b.] Indicates the drug concentration at which colloid radius measurements were made
[c.] Critical Aggregation Concentration

Note: **Table 3.S1** and **3.S2** are available in the published version of the manuscript.

3.8 Materials and Methods

Literature Search and Chemoinformatic Selection of Potential Aggregators.
We used two approaches to identify drugs with the potential to form colloidal aggregates
from repurposing screens: (1) literature searches of published SARS-CoV-2 biochemical
drug screening papers including chemoinformatic analysis of the NCATS COVID-19
OpenData Portal[37] 3CL-Pro and ACE2 biochemical drug screens, and (2)
chemoinformatic predictions of potential aggregators found in the SelleckChem FDA-
approved drug library using the Aggregation Advisor tool.[42] Literature-based keyword
searches were performed using variations of the keywords "SARS-CoV-2" and "drug
repurposing" or "drug screen". Inhibitors from biochemical drug-repurposing screens were
visually inspected and prioritized for testing if they had cLogP values >3 or were highly
conjugated. Next, data from the NCATS COVID-19 OpenData Portal[37] drug-repurposing
screens for modulators of 3CL-Pro and ACE2 activities were retrieved (accessed on
September 28, 2020). In total, 12,262 and 3,405 annotations were found for compounds
screened against 3CL-Pro and ACE2, respectively. Molecules annotated with PubChem[44]
substance identifiers that had activities (AC50s) less than 50 µM but typically greater than
5 µM were selected. Simplified molecular input line entry system (SMILES) data for each
compound were retrieved using the PubChemPy API (https://pubchempy.readthedocs.io)
and used to calculate cLogP values using RDkit-2019.09.3.0 (http://www.rdkit.org).
Molecules with cLogP > 3 were drawn, visually inspected for the presence of molecular
features seen in known aggregators (e.g., multiple conjugated ring systems, overall
hydrophobicity, and no covalent warheads or PAINs), and prioritized for testing. Finally,
the SMILES of 2336 unique desalted molecules were selected from the SelleckChem

library and were analyzed with Aggregation Advisor[42], a command line tool that calculates molecular similarity (Tanimoto coefficients; Tc) between a list of molecules and a database of known aggregators (Table S2). Molecules with $1 > Tc$'s > 0.65 to a known aggregator and cLogP > 4 were drawn, inspected for structural diversity from one another, and for the presence of molecular features seen in known aggregators (e.g., multiple conjugated ring systems), and prioritized for testing. Percentages were calculated relative to the 2336 unique molecules in the library with identified SMILES.

Compounds. All compounds are >95% pure by HPLC, as reported by the vendors. Compounds were ordered from Sigma-Aldrich, Selleck Chem, Cayman Chemical, or Medchem Express.

Dynamic Light Scattering. To detect and quantify colloids, a DynaPro Plate Reader II (Wyatt Technologies) with a 60 mW laser at 830 nm wavelength and a detector angle of 158° was used; the beam size of the instrument was increased by the manufacturer to better enable detection of the colloids, which are larger than protein aggregates for which the instrument was designed. Samples were measured in 384- well plates with 30 µL loading and 10 acquisitions per sample. Compounds were dissolved in DMSO at 100 times their final concentration and were diluted into filtered 50 mM KPi, pH 7.0, to obtain a final 1% DMSO concentration. Compounds were first tested at 3 times the IC50 reported in the literature, and if active, they were further investigated in concentration–response testsIf no IC50 was available, compounds were tested at 100 µM. To calculate a CAC, each compound was serially diluted until substantial scattering

disappeared; aggregating (>106 scattering intensity) and non-aggregating (<106 scattering intensity) portions of the data were fitted with separate nonlinear regression curves, and the point of intersection was determined using GraphPad Prism software version 9.1.1 (San Diego, CA).

Enzyme Inhibition. MDH inhibition assays were performed at room temperature on a HP8453a spectrophotometer in kinetic mode using UV–vis Chemstation software (Agilent Technologies) in methacrylate cuvettes (Fisher Scientific, 14955128) with a final volume of 1 mL for both control and test reactions. MDH (from porcine heart, 901643, Sigma Millipore) was added to a 50 mM KPi pH 7 buffer for a final concentration of 2 nM. Compounds were dissolved in DMSO at 100 times concentration; 10 μL of compound was used for a final DMSO concentration of 1%. After compound addition , the cuvette was mixed by pipetting up and down 5 times with a p1000, and the cuvette was then incubated for 5 min at room temperature. The reaction was initiated by the addition of 200 μM nicotinamide adenine dinucleotide (54839, Sigma- Aldrich) and 200 μM oxaloacetic acid (324427, Sigma Aldrich), and the rate was monitored at 340 nm. A negative control was included in each run, in which 10 μL of DMSO without the compound was added. The reactions were monitored for 90 s, and the initial rates were divided by the initial rate of the negative control to obtain the % inhibition and % enzyme activity. For dose response curves, three replicates were done for each concentration, the graphs were generated using GraphPad Prism version 9.1.1 (San Diego, CA).

3CL-Pro Kinetics Inhibition Assay. A fluorescence-quenched substrate with the sequence rr-K(MCA)-ATLQAIASK(DNP)-COOH was synthesized via the Fmoc solid-phase peptide synthesis as described.[45] Recombinant, active 3CL-Pro was expressed and purified as described.[46] Kinetic measurements were carried out in Corning black 384-well flat- bottom plates and read on a BioTek H4 multimode plate reader. The quenched fluorogenic peptide had a final concentration of KM=10 μM, and 3CL-Pro had a final concentration of 50 nM. The reaction buffer was 20 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.05% Tween-20 (v/v), and 1 mM DTT, pH 7.4. Drugs were incubated with protease prior to substrate addition at 37 °C for 1h After incubation, the substrate was added, and kinetic activity was monitored for 1h at 37°C. Initial velocities were calculated at 1 to 45 min in RFU/s. Velocities were corrected by subtracting the relative fluorescence of a substrate-only control, and fraction activity was calculated using a substrate-corrected no-inhibitor control where DMSO was added instead of a drug. Kinetic measurements were carried out in triplicate.

Colloid Centrifugation. DMSO stocks of drugs were prepared and diluted to 100:1 into 1 mL of 50 mM KPi buffer, pH7, in a 1.5 mL Eppendorf tube. This was mixed by pipetting and centrifuging at 14,000 rpm for 1h at 4°C in a benchtop microfuge. The supernatant (900μL of 1mL) was then tested for MDH inhibition as previously described.

3.9 References

1. Edwards, A. What Are the Odds of Finding a COVID-19 Drug from a Lab Repurposing Screen? *J. Chem. Inf. Model* **60**, 5727–5729 (2020).

2. Edwards, A. & Hartung, I. V. No shortcuts to SARS-CoV-2 antivirals. *Science* **373**, 488–489 (2021).

3. Papapetropoulos, A. & Szabo, C. Inventing new therapies without reinventing the wheel: the power of drug repurposing. *Br. J. Pharmacol.* **175**, 165–167 (2018).

4. Dotolo, S., Marabotti, A., Facchiano, A. & Tagliaferri, R. A review on drug repurposing applicable to COVID-19. *Brief. Bioinform.* **22**, bbaa288- (2020).

5. Bangladesh, D. of B. and G. E., Noakhali Science and Technology University, Noakhali-3814, *et al.* Drug Repurposing for Prevention and Treatment of COVID-19: A Clinical Landscape. *Discoveries* **8**, e121 (2020).

6. Singh, T. U. *et al.* Drug repurposing approach to fight COVID-19. *Pharmacol. Rep.* **72**, 1479–1508 (2020).

7. Saxena, A. Drug targets for COVID-19 therapeutics: Ongoing global efforts. *J. Biosci.* **45**, 87 (2020).

8. 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.* **53**, 2719–2740 (2010).

9. Baell, J. B. Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS). *J. Nat. Prod.* **79**, 616–628 (2016).

10. Dahlin, J. L., Inglese, J. & Walters, M. A. Mitigating risk in academic preclinical drug discovery. *Nat. Rev. Drug Discov.* **14**, 279–294 (2015).

11. Thorne, N., Auld, D. S. & Inglese, J. Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr. Opin. Chem. Biol.* **14**, 315–324 (2010).

12. Lloyd, M. D. High-Throughput Screening for the Discovery of Enzyme Inhibitors. *J. Med. Chem.* **63**, 10742–10772 (2020).

13. Feng, B. Y. *et al.* A High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library. *J. Med. Chem.* **50**, 2385–2390 (2007).

14. Jadhav, A. *et al.* Quantitative Analyses of Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of a Thiol Protease. *J. Med. Chem.* **53**, 37–51 (2010).

15. Doak, A. K., Wille, H., Prusiner, S. B. & Shoichet, B. K. Colloid Formation by Drugs in Simulated Intestinal Fluid. *J. Med. Chem.* **53**, 4259–4265 (2010).

16. Seidler, J., McGovern, S. L., Doman, T. N. & Shoichet, B. K. Identification and Prediction of Promiscuous Aggregating Inhibitors among Known Drugs. *J. Med. Chem.* **46**, 4477–4486 (2003).

17. Bojadzic, D., Alcazar, O. & Buchwald, P. Methylene Blue Inhibits the SARS-CoV-2 Spike–ACE2 Protein-Protein Interaction–a Mechanism that can Contribute to its Antiviral Activity Against COVID-19. *Front. Pharmacol.* **11**, 600372 (2021).

18. Ghahremanpour, M. M. *et al.* Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **11**, 2526–2533 (2020).

19. McGovern, S. L. & Shoichet, B. K. Kinase Inhibitors: Not Just for Kinases Anymore. *J. Med. Chem.* **46**, 1478–1483 (2003).

20. Ganesh, A. N., Donders, E. N., Shoichet, B. K. & Shoichet, M. S. Colloidal aggregation: From screening nuisance to formulation nuance. *Nano Today* **19**, 188–200 (2018).

21. Blevitt, J. M. *et al.* Structural Basis of Small-Molecule Aggregate Induced Inhibition of a Protein–Protein Interaction. *J. Med. Chem.* **60**, 3511–3517 (2017).

22. McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* **46**, 4265–4272 (2003).

23. Coan, K. E. D., Maltby, D. A., Burlingame, A. L. & Shoichet, B. K. Promiscuous Aggregate-Based Inhibitors Promote Enzyme Unfolding. *J. Med. Chem.* **52**, 2067–2075 (2009).

24. Lak, P., O'Donnell, H., Du, X., Jacobson, M. P. & Shoichet, B. K. A Crowding Barrier to Protein Inhibition in Colloidal Aggregates. *J. Med. Chem.* **64**, 4109–4116 (2021).

25. Feng, B. Y. *et al.* Small-molecule aggregates inhibit amyloid polymerization. *Nat. Chem. Biol.* **4**, 197–199 (2008).

26. LaPlante, S. R. *et al.* Monitoring Drug Self-Aggregation and Potential for Promiscuity in Off-Target In Vitro Pharmacology Screens by a Practical NMR Strategy. *J. Med. Chem.* **56**, 7073–7083 (2013).

27. Baker, J. D., Uhrich, R. L., Kraemer, G. C., Love, J. E. & Kraemer, B. C. A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. *PLoS ONE* **16**, e0245962 (2021).

28. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).

29. Zhu, W. *et al.* Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening. *ACS Pharmacol. Transl. Sci.* **3**, 1008–1016 (2020).

30. Olaleye, O. A., Kaur, M., Onyenaka, C. & Adebusuyi, T. Discovery of Clioquinol and analogues as novel inhibitors of Severe Acute Respiratory Syndrome Coronavirus 2 infection, ACE2 and ACE2 - Spike protein interaction in vitro. *Heliyon* **7**, e06426 (2021).

31. Han, Y. *et al.* Identification of SARS-CoV-2 inhibitors using lung and colonic organoids. *Nature* **589**, 270–275 (2021).

32. Tripathi, P. K. *et al.* Screening and evaluation of approved drugs as inhibitors of main protease of SARS-CoV-2. *Int. J. Biol. Macromol.* **164**, 2622–2631 (2020).

33. Fu, W. *et al.* Repurposing FDA-approved drugs for SARS-CoV-2 through an ELISA-based screening for the inhibition of RBD/ACE2 interaction. *Protein Cell* **12**, 586–591 (2021).

34. Ge, S. *et al.* Repositioning of histamine H1 receptor antagonist: Doxepin inhibits viropexis of SARS-CoV-2 Spike pseudovirus by blocking ACE2. *Eur. J. Pharmacol.* **896**, 173897 (2021).

35. Lin, C. *et al.* Ceftazidime is a potential drug to inhibit SARS-CoV-2 infection in vitro by blocking spike protein–ACE2 interaction. *Signal Transduct. Target. Ther.* **6**, 198 (2021).

36. Jang, W. D., Jeon, S., Kim, S. & Lee, S. Y. Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. *Proc. Natl. Acad. Sci.* **118**, e2024302118 (2021).

37. Brimacombe, K. R. *et al.* An OpenData portal to share COVID-19 drug repurposing data in real time. *bioRxiv* 2020.06.04.135046 (2020) doi:10.1101/2020.06.04.135046.

38. White, M. A., Lin, W. & Cheng, X. Discovery of COVID-19 Inhibitors Targeting the SARS-CoV-2 Nsp13 Helicase. *J. Phys. Chem. Lett.* **11**, 9144–9151 (2020).

39. Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **49**, 7274–7277 (2006).

40. McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.* **45**, 1712–1722 (2002).

41. Coan, K. E. D. & Shoichet, B. K. Stability and equilibria of promiscuous aggregates in high protein milieus. *Mol. Biosyst.* **3**, 208–213 (2007).

42. Irwin, J. J. *et al.* An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).

43. Zhang, C.-H. *et al.* Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med. Chem. Lett.* **12**, 1325–1332 (2021).

44. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, gkaa971- (2020).

45. Ravalin, M. *et al.* Specificity for latent C termini links the E3 ubiquitin ligase CHIP to caspases. *Nat. Chem. Biol.* **15**, 786–794 (2019).

46. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).

Gloss to Chapter 4

After everything you've read so far, you may have started to grow weary of the σ receptors as worthwhile drug targets. However, the next chapter of this dissertation describes a legitimate use for σ_2 ligands as efficacious treatments for neuropathic pain. Before delving into this further, I would like to give some context on the timeline of this project. Preceding our work that demonstrated σ receptor ligands confound drug repurposing screens, the lab coincidentally was already working on a large-scale docking project against the, as Brian put it, *famous, if star-crossed,* σ_2 receptor. This work was led by JK Lyu from our lab and Assaf Alon from Andy Kruse's lab at Harvard Medical School. When σ_2 popped up in the proteomics network for SARS-CoV-2, we were excited to test JK's novel ligands, extending the scope of the project beyond drug repurposing into novel chemical territory. At this point, the σ receptor team joined the drug repurposing team and we became a dream team, ready to save the world. However, as with the FDA-approved drugs that bound σ receptors, the novel low nanomolar potency σ_2 ligands only had antiviral effects in the micromolar range. We eventually figured out that some σ receptor ligands induced phospholipidosis, were not efficacious *in vivo,* and we stopped pursuing them as antivirals.

At this time I was simultaneously working with the Basbaum lab on the DARPA Panacea project to discover novel non-opioid pain therapeutics. I stumbled upon a body of literature that showed *in vivo* effects of σ_2 ligands against neuropathic pain models with interesting time-dependent effects. I proposed similar experiments using JK's ligands— controlling for phospholipidosis—to the Basbaum and Kruse labs, and we were able to

replicate these phenotypes in our own hands with our own selective ligands. This work

eventually became a part of JK and Assaf's larger large-scale docking story. I was grateful

to be able to play a supporting role in this project given the expertise I had developed with

these ligands and targets in my other projects.

Chapter 4:Structures of the σ2 receptor enable docking for bioactive

ligand discovery

Structures of the σ2 receptor enable docking for bioactive ligand discovery

Assaf Alon[1,†], Jiankun Lyu[2,†], Joao M. Braz[3,†], Tia A. Tummino[2,4], Veronica Craik[3], Matthew J. O'Meara[5], Chase M. Webb[2,4], Dmytro S. Radchenko[6,11], Yurii S. Moroz[7], Xi-Ping Huang[8,10], Yongfeng Liu[8,10], Bryan L. Roth[8,9,10], John J. Irwin[2], Allan I. Basbaum[3,*], Brian K. Shoichet[2,*] & Andrew C. Kruse[1,*]

† These authors contributed equally.

* Corresponding authors.

[1]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

[2]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA

[3]Department of Anatomy, University of California, San Francisco, San Francisco, CA 94143, USA.

[4]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA

[5]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

[6]Enamine Ltd., Chervonotkatska 78, 02094 Kyiv, Ukraine

[7]Chemspace LLC, 85 Chervonotkatska Street, Kyiv, 02094, Ukraine

[8]Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599-7365, USA

[9]Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7360, USA

[10]National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), School of Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599-7365, USA

[11]Taras Shevchenko National University of Kyiv, 60 Volodymyrska Street, Kyiv 01601, Ukraine

137

4.1 Abstract

The σ_2 receptor has attracted intense interest in cancer imaging[1], psychiatric disease[2], neuropathic pain[3–5], and other areas of biology[6,7]. We determined the crystal structure of this receptor in complex with the clinical candidate roluperidone[2] and the tool compound PB28[8]. These structures templated a large-scale docking screen of 490 million virtual molecules, of which 484 compounds were synthesized and tested. 127 new chemotypes with affinities superior to 1 μM were identified, 31 of which had affinities superior to 50 nM. Hit rate fell smoothly and monotonically with docking score. We optimized three hits for potency and selectivity, achieving affinities ranging from 3 to 48 nM with up to 250-fold selectivity versus the σ_1 receptor. Crystal structures of two new ligands bound to the σ_2 receptor confirmed the docked poses. To investigate the contribution of the σ_2 receptor in pain, two potent σ_2-selective ligands and one potent σ_1/σ_2 non-selective ligand were tested for efficacy in a mouse neuropathic pain model. All three ligands demonstrated time-dependent decreases in mechanical hypersensitivity in the spared nerve injury model[9], supporting a role for the σ_2 receptor in nociception. This study illustrates the opportunities for rapid discovery of *in vivo* probes to study under-explored areas of biology using structure-based screens of diverse, ultra-large libraries.

4.2 Introduction

The σ receptors are integral membrane proteins widely expressed in both the central nervous system (CNS) and in peripheral tissues[10]. They are divided into σ_1 and σ_2 subtypes based on differences in tissue distribution and in pharmacological profile[11], but despite their names, the two proteins are sequence-unrelated. Cloned in 1996, the σ_1 receptor has no paralog within the human genome; its closest homolog of known function is the yeast Δ8,7 sterol isomerase ERG2[12]. Studies conducted on σ_1 knockout mouse tissue[13] showed that the σ_2 is not a splice variant or modified form of σ_1, but rather derives from an unrelated gene. The molecular identity of the σ_2 receptor remained unknown until we purified it from calf liver tissue[14] and showed that it is TMEM97, an ER-resident membrane protein that regulates the sterol transporter NPC1[15,16]. TMEM97 is predicted to be a four-helix bundle protein with both amino and carboxy termini facing the cytoplasm. An EXPERA family[17] member, the σ_2 receptor is distantly related to emopamil-binding protein (EBP), the mammalian Δ8,7 sterol isomerase required for cholesterol synthesis, and to TM6SF2, which regulates liver lipid homeostasis[18].

The σ_2 receptor is overexpressed in proliferating cells and in many tumors[19], and labeled σ_2 ligands have been proposed as tools for cancer diagnosis and therapy[1]. A ternary complex between the σ_2 receptor, PGRMC1, and the LDL receptor was reported to increase the rate of LDL internalization[7]. Consistent with its high expression in the CNS, the σ_2 receptor has also been proposed as a target for the treatment of CNS disorders. The σ_2 receptor ligand Elayta (CT1812) is in clinical trials for mild to moderate Alzheimer's disease[6], and roluperidone (MIN-101) is in clinical development for schizophrenia[2]. When

tested in animal models, σ_2 receptor ligands reduce alcohol-withdrawal symptoms[5,20] and have a neuroprotective effect in brain injury[21]. Finally, recent studies have found σ_2 ligands to be anti-allodynic in models of neuropathic pain[3–5]. As this is also true of σ_1 ligands, and because most σ_2 ligands cross-react with the σ_1 receptor, probe ligands selective for σ_2 over σ_1 would help illuminate σ_2 biology and could be leads for novel therapeutics. However, little is known of the receptor's molecular architecture or the structural bases for ligand recognition, stunting the discovery of selective ligands[22,23]. Here, we employed a biochemical and structural approach combined with computational docking to address these issues.

4.3 Results

Structure determination

The human σ_2 receptor was expressed in *Sf9* insect cells, extracted with detergent, and purified[14]. Size exclusion chromatography multi-angle light scattering (SEC-MALS) showed that the receptor is a dimer in solution. Intriguingly, all members of the EXPERA family are either dimers or pseudo-dimers, although the functional role of dimerization remains unknown. Unlike the σ_1 receptor, which can change oligomeric state in response to ligand binding[24], the presence of ligands did not perturb the oligomeric state of σ_2 (**E.D. Fig. 4.1a**). As the human σ_2 receptor was not tractable in structural studies, further experiments were performed with the homologous bovine σ_2 receptor (**E.D. Fig. 4.1b**). Circular dichroism (CD) experiments showed that the bovine σ_2 receptor is 74% helical (**E.D. Fig. 4.1c**). Thermal unfolding demonstrated that the receptor is remarkably stable, with a midpoint of the unfolding transition (T_m) of 54°C (**E.D. Fig. 4.1d**). Crystals of the σ_2

receptor were grown by the lipidic cubic phase method[25] (**E.D. Fig. 4.1e-g**). Three datasets were collected for the receptor in complex with PB28[8], roluperidone[2], and a ligand tentatively modeled as cholesterol (**E.D. Table 4.1**). Molecular replacement was performed using a model derived from the structure of EBP[26] (see **Methods**).

Overall structure of the σ_2 receptor

The three σ_2 receptor crystal structures are similar, with a backbone root mean square deviation (RMSD) of 0.75 Å. As anticipated from SEC-MALS, the structures showed that σ_2 is an intimately associated homodimer, burying 890 $Å^2$ of surface area in a dimer interface mainly formed by transmembrane helix 3 (TM3; **Fig. 4.1a**). The two protomers adopt the same conformation (backbone RMSD of 0.34 Å, 160 residues), with each adopting the expected four-helix bundle fold.

The four transmembrane helices of the protein are all kinked due to the presence of proline residues in each, creating a ligand-binding cavity near the center of the protein. This cavity is entirely occluded from solvent by extracellular loops 1 and 2, which form a well-ordered cap over the luminal surface of the protein. Asp56, which is crucial for ligand binding[14], bridges extracellular loop 1 to TM helix 4 using a hydrogen bond network (**E.D. Fig. 4.1h**). Hence, Asp56 is likely important for receptor folding and not directly for ligand recognition[14]. Rather than opening to the ER lumen, the pocket opens laterally into the lipid bilayer (**Fig. 4.1b**), reminiscent of lipid-binding G protein-coupled receptors[27,28], and its opening is lined with hydrophobic and aromatic residues. Ligands may enter through this opening in their neutral, deprotonated form and then become protonated in the

binding site, forming a salt bridge with the conserved Asp29 (**Fig. 4.1c-d**). A second conserved acidic residue, Glu73, is located 3 Å away from Asp29, suggesting that these residues are hydrogen-bonded to each another, with Glu73 likely protonated.

The two σ receptors are not homologs and do not share the same fold; the σ_2 receptor is a four-helix bundle, while the σ_1 receptor has a β-barrel cupin fold[29]. Nevertheless, the binding pockets of the two receptors are remarkably similar (**Fig. 4.1c-e**), placing functionally similar amino acids in cognate spatial positions, which is perhaps the result of convergent evolution and explains how two very different folds can share closely overlapping ligand recognition profiles.

Both σ receptors are homologs of proteins that catalyze the same step in sterol biosynthesis. The σ_1 receptor is a homolog of ERG2, the fungal Δ8,7 sterol isomerase; the σ_2 receptor is a homolog of EBP, the mammalian Δ8,7 sterol isomerase. Both EBP and ERG2 rely on two interacting acidic residues in their active site for catalysis, which occurs by protonation of the substrate at carbon 9 (C9) followed by proton abstraction from C7, shifting the double bond into the C8-C7 position. All necessary components for catalysis appear to be present in σ_2 receptor, yet it doesn't catalyze sterol isomerization. It can neither function *in vivo* to rescue a strain of yeast that lacks ERG2 (**E.D. Fig. 4.1k**) nor can it function *in vitro* to convert zymostenol to lathosterol (**E.D. Fig. 4.1l**). The same is true for the σ_1 receptor, which also has all the conserved residues required for catalysis but cannot rescue yeast that lack a sterol isomerase[12] (**E.D. Fig. 4.1k**). It was recently reported that Δ8-9 sterols can serve as signaling molecules[30], which may hint at a

possible physiological function of the σ receptors as sensors of these molecules evolved from enzymes that would modify them.

Docking against the σ₂ receptor

Docking against the σ₂ receptor had two goals: discovering novel and σ₂-selective chemotypes, and investigating how docking scores predict binding likelihood. This has been explored only once before at scale, against the dopamine receptor, revealing a sigmoidal relationship between hit-rate (active ligands/number-tested) and score[31,32]. The promiscuous σ₂ site promised a higher hit-rate, increasing the dynamic range of any relationship observed. Guided by score alone for most molecules picked, supplemented by manual selection among the highest-ranking docked molecules, we prioritized 577 molecules for synthesis, spread among 14 scoring bins, of which 484 compounds were successfully produced. We tested compounds at 1 µM and defined as "hits" those that displaced greater than 50% [³H]-DTG σ₂ binding. 127 of 484 molecules qualified, accounting for 26% of compounds over the full scoring range and a 60% hit-rate among the top-ranked molecules (**Fig. 4.2a**). Hit-rates fell monotonically with score, as with the dopamine receptor[32], with a slope of -4.2%/(kcal/mol) in the inflection region, with one exception (below). The curve dropped from a hit rate of 61% at a docking score of about -60 kcal/mol to 0% at the four lowest scoring bins (-40 to -22.5 kcal/mol) (**Fig. 4.2b, Supplementary Fig. 4.1**).

The highest scoring bin had a hit rate of 27%, much lower than the 61% hit-rate observed in the 2nd-best scoring bin. This dip in the hit-rate curve illuminates defects in the scoring function. Many of the molecules in the top scoring bin had unexpectedly low

desolvation penalties (**E.D. Fig. 4.2a,b,** left column). DOCK3.7 pre-calculates these energies from one conformation among hundreds docked, not necessarily the highest scoring conformation against a target. Indeed, recalculating ligand desolvation using the docked conformation for molecules tested against σ_2 and dopamine receptors increased desolvation penalties for molecules in the top-scoring bin, reducing their ranking and so suggesting a method to improve the scoring function (**E.D. Fig. 4.2d**).

To supplement molecules prioritized by score alone, we picked a comparable number of high-ranking molecules by human inspection[32,33]. In the top three scoring bins (139 molecules) the human-prioritized hit rate (67%) was higher than that by docking score alone (33%) (**E.D. Fig. 4.2e and E.D. Fig. 4.2f**), and the human-prioritized molecules reached higher affinities (**E.D. Fig. 4.2g,h**). Broadly, these patterns reflect what was observed against the dopamine receptor.

Seeking selective probes for the σ_2 receptor, we measured competition binding curves for 14 docking hits with high radioligand displacement at 1 µM. K_i values ranged from 2.4 to 68 nM. In competition binding versus σ_1 receptor (**Fig. 4.2d**, **E.D. Table 4.2**, and **Supplementary Table 4.1**), several of these had substantial selectivity for σ_2 over σ_1, including ZINC450573233 and ZINC895657866, which were 30- and 46-fold selective, respectively.

We sought to improve the affinities of three potent ligands, each representing a different scaffold (**E.D. Fig. 4.3a-c**). 20,000 analogs identified in SmallWorld

(https://sw.docking.org/) from a 28 billion virtual library were docked into the σ_2 site (**Methods**, **Supplementary Table 4.1**). Of these, 105 were synthesized and tested, improving the affinity of each scaffold by 2- to 18-fold (**E.D. Fig. 4.3a-c**, **Supplementary Table 4.1**); for two chemotypes, σ_2 selectivity improved 47- and >250-fold (Z1665845742 and Z4857158944), respectively.

Structures of σ_2 in complex with analogs

To test our docking poses, we determined the crystal structures of σ_2 bound to two high-affinity ligands Z1241145220 (σ_2 K_i = 3.7 nM; PDB ID: 7M95) and Z4857158944 (σ_2 K_i = 4 nM; PDB ID: 7M96). Electron density maps confirmed the docking predictions, with RMSD values between the crystallized and docked poses of 0.88 and 1.4 Å, respectively (**Fig. 4.3a-b**, **E.D. Table 4.1, and E.D. Fig. 4.1i**). Newly predicted hydrogen-bond interactions with the backbone carbonyl of Val146, which was not seen in the roluperidone or PB28 complexes, corresponded well between docked and crystallographic poses. A hydrogen bond interaction with Gln77 is also found in the roluperidone and cholesterol complexes (**Fig. 4.1d, E.D. Fig. 4.1j).** The higher resolution of this structure, 2.4 Å, also revealed an ordered water molecule in one of the binding sub-sites, coordinated by residues His21, Tyr103, and Gln77, and by an azaindole nitrogen in Z1241145220 (**Fig. 4.3b**).

This water was not modeled in the docked structure, so to investigate its role in ligand recognition we tested two analogues that were designed to disrupt the hydrogen bonds between Gln77 and the water **Fig. 4.3c**). Z295861754, which should only

hydrogen-bond with the water but not with Gln77, suffered an 8-fold decrease in affinity, whereas Z163048780, which should not hydrogen bond with either Gln77 or the water, had a K_i value > 10 µM (**Fig. 4.3d**), indicating a crucial role of the water for Z1241145220. We further generated a series of σ_2 mutants in which the coordination of this water molecule was disrupted. Competition binding assays with Z1241145220 showed that mutating either His21 or Gln77 reduces the affinity by about 10-fold (**E.D. Fig. 4.3d-f**). Taken together, these results demonstrate that the ordered water is an integral part of the binding pocket and is required for high-affinity binding of Z1241145220, and likely other ligands.

σ_2 ligands active in mouse pain model

Genetic[34,35] and pharmacological[36–38] evidence supports a role of σ_1 in chronic pain[39]. The discovery of the gene encoding for σ_2[14], made understanding and distinguishing the roles of σ_2 and σ_1 in this indication[3,4] fully possible. However, the limited availability of selective σ_2 probes[4] hinders the ability to distinguish the effect of the two receptors. Accordingly, we treated mice with three high-affinity σ_2 ligands with differing degrees of σ_2/σ_1 selectivity: Z4857158944 (4 nM; >250-fold selective), Z1665845742 (5 nM; 47-fold selective), and Z4446724338 (3 nM non-selective) (**Fig. 4.4a**). We also treated with PB28, a well-established 5 nM non-selective ligand[8]. In pharmacokinetics experiments, the three docking-derived ligands had substantial brain permeability, with brain to plasma ratios ranging from 3 to 16, and brain half-lives ranging from 1.2 to 12 hours (**E.D. Table 4.3**). PB28 also had high brain permeability and a relatively long half-life, though its brain C_{max} was 3- to 8-fold lower than that of the new compounds. The high

brain exposures of all four compounds encouraged us to examine them in a neuropathic pain model in mice.

We tested the efficacy of these ligands in the spared nerve injury (SNI) mouse model of neuropathic pain, in which two out of three branches of the sciatic nerve are transected[9], resulting in mechanical hypersensitivity (allodynia) transmitted by the uninjured peripheral (sural) nerve. *In situ* hybridization of dorsal root ganglia (DRG) sections, where the cell bodies of sensory neurons that transmit the "pain" message to the spinal cord residue, revealed expression of both σ_1 and σ_2 receptors in many DRG neurons, including myelinated and unmyelinated subsets (**E.D. Fig. 4.4**). The expression of σ_1 or σ_2 did not change in DRG neurons seven days after SNI. When administered systemically to SNI mice, both σ_2-selective ligands (Z1665845742 and Z4857158944) were anti-allodynic, increasing mechanical thresholds versus vehicle (**Fig. 4.4b**, **E.D. Fig. 4.4**). This was comparable to the anti-allodynia conferred by a systemic injection of PD-144418, a σ_1-selective ligand. Intriguingly, systemic injection of the non-selective Z4446724338 dose-dependently increased the mechanical thresholds of SNI mice (**Fig. 4.4b**, **E.D. Fig. 4.4**) with the highest dose completely reversing the SNI-induced mechanical allodynia (i.e., thresholds returned to pre-injury levels), a meaningfully higher level of anti-allodynia than observed with the selective σ_2 ligands. Conversely, systemic injections of the non-selective PB28[8] produced mixed results, with anti-allodynic effects observed only in 60% of the mice **(E.D. Fig. 4.4)**. The much stronger anti-allodynia of Z4446724338 versus PB28 may reflect the former's substantially higher brain permeability (**E.D. Table 4.3**). Importantly, none of the new σ_1 and σ_2 ligands were

sedative on the rotarod test (**E.D. Fig. 4.4**), indicating that their anti-allodynic effect was not due to motor impairment.

The anti-allodynia of the σ_2-selective ligands Z1665845742 and Z4857158944 suggest that this receptor is a potential target for managing neuropathic pain. However, because σ_2 ligands are notoriously promiscuous, especially against GPCRs[40,41], we counter-screened the three docking-derived ligands against potential off-targets. In a TANGO screen[42] of 320 GPCRs, the molecules did not act as agonists or inverse agonists against most targets (**E.D. Fig. 4.5a-c**), and the few cases where activity was observed did not replicate in concentration-response assays (**E.D. Fig. 4.5d-f, Supplementary Fig. 4.2-4.3**). We also did not observe substantial activity at the μ-opioid receptor, a key pain target, in a G protein assay (**E.D. Fig. 4.5d**). We further screened the compounds in binding assays against a panel of 19 targets including GPCRs, ion channels, and transporters; no binding was observed for any pain-related targets (**Supplementary Table 4.2**). These observations suggest that the primary mechanism of action of these ligands is via the σ_2 receptor. The stronger activity of the $\sigma_{1/2}$ ligand Z4446724338 suggests that $\sigma_{1/2}$ polypharmacology may further increase anti-allodynia.

σ_2 ligand effects peak after 24 hours

In earlier studies, $\sigma_{1/2}$ ligands showed peak anti-allodynia up to 48 hours after dosing[3]. This unusual behavior was observed with ligands with mid-nanomolar potency and 9 to 14-fold selectivity vs. the σ_1 receptor. We further explored this with the selective ligands, Z4857158944 and Z1665845742, and the non-selective ligand, Z4446724338.

The molecules were tested post SNI, at 1, 24, and 48 hours after dosing. Supporting earlier reports, the anti-allodynia of the three new σ ligands increased over time, peaking 24 hours post-injection (**Fig. 4.4c and E.D. Fig. 4.6**). In contrast, the anti-allodynia of the selective σ_1 ligand PD-144418 was not sustained 24- or 48-hours post-injection. Furthermore, although the σ_2-selective compounds exhibited reduced anti-allodynia efficacy at early time points versus the non-selective ligand Z4446724338, all three compounds conferred similar antinociception by 24 hours. This long-term activity cannot be easily explained by pharmacokinetics, as the brain half-life of all three compounds suggests minimal exposure past 12 hours (**E.D. Table 4.3**). Rather, this time course may reflect longer term signaling or regulatory effects[3].

To investigate tolerance, we also examined the effects of repeated injections of two of the lead compounds, Z4446724338 and Z4857158944. The antinociceptive effect of Z4446724338 persisted for the first three test days, and decreased slightly on the fourth day (**E.D. Fig. 4.4c-d, E.D. Fig. 4.6c-d**). More tolerance was observed for compound Z4857158944; by the third injection, the antinociceptive effect was lost. Taken together, these results suggest that polypharmacology at the σ_1 and σ_2 receptor underlies an enhanced antinociceptive effect compared to selectivity for the σ_2.

4.4 Discussion

The σ_2 receptor has been enigmatic for 30 years. Its involvement in diverse biological processes and the lack of molecular data has clouded its biological role. Four key observations from this study begin to illuminate these issues. **First**, high-resolution

crystal structures of the σ_2 receptor complexed with roluperidone and with PB28 reveal a ligand binding site deeply embedded in the membrane (**Fig. 4.1a, b**), suggesting the possibility of a lipid as an endogenous ligand. The evolutionary connection of σ_2 to EBP and the structure of the receptor bound to cholesterol support an ability to recognize sterols. The structures explain the simple pharmacophore of σ_2 ligands—a cationic amine that ion-pairs with Asp29, while flanking hydrophobic and aromatic moieties are recognized by nearby aromatic residues. The structures also identify nearby polar residues, Gln77 and Thr110 that may aid in recognizing the hydroxyl moiety of sterols. These residues are rarely exploited by classic σ_2 ligands but may provide new selectivity determinants for ligand discovery (**Fig. 4.1c,d, and E.D. Fig. 4.1j**). **Second**, by testing 484 compounds across ranks from a library of 490 million docked, a quantitative relationship emerged between docking score and the likelihood of binding (**Fig. 4.2**). Crystal structures of docking-derived ligands confirmed the docking predictions (**Fig. 4.3a,b**). **Third**, among the top-ranking docking hits were 31 novel scaffolds with potent affinities ($K_i < 100$ nM) (**E.D. Table 4.2**). Optimization of two of these led to potent ligands with 47 to >250-fold selectivity for the σ_2 over the σ_1 receptor (**Supplementary Table 4.1**). **Fourth**, three potent new σ_2 chemotypes were tested for efficacy in a mouse model for neuropathic pain. All three were antiallodynic (**Fig. 4.4**). The expression pattern of the receptor and the activity of the σ_2-selective ligands confirm a contribution of this receptor in pain processing and suggest its potential relevance in pain management.

The σ_2 and the σ_1 receptors are promiscuous, both binding to cationic amphiphiles, leading to receptor cross-reactivity. Although many selective σ_1 ligands, like PD-144418

and (+)-pentazocine, have been described, there are far fewer selective ligands[4,43] for the σ$_2$ receptor. We sought to optimize for such selectivity[,22,44,45] using structure-based analoging, ultimately leading to two selective chemotypes. We combined one of these with a close analog that is σ$_2$ inactive, affording a "probe pair" (Z1665845742 and Z1665798906 available via Sigma-Millipore's probe collection, Cat. Nos. SML3141 and SML3142, respectively) (**Supplementary Fig. 4.8**). Such pairs can interrogate the role of the σ$_2$ receptor in indications for which it has been widely mooted, including cancer[1,19], schizophrenia[2], and Niemann-Pick disease[15,16], with the activity of the non-binding member controlling for inevitable off-targets.

The very promiscuity of the σ$_2$ receptor makes it a good template to investigate how docking score predicts binding likelihood, something only investigated once before at scale, with dopamine receptor[32]. As in that earlier study, a sigmoidal relationship between score and hit-rate emerged, here with hit rates peaking at over 60% (**Fig. 4.2b**). Unlike the dopamine receptor, which suffered from a long hit-rate plateau among the top-ranking molecules, σ$_2$ hit rates continued to rise with docking score through most of the curve. The exception was among a thin slice of the very top scoring molecules, where hit rates actually dropped owing to a subset of molecules that "cheat" the scoring function (**E.D. Fig. 4.2**), affording us the ability to improve it.

After completion of this study, a model of the σ$_2$ receptor was released as part of the AlphaFold protein structure prediction database[46]. This model closely resembles the crystal structures solved here, with an overall backbone RMSD of 0.5 Å (**Supplementary**

Fig. 4.4a). Importantly for ligand discovery, binding site residues have an all-atom RMSD value less than 2 Å (**Supplementary Fig. 4.4b**). Despite the high fidelity of the model to the experimental structure, the 484 new compounds from docking against the crystal structure scored relatively poorly against the AlphaFold model (**Supplementary Fig. 4.4c**), reflecting a slightly contracted pocket in the model. It may yet be true that other ligands could be found that fit the AlphaFold model well and bind to the receptor. To investigate this, new prospective docking will be informative.

Certain caveats bear airing. While our ultimate goal was to find σ_2-selective ligands, a spectrum of affinities and selectivities for both σ receptors emerged, reflecting the similarities of their pockets and their well-known overlapping pharmacology (**Fig. 4.1c-e**). The high hit rates and potencies found here reflect a site unusually well-suited to ligand binding, something unlikely to translate to other targets. While the docking-predicted pose for Z4857158944 and for Z1241145220 were confirmed crystallographically, the important water-bridging interaction for Z1241145220 was missed.

The key observations of this work should not be obscured by these caveats. The crystal structures of σ_2 receptors reveal the basis of its molecular recognition, and template structure-based campaigns for novel ligand discovery. From such campaigns emerged a predictive correlation between docking rank and likelihood of binding, and potent and selective σ_2 ligands that may be used to probe receptor biology.

4.5 Acknowledgements

experiments, supervised, and co-analyzed by A.I.B. M.J.O. conducted the Bayesian analysis of docking scores vs. hit rates; C.M.W. tested molecules for activity against the µOR. X.P.H. and Y.L. tested compounds against the GPCR-ome and other off-targets, with supervision from B.L.R. Y.S.M. supervised the synthesis of molecules from the virtual library, J.J.I. was responsible for the building of the version of the ZINC library that was docked. A.C.K., B.K.S, and A.I.B. supervised the project. The manuscript was written by A.A., J.L., B.K.S, and A.C.K. with input from other authors. **Competing interests:** A.C.K. is a founder and consultant for biotechnology companies Tectonic Therapeutic, Inc., and Seismic Therapeutic, Inc., as well as the Institute for Protein Innovation, a non-profit research institute. B.K.S. is a founder of Epiodyne, a company active in analgesia, and of BlueDolphin, which undertakes fee-for-service ligand-discovery.

4.6 Figures
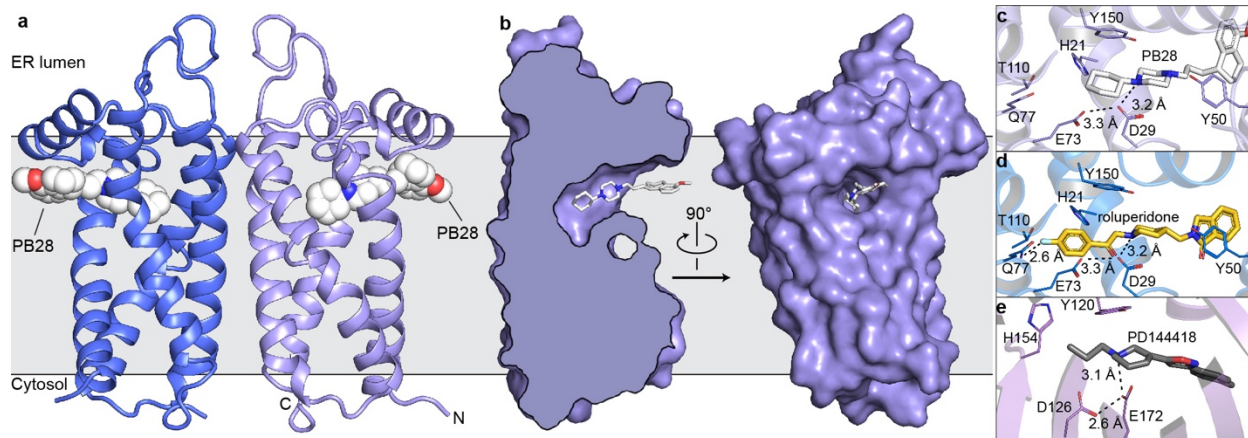
Figure 4.1. Structure of the σ₂ receptor and binding site ligand recognition.
 a, Structure of the σ₂ receptor bound to PB28. Amino- and carboxy-termini are indicated. Membrane boundaries were calculated using the PPM server[31]. **b**, Cross-section of the σ₂ receptor binding pocket (left) and view of the entrance to the binding pocket from the membrane (right). **c**, View of PB28 binding pose, showing charge–charge interaction with Asp29 (black dotted line) and contacts with other binding pocket residues. **d**, Analogous structure of the roluperidone binding pose. **e**, Structure of the σ₁ receptor bound to PD144418 (PDB ID: 5HK1). Amino acids that serve similar roles and positioned in a similar orientation to amino acids in the σ₂ receptor are indicated.
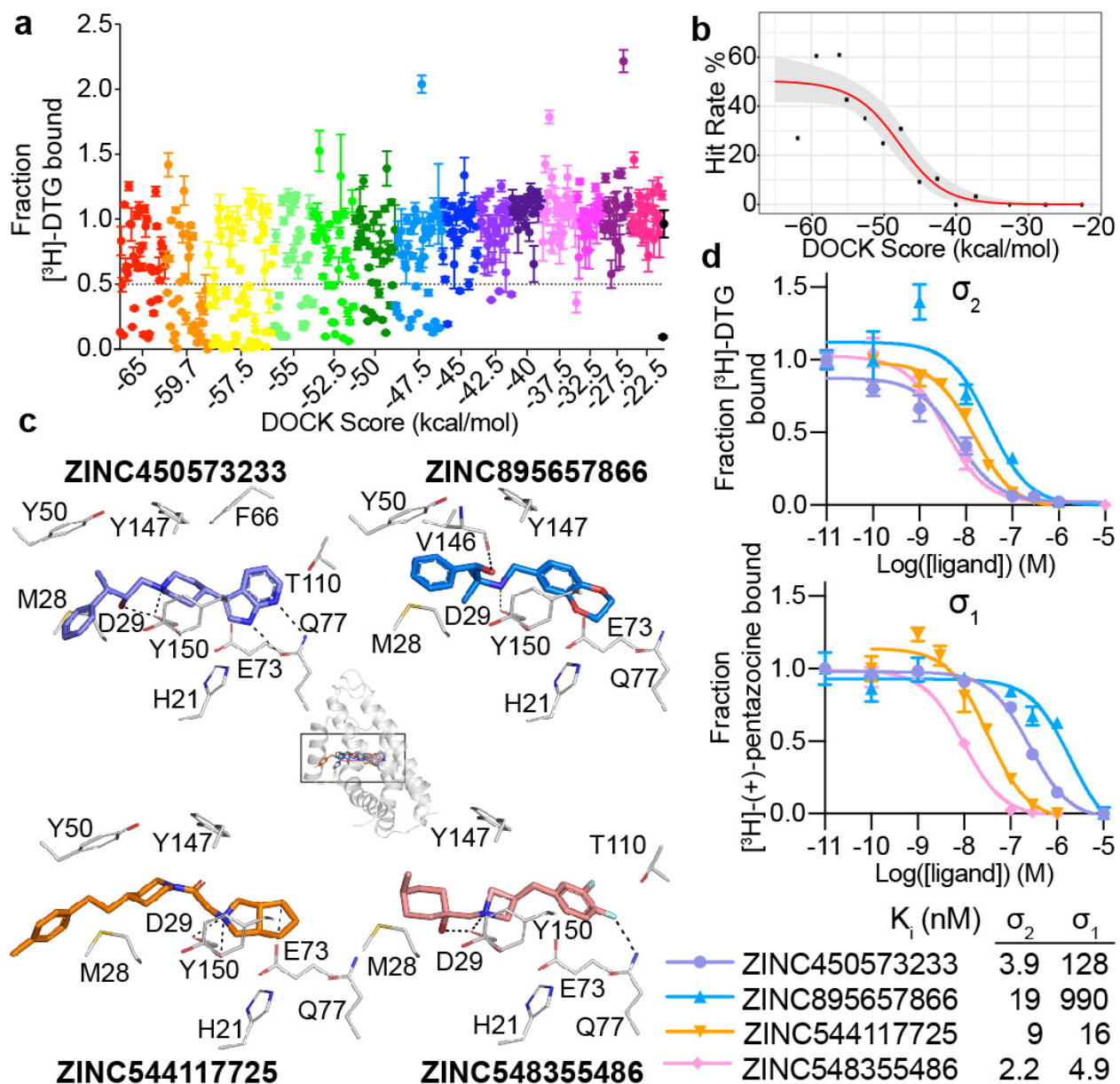
Figure 4.2. Docking 490 million molecules against the σ₂ receptor.

a, Displacement of the radioligand [³H]-DTG by each of the 484 molecules tested at 1 μM (mean ± SEM of three technical replicates). The molecules are colored and grouped by docking score. Dashed line indicates 50% radioligand displacement. Dots below the dashed line represent confirmed binders, whose numbers diminish with worsening docking score. **b,** The hit-rate of 484 experimentally tested compounds was plotted against docking energy. The docking score (dock$_{50}$) and slope at the maximum (slope$_{50}$) are -48 kcal mol^{-1} and -4.2% per kcal mol^{-1}, respectively. The gray band represents the 95% credible interval. **c,** Docked poses of four representative ligands from different scaffolds. **d**, Competition binding curves of the four molecules in **c.** against the σ₂ receptor (upper panel) and the σ₁ receptor (lower panel). The data are the mean ± SEM from three technical replicates.
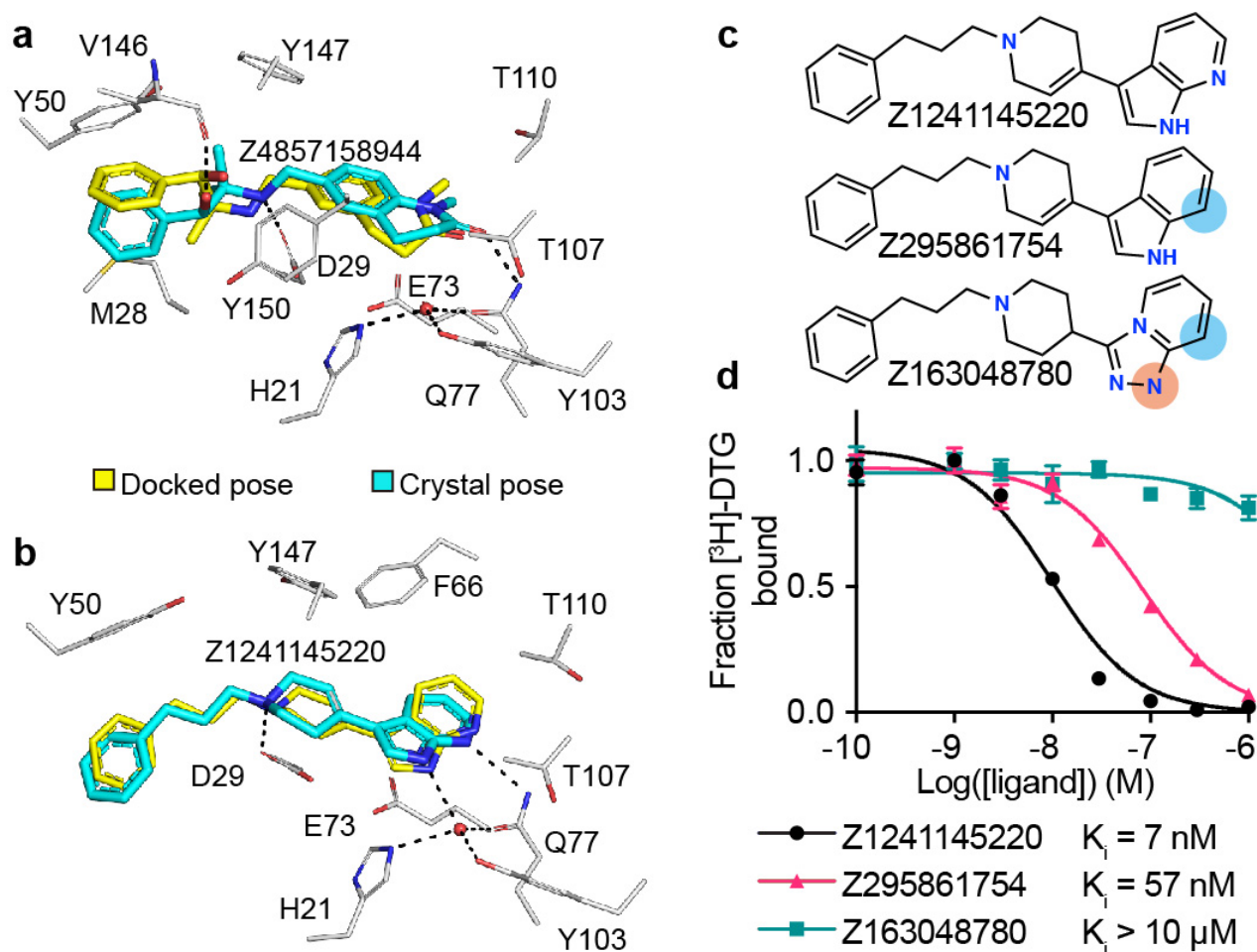
Figure 4.3. High structural fidelity between docked and crystallographic poses of novel σ₂ receptor ligands.

Ligand crystal poses (carbons in cyan) overlaid with respective docked poses (yellow). σ₂ receptor carbons are in grey, oxygens in red, nitrogens in blue, sulfurs in yellow, hydrogen bonds are shown as black dashed lines. **a**, Z4857158944-bound complex (PDB ID: 7M96; RMSD = 1.4 Å). **b,** Z1241145220-bound complex (PDB ID: 7M95; RMSD = 0.88 Å). **c,** Two Z1241145220 analogues that disrupt the hydrogen bonds with Gln77 and the structural water. Blue and apricot circles depict differences between the analogues and the parent compound. **d**, Competition binding curve of compounds from **c**. The data are the mean ± SEM from three technical replicates.
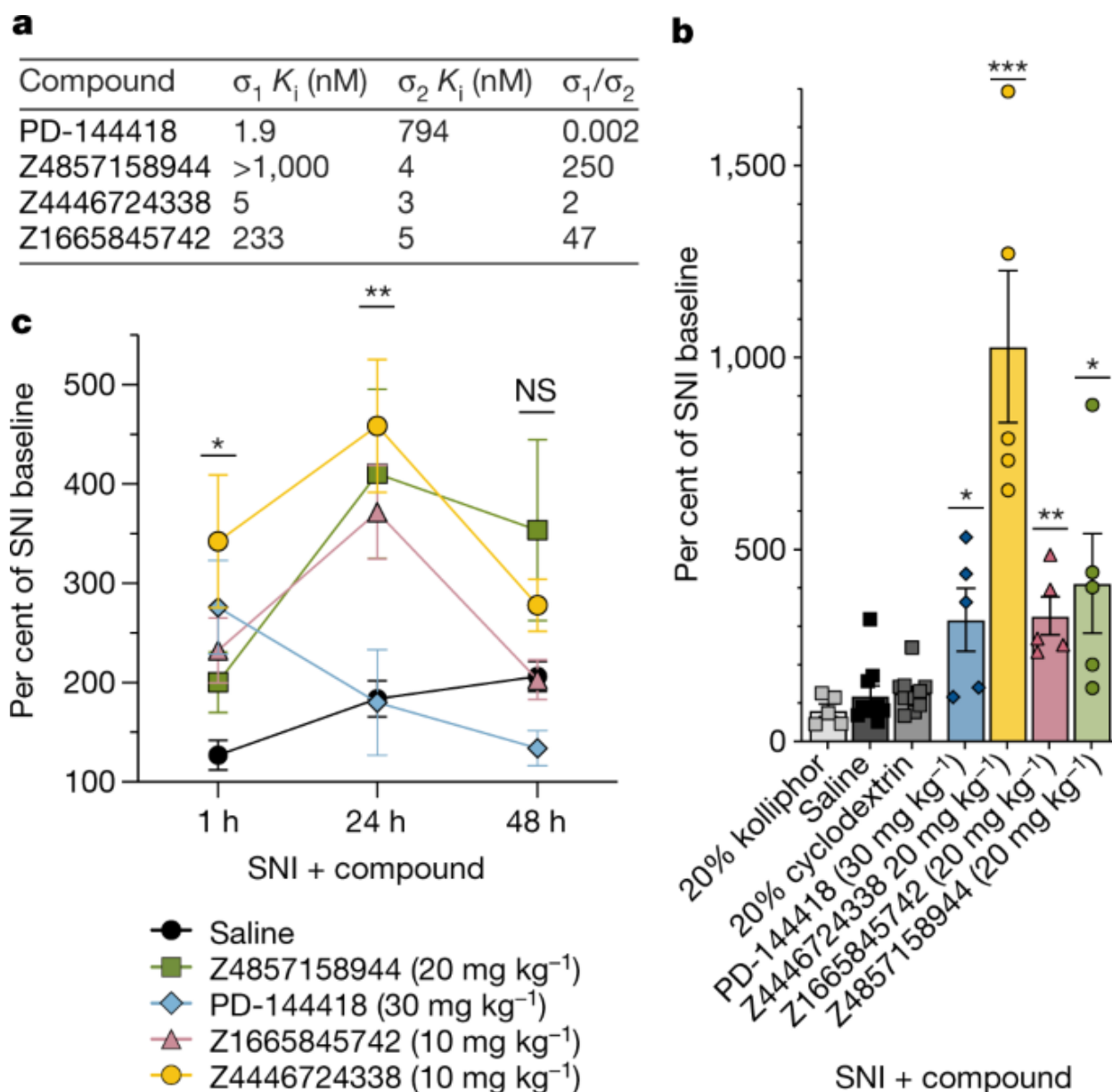
Figure 4.4. $\sigma_{1/2}$ ligands are anti-allodynic in a model of neuropathic pain.
a, Selectivity of four ligands at σ_1 and σ_2. PD-144418 values from the literature[47]. **b**, Response of mice to a von Frey filament after spared nerve injury (SNI). Ligands are compared to their vehicles (PD-144418 30 mg/kg ($n = 5$) vs. kolliphor ($n = 5$), one-way ANOVA, $F_{(2, 12)} = 7.49$, $p = 0.008$; Z4446724338 20 mg/kg ($n = 5$) vs cyclodextrin ($n = 10$), one-way ANOVA, $F_{(2, 22)} = 25.12$, $p = 0.0000021$; Z4857158944 20 mg/kg ($n = 5$) vs cyclodextrin ($n = 10$), one-way ANOVA, $F_{(2, 17)} = 5.10$, $p = 0.02$; Z1665845742 20 mg/kg ($n = 5$) vs saline ($n = 10$), one-way ANOVA, $F_{(3, 31)} = 6.18$, $p = 0.002$; asterisks define individual group differences to respective vehicle control using Dunnett's multiple comparisons Post-hoc test; kolliphor vs. PD-144418 30 mg/kg ($p = 0.009$); cyclodextrin vs. Z4446724338 20 mg/kg ($p < 0.001$); cyclodextrin vs. Z4857158944 20 mg/kg ($p = 0.01$); (Continued on the next page.)

158

(Continued from previous page.) saline vs. Z1665845742 20 mg/kg (p = 0.002); * p < 0.05, ** p < 0.01, *** p < 0.001). Data shown are mean ± SEM. Also see **E.D. Fig. 4.4a**. **c**, The anti-allodynic effects of σ_2, but not σ_1, ligands peak at 24 hours post-injection (two-way ANOVA; time x treatment interaction: $F(8,80)$ = 2.25, p = 0.03; time: $F(2,76)$ = 5.09, p = 0.009; treatment: $F(4,40)$ = 5.40, p = 0.001; four treatment groups (n = 10) except PD-144418 (n = 5); asterisks define difference between Z4446724338 and saline at 1 hr (p = 0.03), 24 hr (p = 0.008), and 48 hr (p = 0.11) for simplicity; ns = not significant, * p < 0.05, ** p < 0.01). Data shown are mean ± SEM.

a

Mass (kDa) / Rayleigh ratio (1/cm)
Retention (ml)

Total | Modifier (LMNG) | Protein
Control | 1 µM PB28 | 1 µM DTG | 1 µM haloperidol

b

```
Bovine   MGTLGARRGLEWFLGFYFLSHIPITLLMDLQGVLPRDLYPVELRNLQQWYIEEFKDPLLQ
Human    MGAPATRRCVEWLLGLYFLSHIPITLFMDLQAVLPRELYPVEFRNLLKWYAKEFKDPLLQ
         **:  .:** :**:**:***********:**** .****:*****:*** :** :********

Bovine   TPPAWFKSFLFCELVFQLPFFPIAAYAFFKGGCKWIRTPAIIYSVHTMTTLIPILSTLLL
Human    EPPAWFKSFLFCELVFQLPFFPIATYAFLKGSCKWIRTPAIIYSVHTMTTLIPILSTFLF
         ***************** *******:**:**.***********************:*.:

Bovine   DDFSKASHFRGQGPKTFQERLFLISVYIPYFLIPLILLLFMVRNPYYKSEEKRKKK
Human    EDFSKASGFKGQRPETLHERLTLVSVYAPYLLLIPFILLIFMLRSPYYKEEKRKKK
         :******* *:** *:*:::*** *:*** **:***:***:**:*.**** *******
```

c

Molar ellipticity [θ] (deg·cm²·dmol⁻¹)
Wavelength (nm)
Control | 12 µM PB28 | 12 µM Haloperidol | 12 µM DTG

d

Molar ellipticity [θ] (deg·cm²·dmol⁻¹)
Temperature (°C)
Control (54 °C) | 12 µM PB28 (75 °C) | 12 µM Haloperidol (64 °C) | 12 µM DTG (67 °C)

i

PB28
Roluperidone
Cholesterol
Z1241145220
Z4857158944

e

A_{280} (mAU)
Retention (ml)
1st run | 2nd run

f

M (kDa)
250, 150, 100, 75, 50, 37, 25, 20, 15, 10
σ₂

g

j

Y150 | H21 | Cholesterol | T110 | Q77 | 3.2 Å | E73 | D29 | Y50

h

D56 | Y50 | L58 | L59 | 2.9 Å | 2.9 Å | 2.7 Å | 2.7 Å | Y147

k

OD600
ERG2 | EBP | σ₁ | σ₂ | EBPL | TM6SF1 | TM6SF2 | Empty vector
without cycloheximide | 50 ng/ml cycloheximide

l

Relative abundance
Retention time (min)
zymostenol | lathosterol
zymostenol + lathosterol
EBP + lathosterol
EBP + zymostenol
σ₂ + lathosterol
σ₂ + zymostenol
zymostenol → lathosterol

160

E.D. Figure 4.1. Characterization of σ₂ receptor.
a, Size-exclusion chromatography with multi-angle light scattering of the human σ₂ receptor. The σ₂ receptor was run either without ligand or with 1 µM of the indicated ligand. Lines indicate calculated total mass (gray), detergent micelle (blue), and protein (purple). **b**, Sequence alignment between the human and bovine σ₂ protein sequences performed using T-coffee[48]. Residues that line the binding pocket are marked in red. **c**, Circular dichroism analysis of the bovine σ₂ receptor alone (black) or with the indicated ligand. Data is representative of multiple experiments. **d**, Circular dichroism melting curves of the bovine σ₂ receptor. Temperature was raised from 20 ºC to 90 ºC and molar ellipticity was measured at 222 nm. Protein was incubated either with or without indicated ligand at 12 µM. Melting temperatures for each measurement are indicated with a circle. Data is representative of multiple experiments **e**, Size-exclusion chromatography (SEC) of the bovine σ₂ receptor. Blue trace is after proteolytic tag removal. Red trace is protein applied on size exclusion after reapplying the tag-free protein on affinity resin to remove proteins with intact tags. The trace presented is representative of multiple purifications. **f**, Analysis of receptor purity after the second SEC using SDS-PAGE. Gray rectangle in **e** represents fractions chosen for analysis. The SDS-PAGE presented here is representative of multiple purifications. See **Source Data** for uncropped version. **g**, Crystals of bovine σ₂ receptor in the lipidic cubic phase. **h**, Aspartate 56 (D56) is important for receptor structure but not for ligand binding. A tight network of hydrogen bonds that bridges extracellular loop 1 to TM helix 4 is depicted with black dashed lines. **i**, Electron density maps for the various ligands. Polder maps[49] were calculated in Phenix. Maps are contoured at a level of 3 σ. **j**, View of cholesterol binding pose, showing contacts with other binding pocket residues. Hydrogen bonds are marked with black dashed lines. **k**, Yeast complementation assay. A ΔERG2 yeast strain was transformed with plasmids harboring the indicated genes. Yeast cells were grown to logarithmic phase and diluted to OD600 of 0.1, and then further diluted in a five-fold serial dilution series. Two microliters of each dilution were spotted on plates. Yeast cells were grown either in permissive conditions of no cycloheximide or in the restrictive conditions of 50 ng/ml cycloheximide, which requires functional Δ8-9 sterol isomerase activity for viability. ERG2 and EBP can act as sterol isomerases and rescue the growth of ΔERG2 *Saccharomyces cerevisiae* while the σ₂ receptor, the σ₁ receptor, or any other member of the EXPERA family cannot. **l**, EBP can catalyze the conversion of zymostenol to lathosterol while σ₂ cannot. Standards are in dark gray. EBP converts zymostenol to lathosterol (apricot) but does not convert lathosterol to zymostenol (dark red). The σ₂ receptor does not convert lathosterol to zymostenol (dark blue) or zymostenol to lathosterol (light purple). Structures of zymostenol and lathosterol are depicted below the traces.
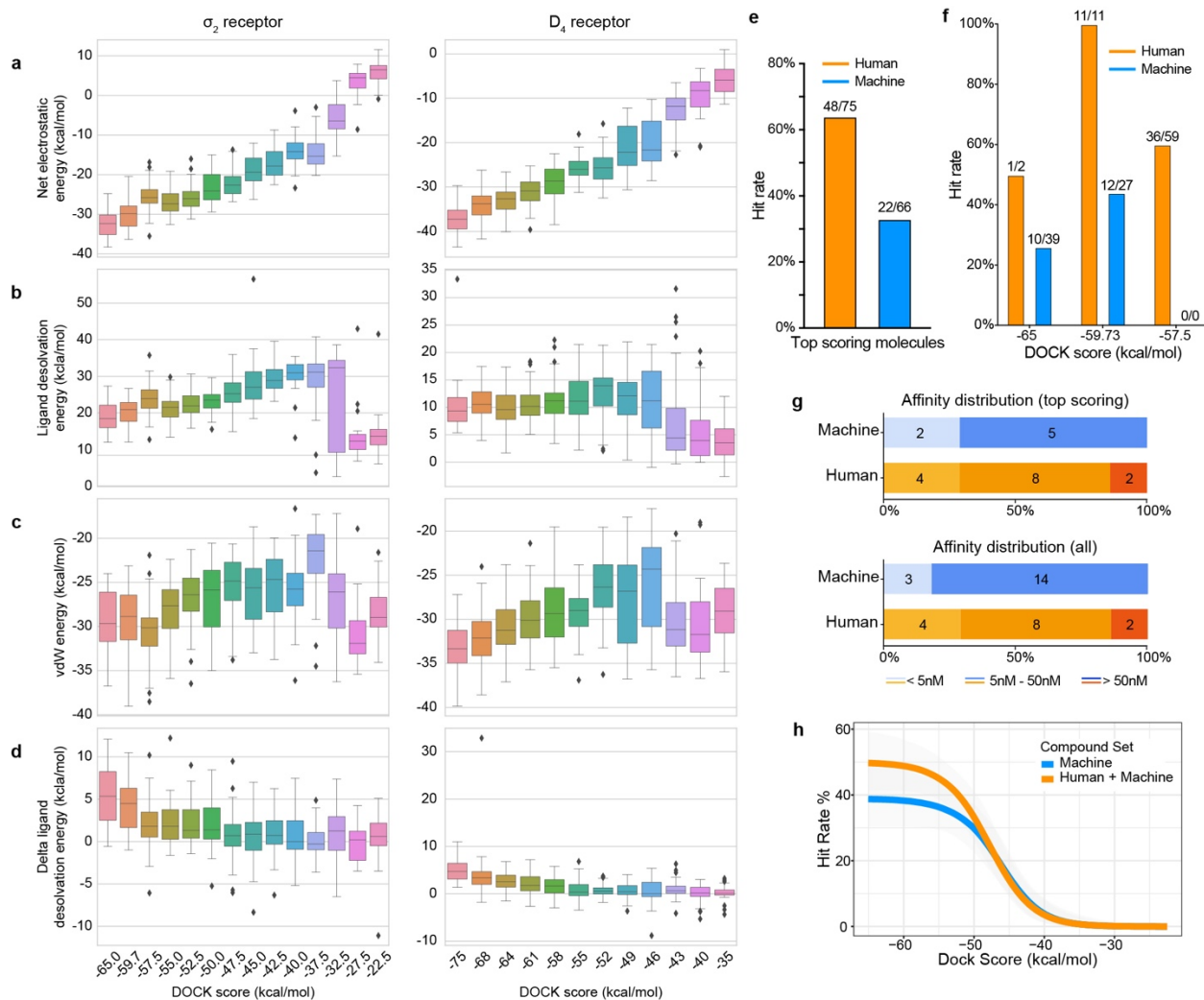
E.D. Figure 4.2. Comparisons of the distribution of docking scores.
a-d, The distribution of docking scores of tested molecules for hit rate curves against σ_2 (left column) and D_4 (right column) receptors. All tested molecules are grouped based on docking score bins. The distributions are shown in box plots for **a**, net electrostatic energy, **b,** ligand desolvation energy, **c,** van der Waals (vdW) energy and **d,** delta ligand desolvation energy after recalculating atomic desolvation energy based on the docked pose. **e-h**, Comparison of hit rates and affinities achieved by combined docking score and human inspection and these achieved by docking score alone. **e**, Overall hit rates for selecting compounds from the first 3 scoring bins by each strategy: human prioritization and docking score (orange), or docking score alone (blue). Hit rate is the ratio of active compounds/tested compounds; the raw numbers appear at the top of each bar. **f**, Hit rates for selecting compounds at different scoring ranges by each strategy: human prioritization and docking score (orange) or docking score alone (blue). **g**, Distribution of the binding affinity level among the hits from **e** (top panel). We measured competition binding curves for 14 docking hits from human prioritization and docking score, and 7 hits from the docking score alone. (Continued on the next page.)

(Continued from previous page.) These are divided into three affinity ranges: <5 nM; 5 nM–50 nM; >50 nM; Distribution of the binding affinity level among the hits from all different scoring ranges (bottom panel). We measured competition binding curves for 14 docking hits from human prioritization and docking score, and 17 hits from the docking score alone. **h**. Hit-rate curve comparison with/without human picks. The hit rate without human picks at the top plateau is 39% and at the bottom plateau is 0%, and the docking score ($dock_{50}$) and slope at the maximum ($slope_{50}$) are -46.5 kcal mol^{-1} and -3.5% per kcal mol^{-1}, respectively.

a

Fraction [³H]-DTG bound

ZINC548355486 K_i = 8 nM
Z4510628128 K_i = 14 nM
Z4430417979 K_i = 5 nM
Z4430418526 K_i = 3 nM

ZINC548355486

Z4510628128

Z4430417979

Z4430418526

b

Fraction [³H]-DTG bound

ZINC895657866 K_i = 70 nM
Z4857158944 K_i = 4 nM
Z4510627652 K_i = 14 nM
Z4510628191 K_i = 13 nM

ZINC895657866

Z4857158944

Z4510627652

Z4510628191

c

Fraction [³H]-DTG bound

ZINC450573233 K_i = 9.4 nM
Z1241145220 K_i = 3.7 nM
Z1665845742 K_i = 4.6 nM
Z4493929234 K_i = 3.9 nM

ZINC450573233

Z1241145220

Z1665845742

Z4493929234

d

H21
Z1241145220
Y17
3.3 Å
2.7 Å
2.7 Å
2.9 Å
3.2 Å
2.4 Å
Q77
Y103

e

Wildtype σ_2 K_d = 21 nM
Y17F K_d = 14 nM
Q77N K_d = 27 nM
H21Q K_d = 57 nM
H21N K_d = 40 nM
Y103F K_d = 120 nM

f

Wildtype σ_2 K_i = 8 nM
Y17F K_i = 5 nM
Q77N K_i = 77 nM
H21Q K_i = 102 nM
H21N K_i = 50 nM
Y103F K_i = 3 nM

164

E.D. Figure 4.3. Analogs of σ_2 receptor ligands and the effect of a structural water molecule.

a-c, Initial hits and selected analogs of σ_2 receptor ligands. Competition binding curves on the top panel, 2D drawings of compounds are on the bottom panel. Parent compound is indicated by gray background. Points shown as mean ± SEM from three technical replicates. **a**, Parent compound ZINC548355486 and its three potent analogues. **b**, Parent compound ZINC895657866 and its three potent analogues. **c**, Parent compound ZINC450573233 and its three potent analogues. **d-f**, The binding site of the σ_2 receptor contains a structural water. **d**, Water coordination at the binding site of the σ_2 receptor. Water molecule is depicted as a red sphere. Hydrogen bonds are indicated by black dashed lines. **e**, Saturation binding curve to measure the dissociation constant (K_d) of [^{3}H]-DTG for the various mutants of σ_2 receptor meant to disrupt water coordination. Residues proximal to the structural water were chosen for mutation. Residues were mutated to the indicated amino acid. Points shown as mean ± SEM from three technical replicates. **f**, Competition binding measurement of affinity of Z1241145220 in various mutants of σ_2. Points shown as mean ± SEM from three technical replicates.
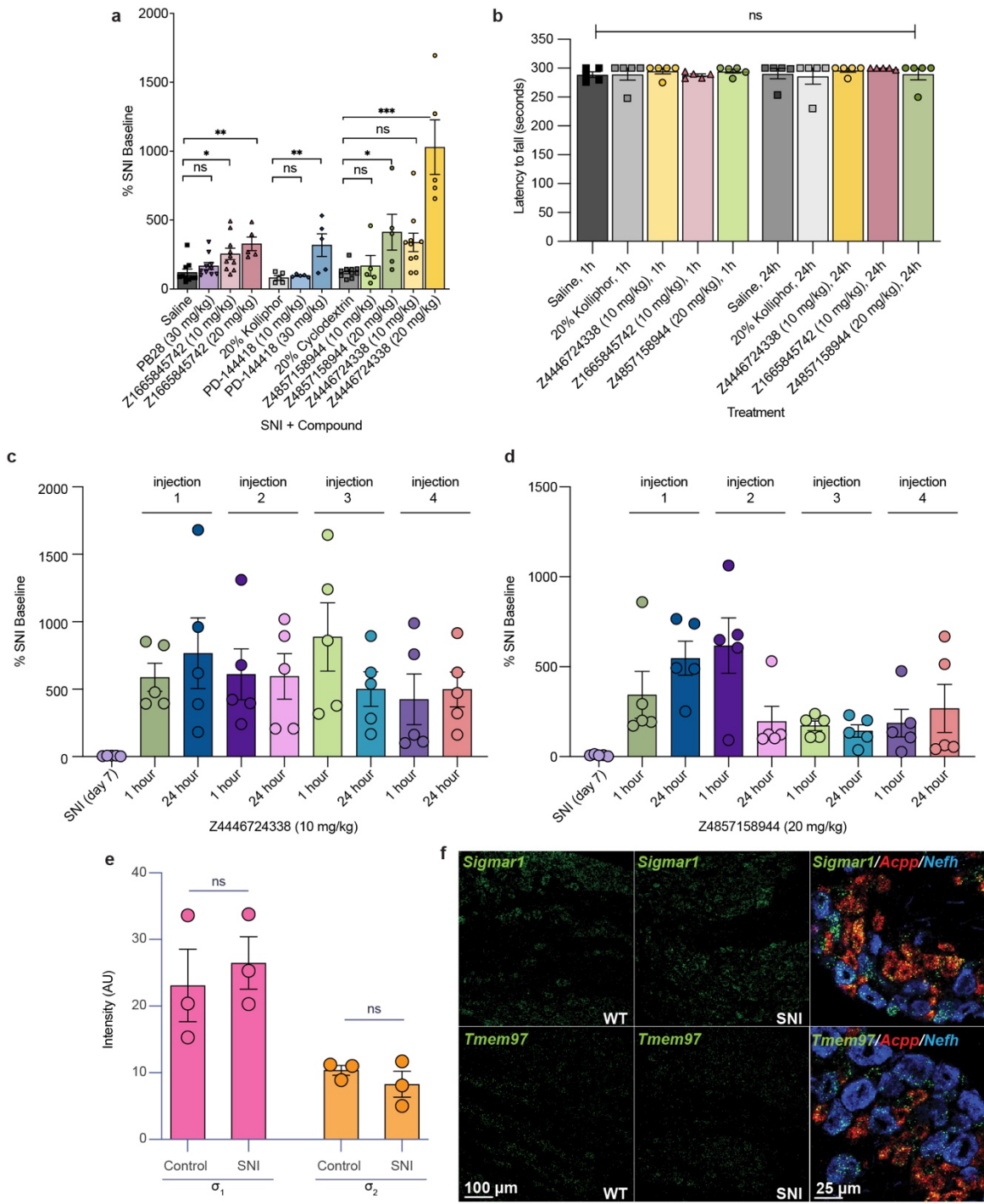
E.D. Figure 4.4. Effect of systemic σ receptor ligands on motor behavior.
a, Response of mice to a von Frey filament after spared nerve injury (SNI). All five ligands are compared to their respective vehicles (PD-144418 10 mg/kg (*n* = 5) and 30 mg/kg (*n* = 5) vs. kolliphor (*n* = 5), one-way ANOVA, F(2, 12) = 7.49, *p* = 0.008; Z4446724338 10 mg/kg (*n* = 10) (Continued on the next page.)

166

(Continued from previous page.) and 20 mg/kg (n = 5) vs cyclodextrin (n = 10), one-way ANOVA, $F(2, 22)$ = 25.12, $p < 0.001$; Z4857158944 10 mg/kg (n = 5) and 20 mg/kg (n = 5) vs cyclodextrin (n = 10), one-way ANOVA, $F(2, 17)$ = 5.10, $p = 0.02$; Z1665845742 10 mg/kg (n = 10) and 20 mg/kg (n = 5) and PB28 30 mg/kg (n = 10) vs saline (n = 10), one-way ANOVA, $F(3, 31)$ = 6.18, $p = 0.002$; asterisks define individual group differences to respective vehicle control using Dunnett's multiple comparisons Post-hoc test; ns = not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Data shown are mean ± SEM. Data for higher doses and vehicles is replotted from **Fig. 4.4**. **b,** No sedation or motor impairment on the rotarod was observed after drug treatments compared to vehicle at 1 hour (Z1665845742 10 mg/kg (n = 5) and Z4857158944 20 mg/kg (n = 5) vs saline (n = 5), one-way ANOVA, $F(2, 12)$ = 1.04, $p = 0.38$; Z4446724338 10 mg/kg (n = 5) vs kolliphor (n = 5), unpaired two-tailed Student's t-test, $t(8)$ = 0.47, $p = 0.65$) or 24 hours post-injection (Z1665845742 10 mg/kg (n = 5) and Z4857158944 20 mg/kg (n = 5) vs saline (n = 5), one-way ANOVA, $F(2, 12)$ = 0.45, $p = 0.65$; Z4446724338 10 mg/kg (n = 5) vs kolliphor (n = 5), unpaired two-tailed Student's t-test, $t(8)$ = 0.72, $p = 0.49$); ns = not significant. Data shown are means ± SEM. **c,** Response of SNI mice to a von Frey filament after repeated injections of Z4446724338 10 mg/kg (n = 5). Mechanical thresholds were assessed 1 hour and 24 hours after four separate injections. Data shown are means ± SEM normalized to each mouse's SNI baseline. **d**, Response of SNI mice to a von Frey filament after repeated injections of Z4857158944 10 mg/kg (n = 5). Mechanical thresholds were assessed 1 hour and 24 hours after four separate injections. Data shown are means ± SEM normalized to each mouse's SNI baseline. **e.** Quantification of the expression levels of Sigmar1 (σ_1) and Tmem97 (σ_2) in wildtype (WT) and SNI mice detected by *in situ* hybridization (n = 3 mice per group). Representative images can be found in panel **f**. Data shown are mean ± SEM; unpaired two-tailed Student's t-test— Sigmar1: $t(4)$ = 0.5, p = 0.64; Tmem97: $t(4)$ = 1.0, p = 0.37; ns = not significant. AU = arbitrary units. **f**, *in situ* hybridization of mouse dorsal root ganglion (DRG) sections for *Sigmar1* (σ_1) and *Tmem97* (σ_2) genes illustrates expression in myelinated (Nefh-positive; blue) and unmyelinated (Acpp-positive; red) subsets of sensory neurons and no change after SNI.
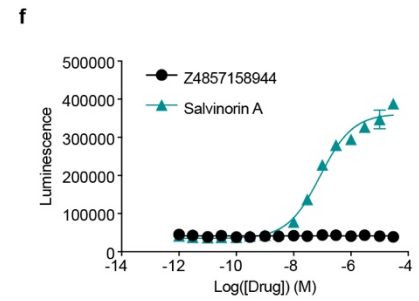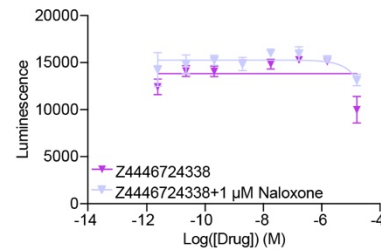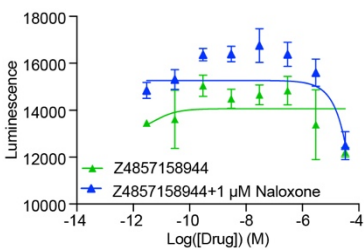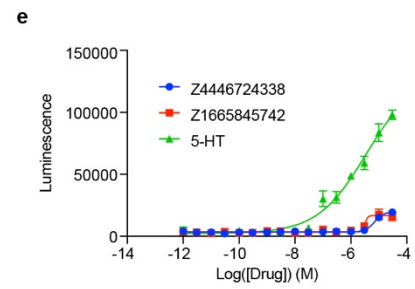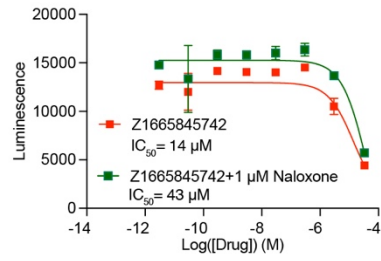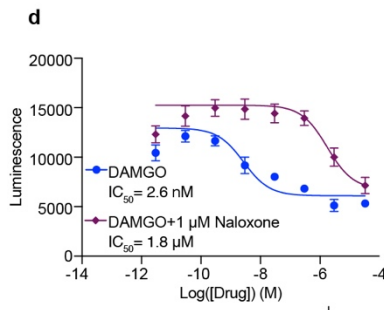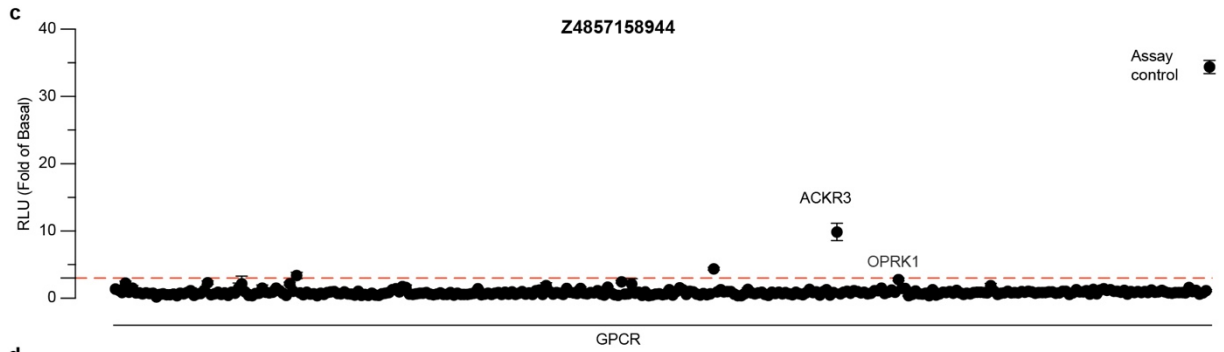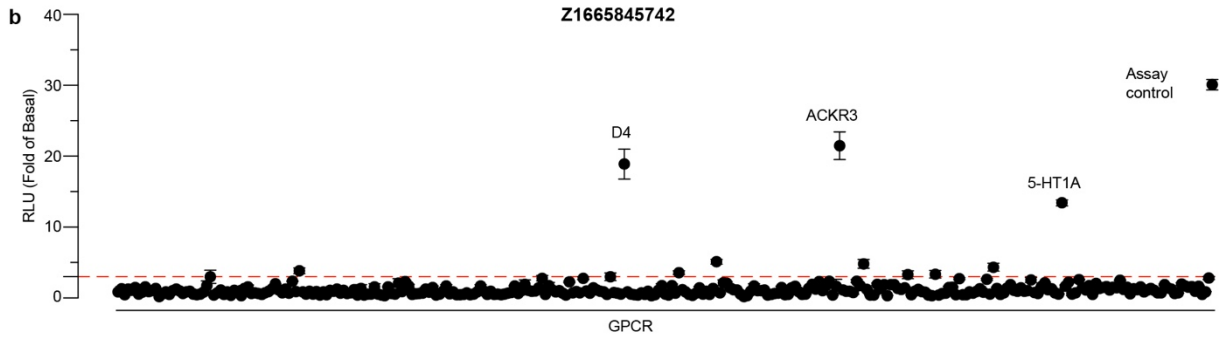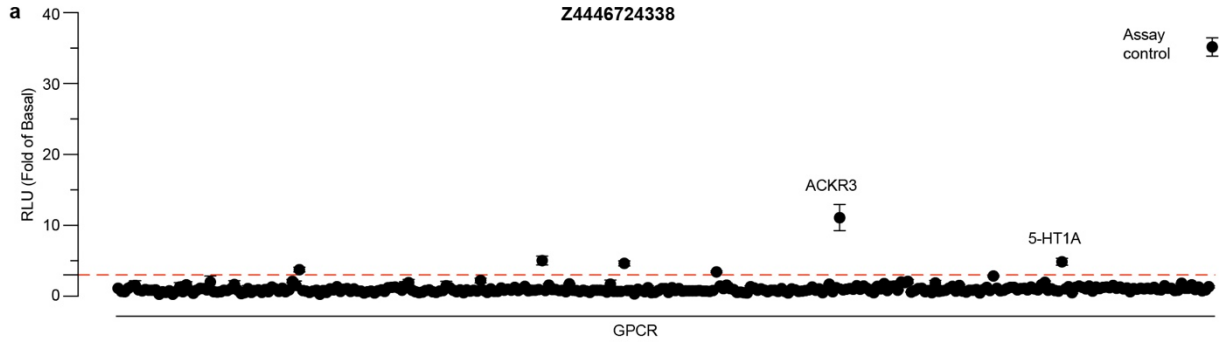
E.D. Figure 4.5. Off-target profiling of Z4446724338, Z1665845742, and Z4857158944.

a-c, TANGO screens against a panel of 320 GPCRs for the indicated σ$_2$ ligand. **a**, Z4446724338, **b**, Z1665845742, **c**, Z4857158944. **d**, GloSensor μOR-mediated cAMP inhibition (G$_i$ activation) by DAMGO, Z4446724338, Z1665845742, and Z4857158944. **e-f**, Follow-up does-response curves for pain-related receptors that showed activation in **a-c**. **e**, Z4446724338 and Z1665845742 against 5HT1A. **f**, Z4857158944 against κOR. Data shown are means ± SEM.
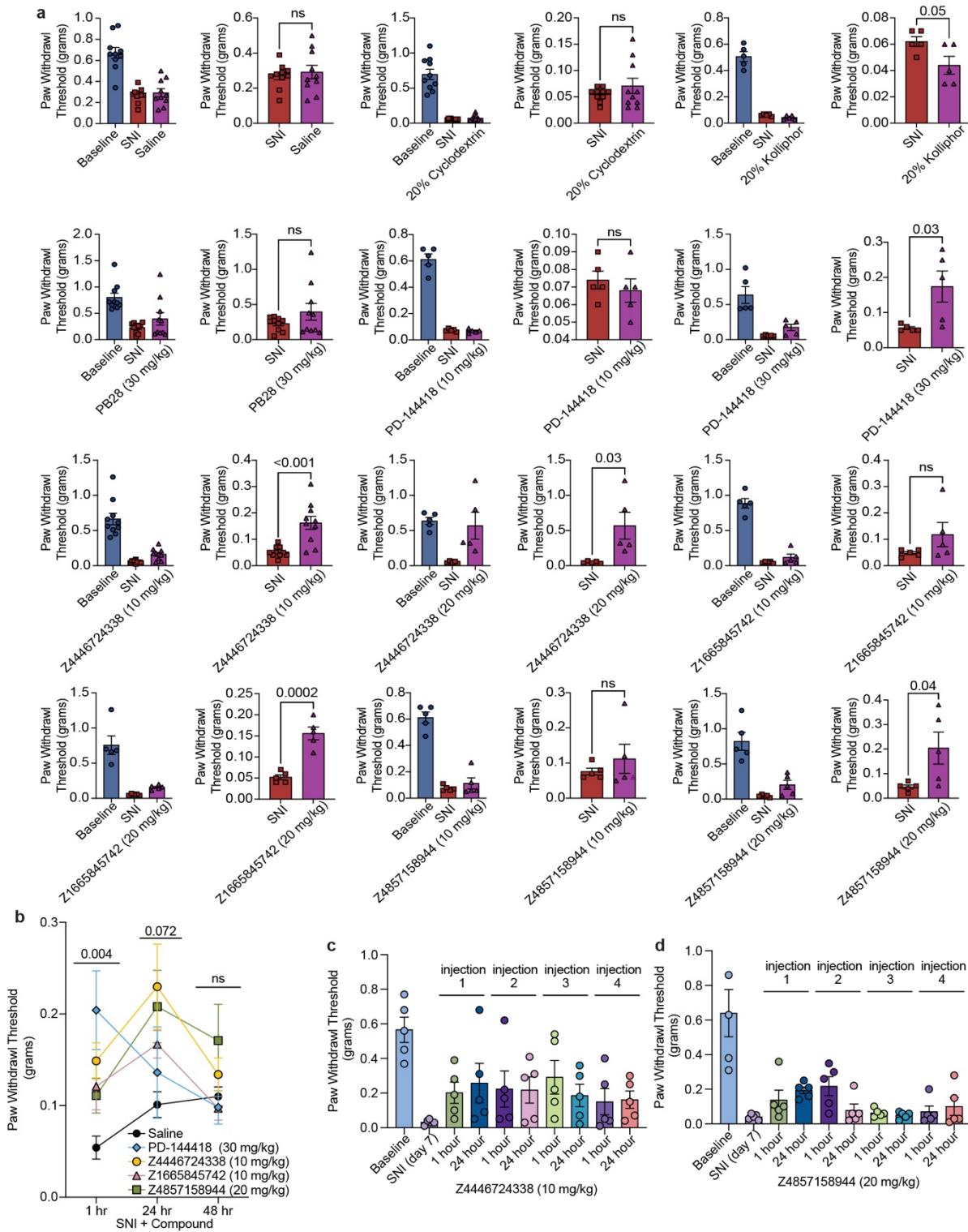
E.D. Figure 4.6. Paw withdrawal thresholds.
a, Paw withdrawal thresholds (PWT) before (blue bar) and after (red bar) spared nerve injury (SNI), as well as after SNI + treatment (purple bar). (Continued on the next page.)

(Continued from previous page.) For easier visualization of individual data points, data was also plotted without the pre-SNI baseline. Data are the same as in **Figure 4.4b** and **E.D. Fig. 4.4a**, but without the normalization to the individual post-SNI baselines and are expressed as mean ± SEM; mice per group: saline (n = 10); cyclodextrin (n = 10); kolliphor (n = 5); PB28 30 mg/kg (n = 10); PD-144418 10 mg/kg (n = 5) and 30 mg/kg (n = 5); Z4446724338 10 mg/kg (n = 10) and 20 mg/kg (n = 5); Z1665845742 10 mg/kg (n = 5) and 20 mg/kg (n = 5); Z4857158944 10 mg/kg (n = 5) and 20 mg/kg (n = 5); unpaired two-tailed Student's t-test. **b**, PWTs 1 hour, 24 hours, and 48 hours after saline or drug treatment. Data are the same as in **Figure 4.4c**, but without the normalization to the individual post-SNI baselines, and are expressed as mean ± SEM. Significance levels determined using Dunnett's multiple comparisons Post-hoc test reflect the difference between Z4446724338 and saline for simplicity (two-way ANOVA; time x treatment interaction: $F(8, 80)$ = 2.4, p = 0.02; time: $F(2, 74)$ = 5.2, p = 0.009; treatment: $F(4, 40)$ = 3.3, p = 0.02; four treatment groups (n = 10) except PD-144418 (n = 5); ns = not significant. **c**, Response of SNI mice to a von Frey filament after repeated injections of Z4446724338 10 mg/kg (n = 5). Mechanical thresholds were assessed 1 hour and 24 hours after four separate injections. Data shown are paw withdrawal thresholds in grams, expressed as mean ± SEM. **d**, Response of SNI mice to a von Frey filament after repeated injections of Z4857158944 10 mg/kg (n = 5). Mechanical thresholds were assessed 1 hour and 24 hours after four separate injections. Data shown are paw withdrawal thresholds in grams, expressed as mean ± SEM.
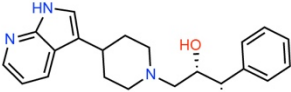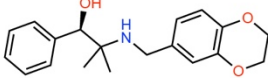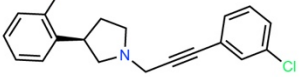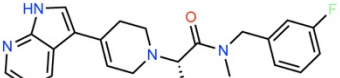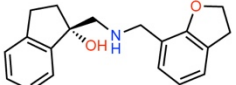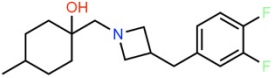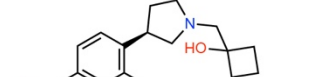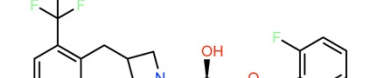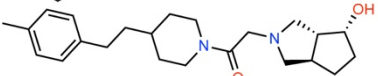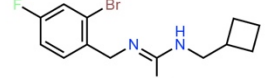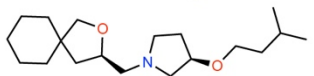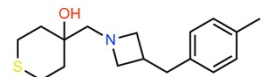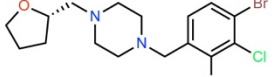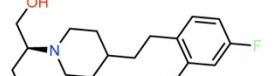
Note: **Supplementary Figures 4.1-4.8** can be found in the online version of the manuscript.

4.7 Tables

E.D. Table 4.1. Data collection and refinement statistics.

	PB28-bound (Se-labeled)	Roluperidone-bound (Native)	Z1241145220-bound (Native)	Z4857158944-bound (Native)	Cholesterol-bound (Native)
Data collection					
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_1$
Number of crystals	1	1	1	1	1
Cell dimensions					
a, b, c (Å)	70.6, 55.2, 93.0	69.1, 54.2, 99.7	55.4, 61.5, 110.4	70.7, 55.4, 93.0	70.9, 54.9, 93.0
α, β, γ (°)	90, 95.0, 90	90, 91.1, 90	90, 90, 90	90, 94.5, 90	90, 94.3, 90
Wavelength (Å)	1.255	1.03320	1.03321	1.033167	1.03320
Resolution (Å)	33.88 - 2.942	42.61 - 2.71	49.5 - 2.41	40.2 - 2.41	47.24 - 2.8
	(3.047 - 2.942)	(2.81 - 2.71)	(2.55 - 2.41)	(2.55 - 2.41)	(2.9 - 2.8)
R_{sym}	24.75 (88.16)	26.11 (205.9)	18.4 (177.4)	19.67 (227.6)	32.2 (220.2)
$I/\sigma I$	5.73 (0.93)	5.90 (0.71)	7.50 (0.7)	5.18 (0.56)	5.35 (0.94)
Completeness (%)	98.67 (90.76)	99.54 (99.87)	99.55 (99.36)	97.9 (88.3)	99.7 (98.5)
Redundancy	4.0 (3.4)	6.8 (6.5)	6.2 (4.4)	4.4 (4.5)	6.7 (6.9)
$CC_{1/2}$	98.7 (49.5)	99.4 (36.6)	99.6 (26.9)	99.5 (28.3)	99.2 (42.6)
Refinement					
Resolution (Å)	2.94	2.71	2.41	2.41	2.8
No. reflections	15228	20340	15165	27448	17720
No. reflection used for R_{free}	1524 (10%)	2004 (9.85%)	1063 (7%)	1370 (5%)	1752 (9.89%)
R_{work} / R_{free}	20.39 / 24.26	22.18 / 25.2	21.36 / 24.6	25.0 / 28.8	24.06 / 27.81
No. atoms					
Protein	5490	5472	2761	5393	5292
Lipid/ion	231	250	148	231	223
Ligand	108	108	48	100	112
Water	27	7	46	37	21
B-factors (Å^2)					
Protein	49.68	67.89	50.19	57.24	64.29
Lipid/ion	52.48	65.57	59.49	62.32	62.33
Ligand	56.89	79.39	49.32	66.07	71.96
Water	45.66	62.72	57.08	57.77	61.02
R.m.s. deviations					
Bond lengths (Å)	0.003	0.003	0.005	0.003	0.003
Bond angles (°)	0.61	0.61	1.04	0.58	0.59

E.D. Table 4.2. Fourteen of the highest-affinity direct docking hits for the σ_2 receptor.

2D drawing	ZINC ID	Rank	DOCK score (kcal/mol)	TC*	K_i (nM) σ_2	K_i (nM) σ_1	Selectivity (σ_1/σ_2)
	ZINC000450573233	4429	-57.25	0.32	4.3	128	30
	ZINC000895657866	19047	-55.35	0.31	21.4	989.6	46
	ZINC001170548029	4945	-57.11	0.35	22.6	727.2	32
	ZINC000533478938	18545	-55.38	0.30	34.5	1470	43
	ZINC000921927365	983	-59.01	0.31	67.3	1186	18
	ZINC000548355486	7007	-56.68	0.29	2.4	4.9	2
	ZINC000348332392	931	-59.07	0.28	33.7	2.9	0.1
	ZINC001254761628	16059	-55.58	0.27	4.7	53	11.3
	ZINC000544117725	3522	-57.52	0.28	10	16.25	1.6
	ZINC000170908795	13281	-55.84	0.29	6.7	32.7	4.9
	ZINC001196519317	9290	-56.3	0.29	2.4	13.4	5.6
	ZINC000656714762	1276	-58.68	0.26	67.8	4.6	0.1
	ZINC001237901728	11409	-56.03	0.30	27	1.6	0.1
	ZINC001460312963	11817	-55.99	0.29	5.2	1.7	0.3

* TC, Tanimoto coefficient to sigma ligands from ChEMBL.

E.D. Table 4.3. Measured pharmacokinetic parameters for PB28, Z1665845742, Z4446724338 and Z4857158944 in male CD-1 mice by 10 mg/kg subcutaneous administration.

				Pharmacokinetic Parameters			
Type	Name	T_{max} min	C_{max} ng/ml (g)	$AUC_{0\to t}$ (AUC_{last}) ng*min/ml (g)	$AUC_{0\to\infty}$ (AUC_{inf_obs}) ng*min/ml (g)	$T_{1/2}$ (HL_Lambda_z), min	K_{el} (Lambda_z), min^{-1}
Plasma	Z1665845742	20	968	99000	112000	185	0.00374
	Z4446724338	20	449	58300	60500	47.4	0.0146
	Z4857158944	20	228	13300	14200	27.8	0.0249
	PB28	60	42	8640	45900	740	0.000937
Brain	Z1665845742	20	3150	436000	509000	747	0.000928
	Z4446724338	20	7390	1140000	1150000	69.7	0.00995
	Z4857158944	20	2960	247000	327000	452	0.00153
	PB28	60	948	229000	240000	98.1	0.00706

Note: **Supplementary Tables 4.1-4.5** can be found in the published version of the

manuscript.

4.8 Materials and Methods

Protein expression and purification for crystallography. The bovine σ_2 receptor was cloned into pVL1392 with an N-terminal human protein C epitope tag followed by a 3C protease cleavage site. The construct was truncated after residue 168 to exclude the ER localization signal for better expression and to facilitate crystallization. This receptor construct was expressed in *Sf9* insect cells (Expression Systems) using the BestBac baculovirus system (Expression Systems) according to manufacturer's instruction. Infection was performed when cell density reached 4×10^6 cells per milliliter. Cells were shaken at 27 °C for 60 hours before harvest by centrifugation. Cell pellets were stored at -80 °C until purification.

During all purification steps ligands (PB28, roluperidone, Z1241145220, and Z4857158944) were present in all buffers at 1 µM. For the cholesterol-bound structure the protein was purified in the presence of 1 µM DTG. Cell paste was thawed and cells were disrupted by osmotic shock in 20 mM HEPES pH 8, 2 mM magnesium chloride, 1:100,000 (v:v) benzonase nuclease (Sigma Aldrich), and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). Lysed cells were centrifuged at 50,000 x g for 15 minutes. Following centrifugation, supernatant was discarded, and the membrane pellets were solubilized with a glass Dounce tissue homogenizer in 20 mM HEPES pH 8, 250 mM NaCl, 10% (v/v) glycerol, 1% (w/v) lauryl maltose neopentyl glycol (LMNG; Anatrace), and 0.1% (w/v) cholesterol hemisuccinate (CHS; Steraloids). Samples were stirred at 4 °C for 2 hours and then non-solubilized material was removed by centrifugation at 50,000 x g for 30 min. Supernatant was supplemented with 2 mM calcium chloride and filtered by a

glass microfiber filter (VWR). Samples were then loaded by gravity flow onto 5 ml anti-protein C antibody affinity resin. Resin was washed with 10 column volumes of 20 mM HEPES pH 8, 250 mM NaCl, 2 mM calcium chloride, 1% (v/v) glycerol, 0.1% (w/v) LMNG, and 0.01% (w/v) CHS, and then with 10 column volumes of 20 mM HEPES pH 8, 250 mM NaCl, 2 mM calcium chloride, 0.1% (v/v) glycerol, 0.01% (w/v) LMNG, and 0.001% (w/v) CHS. The receptor was eluted with buffer containing 20 mM HEPES pH 8, 250 mM NaCl, 5 mM EDTA, 0.1% (v/v) glycerol, 0.01% (w/v) LMNG, 0.001% (w/v) CHS, and 0.2 mg/ml protein C peptide, in 1 ml fractions. Peak fractions were pulled and 3C protease was added (1:100 w:w) and incubated with the receptor at 4 °C overnight. Next the receptor was purified by size exclusion chromatography on a Sephadex S200 column (Cytiva) in 20 mM HEPES pH 8, 250 mM NaCl, 0.1% glycerol, 0.01% LMNG, and 0.001% CHS. Peak fractions were pulled, calcium chloride was added to 2 mM and the sample was reapplied on the anti-protein C resin to remove uncleaved receptor. The column was washed with 5 column volumes and flow-through and wash fractions were pulled, concentrated, and reapplied on SEC. Peak fractions were pulled, concentrated to 50 mg/ml, and aliquoted. Protein aliquots were flash frozen in liquid nitrogen and stored in -80 °C until use. Purity was evaluated by SDS-PAGE.

Crystallography and data collection. Purified σ_2 receptor was reconstituted into lipidic cubic phase (LCP) by mixing with a 10:1 (w:w) mix of monoolein (Hampton Research) with cholesterol (Sigma Aldrich) at a ratio of 1.5:1.0 lipid:protein by mass, using the coupled syringe reconstitution method[25]. All samples were mixed at least 100 times. The resulting phase was dispensed in 30–40 nl drops onto a hanging drop cover and

overlaid with 800 nl of precipitant solution using a Gryphon LCP robot (Art Robbins Instruments). The PB28-bound crystals grew in 20–30% PEG 300, 0.1 M MES pH 6, 600 mM NaCl. The Roluperidone-bound crystals grew in 20% PEG 300, 0.1 M MES pH 6, 500 mM NaCl, 60 mM succinate. The Z1241145220-bound crystals grew in 30% PEG 300, 0.1 M MES pH 6, 210 mM ammonium phosphate. The Z4857158944-bound crystals grew in 30% PEG 300, 0.1 M MES pH 6, 560 mM ammonium phosphate. The cholesterol-bound crystals grew in 25% PEG300, 0.1 M MES pH 6, 400 mM sodium citrate, and 1% 1,2,3-heptanetriol. All crystals grew in the presence of 1 µM of ligand, except for the cholesterol structure, which had no ligand present during crystal growth. Crystals were harvested using either MicroLoops LD or mesh loops (MiTeGen) and stored in liquid nitrogen until data collection. Data collection was performed at Advanced Photon Source GM/CA beamlines 23ID-B and 23ID-D. Data collection used a 10 µm beam and diffraction images were collected in 0.2° oscillations at a wavelength of 1.254858 Å for the PB28-bound crystals and a wavelength of 1.033167 Å for all other crystals. A complete data set was obtained from a single crystal in each case.

Data reduction and refinement. Diffraction data were processed in HKL2000[50] and in XDS[51], and statistics are summarized in Table 1. The PB28-bound structure was solved using molecular replacement starting with a Rosetta[52] homology model generated using the structure of EBP (Protein Data Bank accession 6OHT). Matthews probability predicted four copies in the asymmetric unit. Initially, a single copy of this model was placed using Phaser[53] giving a marginally interpretable electron density map. This model did not fit well into density and was replaced with Idealized helices that were used as a

search model for an additional copy. The resulting dimer was duplicated and manually placed into unmodeled density. The resulting structure was iteratively refined in Phenix[54] and manually rebuilt in Coot[55]. Final refinement statistics are summarized in **E.D. Table 4.1**. The PB28 structure was used as a model for molecular replacement for all other datasets. In the case of the structure modeled as cholesterol-bound, electron density for a sterol-shaped ligand was observed **E.D. Fig. 4.1i**) and tentatively modeled as cholesterol based on the high (millimolar) concentration of cholesterol in the crystallization conditions and the compatibility of cholesterol with the shape of the electron density in the binding pocket. The receptor was purified in the presence of ditolylguanidine (DTG), but no DTG was present in the precipitating solution, and electron density was clearly incompatible with bound DTG. We cannot exclude the possibility that some other compound structurally similar to cholesterol was carried through the purification and is the ligand observed in the binding pocket. Figures containing electron density or structures were prepared in PyMOL[56] v2.5 or UCSF Chimera[57] v1.15.

Preparation of membranes for radioligand binding. The human σ_2 receptor was cloned into pcDNA3.1 (Invitrogen) mammalian expression vector with an amino-terminal protein C tag followed with a 3C protease cleavage site. Mutations were introduced by Site-directed mutagenesis using HiFi HotStart DNA Polymerase (Kapa Biosystems). Expi293 cells were transfected using FectoPRO (Polyplus-transfection) according to manufacturer instruction. Cells were harvested by centrifugation and lysed by osmotic shock in a buffer containing 20 mM HEPES, pH 7.5, 2 mM MgCl2,1:100,000 (vol/vol) benzonase nuclease (Sigma Aldrich), and cOmplete Mini EDTA-free protease-

inhibitor tablets (Sigma Aldrich). The lysates were homogenized with a glass dounce tissue homogenizer and then centrifuged at 20,000 x g for 20 min. After centrifugation, the membranes were resuspended in 50 mM Tris, pH 8.0, divided into 100 µL aliquots, flash frozen in liquid nitrogen, and stored at –80 °C until use.

Saturation and competition binding in Expi293 membranes. Saturation binding was performed with a method similar to that of Chu and Ruoho[58]. Briefly, membrane samples from Expi293 cells (Thermo Fisher Scientific) expressing wild-type or mutant σ_2 receptor, prepared as described above, were thawed, homogenized with a glass dounce, and diluted in 50 mM Tris, pH 8.0. Binding reactions were done in 100 µL, with 50 mM Tris pH 8.0, [^{3}H]-DTG (PerkinElmer), and supplemented with 0.1% bovine serum albumin to minimize non-specific binding. To assay non-specific binding, equivalent reactions containing 10 µM haloperidol were performed in parallel. Competition assays were performed in a similar fashion with 10 nM [^{3}H]-DTG and the indicated concentration of the competing ligand. Samples were shaken at 37 °C for 90 min. Afterward, the reaction was terminated by massive dilution and filtration over a glass microfiber filter with a Brandel harvester. Filters were soaked with 0.3% polyethyleneimine for at least 30 min before use. Radioactivity was measured by liquid scintillation counting. Data analysis was done in GraphPad Prism 9.0, with K_i values calculated by Cheng-Prusoff correction using the experimentally measured probe dissociation constant.

Circular dichroism. Far-UV circular dichroism (CD) spectra (185–260 nm) were measured with a JASCO J-815 (JASCO Inc., Tokyo, Japan), with a Peltier temperature

controller and single cuvette holder and Spectra Manager II software for data collection and analysis. Data was collected using 1 mm path length cuvette, bandwidth of 1 nm, data pitch of 0.5 nm, scanning speed of 50 nm/min, continuous scanning mode, and with 5 accumulations. Protein concentration was 0.25 mg/ml (10 µM) in 10 mM potassium phosphate pH 7.4, 250 mM potassium fluoride. Ligands were at 12 µM. Melt curves were measured at 222 nm between temperatures 20-95 ºC, bandwidth of 1 nm, and a ramp rate of 1 ºC/min with 10 s wait time. Calculation of T_m was done in Spectra Manager II by finding the peak of the first derivative of the melt curves, calculated using the Savitzky-Golay filter.

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS). The oligomeric state of σ_2 receptor was assessed by SEC–MALS using a Wyatt Dawn Heleos II multi-angle light scattering detector and Optilab TrEX refractive index monitor with an Agilent isocratic HPLC system Infinity II 1260. Receptor was prepared as described above, but with no ligand added during purification. The ligand-free receptor was diluted to 1 mg/ml in SEC–MALS buffer (0.01% LMNG, 20 mM HEPES pH 7.5, 150 mM sodium chloride). Ligands were added to a final concentration of 1 µM and the sample was incubated with ligand for 2 h at room temperature (21 °C). Separation steps were performed in SEC–MALS buffer with a Tosoh G4SWxl column at a flow rate of 0.5 ml min^{-1}. Data analysis used the Astra software package version 6.1.4.25 (Wyatt) using the protein conjugate method with a dn/dc value of 0.21 (mL/g) for detergent and 0.185 (mL/g) for protein.

Molecular docking. The σ_2 receptor bound to cholesterol (PDB ID: 7MFI) was used in the docking calculations. The structure was protonated at pH 7.0 by Epik and PROPKA in Maestro[59] (2019 release). Based on the mutagenesis data[14], E73 was modeled as a neutral residue. AMBER united atom charges were assigned to the structure. To model more realistic low protein dielectric boundary of this site, we embedded the receptor into a lipid-bilayer to capture its native environment in endoplasmic reticulum (ER) membrane, then followed by a 50 ns coarse-grained molecular dynamic (MD) simulation with a restricted receptor conformation. A more detailed protocol can be found on the DISI wiki page (http://wiki.docking.org/index.php/Membrane_Modeling). The volume of the low dielectric and the desolvation volume was extended out 2.2 Å and 1.2 Å, respectively, from the surface of protein and modelled lipid-bilayer using spheres calculated by SPHGEN. Energy grids were pre-generated with AMBER force fields using CHEMGRID for van der Waals potential[60], QNIFFT[61] for Poisson–Boltzmann-based electrostatic potentials, and SOLVMAP[62] for ligand desolvation.

The resulting docking setup was evaluated for its ability to enrich known σ_2 ligands over property-matched decoys. Decoys are unlikely to bind to the receptor because despite their similar physical properties to known ligands, they are topologically dissimilar. We extracted 10 known σ_2 ligands from ChEMBL(https://www.ebi.ac.uk/chembl/) including PB28 and roluperidone whose crystallographic poses were report here. Five-hundred and forty-two property-matched decoys were generated by the DUDE-Z pipeline[63]. Docking performance was evaluated based on the ability to enrich the knowns

over the decoys by docking rank, using log adjusted AUC values (logAUC). The docking setup described above was able to achieve a high logAUC of 39 and to recover the crystal poses of PB28 and roluperidone with RMSD values of 0.93 and 0.77 Å, respectively. This docking setup gave the best retrospective enrichment and pose reproduction among three ligand-bound σ_2 structures (**Supplementary Fig. 4.5**). We also constructed an 'extrema' set[63] of 61,687 molecules using the DUDE-Z web server (http://tldr.docking.org) to ensure that molecules with extreme physical properties were not enriched. The docking setup enriched close to 90% mono-cations among the top1000 ranking molecules. To check if the limited amounts of knowns and property-matched decoys over-trained the docking parameters, the enrichment test was run using 574 additional σ_2 ligands from S2RSLDB[42] (http://www.researchdsf.unict.it/S2RSLDB) against the 'extrema' set. The resulting high logAUC of 41 demonstrated the docking setup was still able to enrich knowns over decoys on a 112-fold larger test set, indicating the favorable docking parameters for launching an ultra-large-scale docking campaign.

Four-hundred and ninety million cations from ZINC15 (http://zinc15.docking.org), characterized by similar physical properties as $\sigma_{1/2}$ known ligands (for instance, with calculated octanol-water partition coefficients (cLogP) <=5 and with 250 Da <molecular weight <=400 Da), was then docked against the σ_2 ligand binding site using DOCK3.8. Of these, 469 million molecules were successfully docked. On average, 3,502 orientations were explored and for each orientation, 183 conformations were averagely sampled. In total, more than 314 trillion complexes were sampled and scored. The total calculation time was 177,087 hours, or 3.7 calendar days on a cluster of 2,000 cores.

The top-ranking 300,000 molecules were filtered for novelty using the ECFP4-based Tanimoto coefficient against 2,232 $\sigma_{1/2}$ ligands in ChEMBL (https://www.ebi.ac.uk/chembl/) and 574 σ_2 ligands from S2RSLDB (http://www.researchdsf.unict.it/S2RSLDB). Molecules with Tanimoto coefficient (T_c) ≥ 0.35 were eliminated. The remaining 196,170 molecules were clustered by ECFP4-based T_c of 0.5, resulting in 33,585 unique clusters. From the top 5,000 novel chemotypes, molecules with > 2 kcal/mol internal strains were filtered out using strain_rescore.py in Macromodel[64]. After filtering for novelty and diversity, the docked poses of the best-scoring members of each chemotype were manually inspected for favorable and diversified interactions with the σ_2 site, such as the salt bridge with Asp29, the hydrogen bond with His21/Val146 and the π-π stacking with Tyr50/Trp49. Ultimately, 86 compounds were chosen for testing, 79 of which were successfully synthesized.

Hit-rate curve prediction. To guide the design of scoring bins for the hit rate curve, 1,000 docked poses were sampled in bins every 2.5 kcal/mol from the best score of -65 kcal/mol up to -22.5 kcal/mol. We chose this 2.5 kcal/mol distance between the bins to span the range with enough points (bins) to define a potential hit-rate vs. docking score curve. At the top of what we expected to be the curve, we increased the bin sizes because the density of molecules at these very highest ranks was relatively low. Correspondingly, at the lowest scores we added several more bins, also at a larger spacing, to help us get a robust lower baseline. The estimated hit rate was calculated by the number of sensible docked poses divided by 1,000. The criteria to define a sensible docked pose contains 1) no unsatisfied hydrogen bond donors; 2) less than 3 unsatisfied

hydrogen acceptors; 3) forms a salt bridge with Asp29; 4) total torsion strain energy < 8 units; 5) maximum strain energy per torsion angle < 3 units. The first three filters were implemented based on LUNA (https://github.com/keiserlab/LUNA), which calculated all the intra- and interactions of a docked pose with the receptor, then hashed them into a binary fingerprint. The strain energy was calculated by an in-house population-based method[62]. Based on the shape of the estimated prior curve (**Supplementary Fig. 4.6**), more scoring bins are selected in the higher estimated hit-rate region: -65, -59.73 and -57.5 kcal/mol. After that, every scoring bin was 2.5 kcal/mol from each other till -37.5. The last four bins were 5 kcal/mol from each other. 13,000 molecules sampled were from these 14 scoring bins were filtered by novelty and internal torsion strain described above. The remaining 9,216 novel and non-strained molecules were cluster by the LUNA 1024-length binary fingerprint of a $T_c = 0.32$, resulting in 6,681 clusters. The first 40 chemotypes were attempted to be purchased from each scoring bin. After the evaluation of synthesis availability from the vendors, 491 molecules were ordered (**Supplementary Tables 4.1 and 4.3**).

Hit-rate curve fitting. To fit the Bayesian hit-rate models we used Stan[65] (v2.21.2) via BRMS[66] (v2.14.4), with generic parameters: iter=4000, and cores=4. Here are the model specific parameters. For both hit-picking prior and posterior Sigmoid models formula=bmrs::formula(hit ~ top * inv_logit(hill*4/top*(dock_energy - dock50)), top + hill + $dock_{50}$ ~ 1, nl=TRUE), where hill is scaled by 4/top so it is the slope of the curve at the $dock_{50}$ irrespective of the value of Top. For Prior Sigmoid model, prior=c(brms::prior(normal(.5, .2), lb=0, ub=1, nlpar="top"), brms::prior(normal(-50, 10),

nlpar="dock$_{50}$"),   brms::prior(normal(-.1,   .1),   ub=-.001,   nlpar="hill")),

inits=function(){list(top=as.array(.5),   dock$_{50}$=as.array(-50),   hill=as.array(-.1))},

family=gaussian(). Updating the Prior sigmoid model with the mean expected hit-rate for

each computationally analyzed tranche yielded an estimate and 95% credible interval for

the sigma parameter for the Gaussian response of 20 [15, 30]%, but did not significantly

adjust the distributions for Top, Hill, or Dock$_{50}$ (**Supplementary Fig. 4.7**). Therefore, to

estimate the posterior sigmoid model, we transferred the per-parameter prior distributions

and initial values and used the family=bernoulli("identity"). To compare models, we used

the loo package to add the Pareto smoothed importance sampling leave-one-out (PSIS-

LOO) and Bayesian version of the R2[67] (loo_R2) information criteria. Figures were

generated using tidybayes[68], ggplot2[69], and tidyverse[70] packages in R[71].

Analoging within the make-on-demand library. Using 4 primary docking hits

(ZINC450573233, ZINC533478938, ZINC548355486 and ZINC895657866) as queries in

SmalWorld (https://sw.docking.org/) from the 28B make-on-demand library, a subset of

Enamine REAL space, 20,005 analogues were selected by its default settings, then

docked into the σ_2 site for potential favorable interactions with His21, Tyr50, Gln77, and

Val146.

Make-on-demand synthesis. 79 molecules that were prioritized by human

inspection were delivered within 7 weeks with a 93% fulfilment rate, and 412 molecules

by docking score alone were delivered within 4 weeks with an 82% fulfilment rate after a

single synthesis attempt (**Supplementary Tables 4.1 and 4.3-4.4**). Most of the make-on-

demand molecules were derived from Enamine REAL database (https://enamine.net/compound-collections/real-compounds). See **Supplementary Information** (available in online publication) for synthesis procedure and characterization of compounds.

Yeast isomerase complementation assay. The human σ_2 receptor, ERG2, and EBP were subcloned into the URA3 shuttle vector p416GPD. The plasmids were transformed into the Erg2-deficient *Saccharomyces cerevisiae* strain Y17700 (BY4742; MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; YMR202w::kanMX4) (Euroscarf) by the lithium acetate/single-stranded carrier DNA/polyethylene glycol method. A single colony was picked from a URA-selective plate and grown in suspension. Yeast were diluted in sterile water in a five-fold serial dilution starting from O.D. 0.1. Two microliters of the yeast dilutions were spotted on a URA−selective plate either in the absence or the presence of sub-inhibitory concentrations of cycloheximide (50 ng/ml) and grown at 30°C for 24-48 h before imaging.

Sterol isomerization enzymatic assay. EBP and σ_2 were cloned into pcDNA3.1 (Invitrogen) mammalian expression vector with FLAG and protein C affinity tag, respectively. Proteins were purified as described for crystallography preparations, except no ligand was present during purification. Following size exclusion chromatography proteins were flash frozen in liquid nitrogen and kept at -80 °C until use. Zymostenol (CAS #566-97-2) and lathosterol (CAS #80-99-9) were purchased from Avanti Polar Lipids. For each sterol, a 2x solution was prepared by first dissolving DDM in isopropanol to 1% (w/v)

186

and dissolving sterols in chloroform to a concentration of 1 mg/ml, followed by transferring 500 µM of the sterols to a new vial, evaporating under argon, and dissolving with DDM in a 1:20 (w/w) detergent to sterol ratio and a final 0.2% detergent in HEPES buffered saline (HBS; 20 mM HEPES pH 7.5, 150 mM NaCl). Proteins were diluted in HBS to 5 µM. Individual sterol standards were prepared by mixing each sterol 1:1 with HBS. A mixed sterol standard was prepared by mixing both sterols in a 1:1 ratio. For the enzymatic reactions, sterols were mixed in 1:1 ratio with the protein sample to give a final protein concentration of 2.5 µM, sterol concentration of 250 µM, and detergent concentration of 0.1%, in HBS. Reactions were incubated for 1 hour at 37 °C and then diluted 1:10 in methanol and kept at -20 °C until analysis by LC-MS. Samples were analyzed on a QE-plus mass spectrometer coupled to an Ultimate 3000 LC (Thermo fisher) in a method modified from Skubic *et al*[72]. Five microliters were injected on a Force PFPP column coupled with an Allure PFPP column (both 2mm x 150 mm, Restek) maintained at 40°C. The mobile phases were A: methanol:isopropyl alcohol:water:formic acid (80:10:10:0.02) 5 mM ammonium formate, and B: isopropyl alcohol. The gradient was as follows: 0% B for 15 min, then 100% B in 1 second, maintained at 100% B for 5 min, followed by 5 min re-equilibration at 0% B. The flow rate was 0.15 mL min^{-1}. The mass spectrometer was acquiring in t-SIM mode for the [M-H2O+H]+ ion (369.35158) with 70,000 resolution, and 0.5 m/z isolation. Standard samples for each compound were run first separately to obtain the retention time of each of the two isobaric compounds.

µOR activation assay. To measure µOR $G_{i/o}$-mediated cAMP inhibition, 2.5 million HEK-293T cells (ATCC) were seeded in 10-cm plates. Eighteen to 24 hours later,

upon reaching 85-90% confluency, cells were transfected using a 1:3 ratio of human μOR and a split-luciferase based cAMP biosensor (pGloSensorTM-22F; Promega). TransIT 2020 (Mirus Biosciences) was used to complex the DNA at a ratio of 3 μL TransIT per μg DNA, in OptiMEM (Gibco-ThermoFisher) at a concentration of 10 ng DNA per μL OptiMEM. Twenty-four hours later, cells were harvested from the plate using Versene (PBS + 0.5 mM EDTA, pH 7.4) and plated in poly-D-lysine-coated white, clear-bottom 96-well assay plates (Corning Costar #3917) at a density of 35,000 cells per well and incubated at 37 °C with 5% CO_2 overnight. The next day, after aspiration of the culture medium, cells were incubated for 2 hours covered, at room temperature, with 40 μL assay buffer (CO_2-independent medium, 10% FBS) supplemented with 2% (v/v) GloSensor™ reagent (Promega). To stimulate endogenous cAMP via β adrenergic-G_s activation, 5x drugs were prepared in 10x isoproterenol containing assay buffer (200 nM final concentration). For naloxone competition experiments, 5x naloxone (1 μM final concentration) was also added to each well. Luminescence was immediately quantified using a BMG Clariostar microplate reader. Data were analyzed using nonlinear regression in GraphPad Prism 9.0 (Graphpad Software Inc., San Diego, CA).

Off-target counterscreens. Screening of compounds in the PRESTO-Tango GPCRome was accomplished as previously described[41] with several modifications. First, HTLA cells were plated in DMEM with 10% FBS and 10 U ml−1 penicillin–streptomycin. Next, the cells were transfected using an in-plate PEI method[73]. PRESTO-Tango receptor DNAs were resuspended in OptiMEM and hybridized with PEI before dilution and distribution into 384-well plates and subsequent addition to cells. After overnight

incubation, drugs were added to cells at 10 μM final concentration without replacement of the medium. The remaining steps of the PRESTO-Tango protocol were followed as previously described. For those six receptors for which activity was reduced to less than 0.5-fold of basal levels of relative luminescence units or for the one receptor for which basal signaling was increased greater than 3-fold of basal levels, assays were repeated as a full dose–response assay. Activity for none of the seven could be confirmed, and we discount the apparent activity seen in the single-point assay.

Radioligand binding screen of off-targets was performed by the National Institutes of Mental Health Psychoactive Drug Screen Program (PDSP)[74]. Detailed experimental protocols are available on the NIMH PDSP website at https://pdsp.unc.edu/pdspweb/content/PDSP%20Protocols%20II%202013-03-28.pdf.

Cell lines

All cell lines in this study were not authenticated. All cells used in this study are commercial and were obtained from vendors as indicated. Cells were confirmed to be mycoplasma free.

Animals

Animal experiments were approved by the UCSF Institutional Animal Care and Use Committee and were conducted in accordance with the NIH Guide for the Care and Use of Laboratory animals. Adult (8-10 weeks old) male C56BL/6 mice (strain #664) were purchased from the Jackson Laboratory. Mice were housed in cages on a standard 12:12 hour light/dark cycle with food and water *ad libitum*. We did not perform sample-size

calculations. We modeled our sample sizes for behavioral studies on previous studies using a similar approach to our own, which have been demonstrated to be capable of detecting significant changes[75,76]. The animals were randomly assigned to the treatment group and control group. For behavioral experiments, animals were initially placed into one cage and allowed to free run for a few minutes. Next, each animal was randomly picked up, injected with the drug or vehicle control, and placed into a separate cylinder before the behavior test. All experiments were for animal behavior and followed this randomization protocol. For all behavioral testing the experimenter was always blind to treatment. All experiments were in animals and under blinding conditions.

Compounds

All ligands used in the animal studies were synthesized by Enamine (https://enamine.net/) (**Supplementary Table 4.5**) and dissolved 30 minutes prior testing. PB28 and Z1665845742 were resuspended in 0.9% NaCl. Z4857158944 and Z4446724338 were resuspended in 20% cyclodextrin. PD-144418 was resuspended in 20% Kolliphor.

Behavioral analyses

For all behavioral tests, animals were first habituated for 1 hour in Plexiglas cylinders. The experimenter was always blind to treatment. All tests were conducted 30 minutes after subcutaneous injection of the compounds. Hindpaw mechanical thresholds were determined with von Frey filaments using the up-down method[77]. For the ambulatory (rotarod) test, mice were first trained on an accelerating rotating rod, 3 times for 5 min,

before testing with any compound.

Spared-nerve injury (SNI) model of neuropathic pain

Under isoflurane anesthesia, two of the three branches of the sciatic nerve were ligated and transected distally, leaving the sural nerve intact. Behavior was tested 7 to 14 days after injury and *in situ* hybridization was performed one week post-injury.

***In situ* hybridization**

In situ hybridization was performed using fresh DRG tissue from adult mice (8-10 week old), following Advanced Cell Diagnostics' protocol and as previously described[78]. All images were taken on an LSM 700 confocal microscope (Zeiss) and acquired with ZEN 2010 (Zeiss). Adjustment of brightness/contrast and changing of artificial colors (LUT) were done with Photoshop. The same imaging parameters and adjustments were used for all images within an experiment.

Statistical analyses of animal studies

All animal statistical analyses were performed with GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA) unless otherwise noted. All data are reported as means ± SEM unless otherwise noted. Dose-response experiments were analyzed with one-way ANOVA and time-course experiments were analyzed with two-way ANOVA, and both experiments used Dunnett's multiple comparison post-hoc test to determine differences between specific treatments and vehicle controls visualized in the figures. Rotarod experiments were analyzed using one-way ANOVA (saline, Z1665845742, and

Z4857158944) or unpaired two-tailed Student's t-test (kolliphor and Z4446724338). Details of analyses, including number of tested animals and groups, degrees of freedom, and *p*-values can be found in the figure legends.

Code Availability

DOCK3.7 is freely available for non-commercial research http:// dock.compbio.ucsf.edu/DOCK3.7/. A web-based version is available at http:// blaster.docking.org/.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The coordinates and structure factors for PB28-bound σ_2, roluperidone-bound σ_2, Z1241145220-bound σ_2, Z4857158944-bound σ_2, and cholesterol-bound σ_2 have been deposited in the PDB with accession codes 7M93, 7M94, 7M95, 7M96, and 7MFI respectively. The identities of the compounds docked in this study are freely available from the ZINC database (http://zinc15.docking.org) and active compounds may be purchased from Enamine. Any other data relating to this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

Ethical compliance

All animal experiments were approved by the Institutional Animal Care and Use Committee at UCSF and were conducted in accordance with the NIH Guide for the Care and Use of Laboratory animals.

4.9 References

1. Waarde, A. van *et al.* Potential applications for sigma receptor ligands in cancer diagnosis and therapy. *Biochimica Et Biophysica Acta Bba - Biomembr.* **1848**, 2703–2714 (2015).

2. Harvey, P. D. *et al.* Effects of Roluperidone (MIN-101) on two dimensions of the negative symptoms factor score: Reduced emotional experience and reduced emotional expression. *Schizophr. Res.* **215**, 352–356 (2020).

3. Sahn, J. J., Mejia, G. L., Ray, P. R., Martin, S. F. & Price, T. J. Sigma 2 Receptor/Tmem97 Agonists Produce Long Lasting Antineuropathic Pain Effects in Mice. *Acs Chem. Neurosci.* **8**, 1801–1811 (2017).

4. Intagliata, S. *et al.* Discovery of a Highly Selective Sigma-2 Receptor Ligand, 1-(4-(6,7-Dimethoxy-3,4-dihydroisoquinolin-2(1H)-yl)butyl)-3-methyl-1H-benzo[d]imidazol-2(3H)-one (CM398), with Drug-Like Properties and Antinociceptive Effects In Vivo. *AAPS J.* **22**, 94 (2020).

5. Quadir, S. G. *et al.* The Sigma-2 receptor / transmembrane protein 97 (σ2R/TMEM97) modulator JVW-1034 reduces heavy alcohol drinking and associated pain states in male mice. *Neuropharmacology* **184**, 108409 (2021).

6. Grundman, M. *et al.* A phase 1 clinical trial of the sigma-2 receptor complex allosteric antagonist CT1812, a novel therapeutic candidate for Alzheimer's disease. *Alzheimer's Dementia Transl. Res. Clin. Interventions* **5**, 20–26 (2019).

7. Riad, A. *et al.* Sigma-2 Receptor/TMEM97 and PGRMC-1 Increase the Rate of Internalization of LDL by LDL Receptor through the Formation of a Ternary Complex. *Sci. Rep.* **8**, 16845 (2018).

8. Abate, C. *et al.* PB28, the Sigma-1 and Sigma-2 Receptors Modulator With Potent Anti–SARS-CoV-2 Activity: A Review About Its Pharmacological Properties and Structure Affinity Relationships. *Front Pharmacol.* **11**, 589810 (2020).

9. Shields, S. D., Eckert, W. A. & Basbaum, A. I. Spared nerve injury model of neuropathic pain in the mouse: a behavioral and anatomic analysis. *J. Pain* **4**, 465–470 (2003).

10. Hellewell, S. B. *et al.* Rat liver and kidney contain high densities of σ1 and σ2 receptors: characterization by ligand binding and photoaffinity labeling. *European J. Pharmacol. Mol. Pharmacol.* **268**, 9–18 (1994).

11. Hellewell, S. B. & Bowen, W. D. A sigma-like binding site in rat pheochromocytoma (PC12) cells: decreased affinity for (+)-benzomorphans and lower molecular weight suggest a different sigma receptor form from that of guinea pig brain. *Brain Res.* **527**, 244–253 (1990).

12. Hanner, M. *et al.* Purification, molecular cloning, and expression of the mammalian sigma1-binding site. *Proc. National Acad. Sci.* **93**, 8072–8077 (1996).

13. Langa, F. *et al.* Generation and phenotypic analysis of sigma receptor type I (σ1) knockout mice. *Eur. J. Neurosci.* **18**, 2188–2196 (2003).

14. Alon, A. *et al.* Identification of the gene that codes for the σ2 receptor. *Proc. National Acad. Sci.* **114**, 7160–7165 (2017).

15. Ebrahimi-Fakhari, D. *et al.* Reduction of TMEM97 increases NPC1 protein levels and restores cholesterol trafficking in Niemann-pick type C1 disease cells. *Hum. Mol. Genet.* **25**, 3588–3599 (2016).

16. Bartz, F. *et al.* Identification of Cholesterol-Regulating Genes by Targeted RNAi Screening. *Cell Metab.* **10**, 63–75 (2009).

17. Sanchez-Pulido, L. & Ponting, C. P. TM6SF2 and MAC30, new enzyme homologs in sterol metabolism and common metabolic disease. *Frontiers Genetics* **5**, 439 (2014).

18. Mahdessian, H. *et al.* TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc. National Acad. Sci.* **111**, 8913–8918 (2014).

19. Vilner, B. J., John, C. S. & Bowen, W. D. Sigma-1 and sigma-2 receptors are expressed in a wide variety of human and rodent tumor cell lines. *Cancer Res.* **55**, 408–13 (1995).

20. Scott, L. L. *et al.* Small molecule modulators of σ2R/Tmem97 reduce alcohol withdrawal-induced behaviors. *Neuropsychopharmacol.* **43**, 1867–1875 (2018).

21. Vázquez-Rosa, E. *et al.* Neuroprotective Efficacy of a Sigma 2 Receptor/TMEM97 Modulator (DKR-1677) after Traumatic Brain Injury. *Acs Chem. Neurosci.* **10**, 1595–1602 (2019).

22. Stein, R. M. *et al.* Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 1–8 (2020).

23. Schuller, M. *et al.* Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).

24. Schmidt, H. R. & Kruse, A. C. The Molecular Function of σ Receptors: Past, Present, and Future. *Trends Pharmacol. Sci.* **40**, 636–654 (2019).

25. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009).

26. Long, T. *et al.* Structural basis for human sterol isomerase in cholesterol biosynthesis and multidrug recognition. *Nat. Commun.* **10**, 2452 (2019).

27. Audet, M. & Stevens, R. C. Emerging structural biology of lipid G protein-coupled receptors. *Protein Sci.* **28**, 292–304 (2019).

28. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

29. Schmidt, H. R. *et al.* Crystal structure of the human σ1 receptor. *Nature* **532**, 527–530 (2016).

30. Hubler, Z. *et al.* Accumulation of 8,9-unsaturated sterols drives oligodendrocyte formation and remyelination. *Nature* **560**, 372–376 (2018).

31. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).

32. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

33. Fischer, A., Smieško, M., Sellner, M. & Lill, M. A. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *J. Med. Chem.* **64**, 2489–2500 (2021).

34. Cendán, C. M., Pujalte, J. M., Portillo-Salido, E., Montoliu, L. & Baeyens, J. M. Formalin-induced pain is reduced in σ1 receptor knockout mice. *Eur. J. Pharmacol.* **511**, 73–74 (2005).

35. Puente, B. de la *et al.* Sigma-1 receptors regulate activity-induced spinal sensitization and neuropathic pain after peripheral nerve injury. *Pain* **145**, 294–303 (2009).

36. Cendán, C. M., Pujalte, J. M., Portillo-Salido, E. & Baeyens, J. M. Antinociceptive effects of haloperidol and its metabolites in the formalin test in mice. *Psychopharmacology* **182**, 485–493 (2005).

37. Romero, L. *et al.* Pharmacological properties of S1RA, a new sigma-1 receptor antagonist that inhibits neuropathic pain and activity-induced spinal sensitization. *Brit. J. Pharmacol.* **166**, 2289–2306 (2012).

38. Bruna, J. *et al.* Efficacy of a Novel Sigma-1 Receptor Antagonist for Oxaliplatin-Induced Neuropathy: A Randomized, Double-Blind, Placebo-Controlled Phase IIa Clinical Trial. *Neurotherapeutics* **15**, 178–189 (2018).

39. Vela, J. M., Merlos, M. & Almansa, C. Investigational sigma-1 receptor antagonists for the treatment of pain. *Expert Opin. Inv. Drug* **24**, 883–896 (2015).

40. Kooistra, A. J. *et al.* GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, gkaa1080- (2020).

41. Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41-54.e19 (2018).

42. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biology* **22**, nsmb.3014 (2015).

43. Nastasi, G. *et al.* S2RSLDB: a comprehensive manually curated, internet-accessible database of the sigma-2 receptor selective ligands. *J. Cheminformatics* **9**, 3 (2017).

44. Huang, X.-P. *et al.* Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **527**, 477–483 (2015).

45. Wang, S. *et al.* Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* **555**, 269–273 (2018).

46. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

47. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

48. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *J. Mol. Biol.* **302**, 205–217 (2000).

49. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D* **73**, 148–157 (2017).

50. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

51. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biological Crystallogr.* **66**, 125–132 (2010).

52. Alford, R. F. *et al.* An Integrated Framework Advancing Membrane Protein Modeling and Design. *Plos Comput. Biol.* **11**, e1004398 (2015).

53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

54. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).

55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biological Crystallogr.* **66**, 486–501 (2010).

56. Schrödinger & DeLano, W. *PyMOL.* (2020).

57. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

58. Chu, U. B. & Ruoho, A. E. Sigma Receptor Binding Assays. *Curr. Protoc. Pharmacol.* **71**, 1.34.1-1.34.21 (2015).

59. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).

60. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **13**, 505–524 (1992).

61. Gallagher, K. & Sharp, K. Electrostatic Contributions to Heat Capacity Changes of DNA-Ligand Binding. *Biophys. J.* **75**, 769–776 (1998).

62. Mysinger, M. M. & Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model* **50**, 1561–1573 (2010).

63. Stein, R. M. *et al.* Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model* **61**, 699–714 (2021).

64. Gu, S., Smith, M. S., Yang, Y., Irwin, J. J. & Shoichet, B. K. Ligand Strain Energy in Large Library Docking. *J. Chem. Inf. Model* **61**, 4331–4341 (2021).

65. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *Journal of Statistical Software* **76**, (2017).

66. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Soft.* **80**, 1–28 (2017).

67. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. R-squared for Bayesian Regression Models. *Am. Statistician* **73**, 307–309 (2019).

68. Kay, M. *tidybayes: Tidy Data and Geoms for Bayesian Models*. (2020).

69. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

70. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

71. Team, R. C. *R: A Language and Environment for Statistical Computing*. (2018).

72. Skubic, C., Vovk, I., Rozman, D. & Križman, M. Simplified LC-MS Method for Analysis of Sterols in Biological Samples. *Molecules* **25**, 4116 (2020).

73. Longo, P. A., Kavran, J. M., Kim, M.-S. & Leahy, D. J. Chapter Eighteen Transient Mammalian Cell Transfection with Polyethylenimine (PEI). *Methods Enzymol.* **529**, 227–240 (2013).

74. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215 (2012).

75. Scherrer, G. *et al.* Dissociation of the Opioid Receptor Mechanisms that Control Mechanical and Heat Pain. *Cell* **137**, 1148–1159 (2009).

76. Muralidharan, A. *et al.* Identification and characterization of novel candidate compounds targeting 6- and 7-transmembrane μ-opioid receptor isoforms. *Brit. J. Pharmacol.* **178**, 2709–2726 (2021).

77. Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M. & Yaksh, T. L. Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Meth.* **53**, 55–63 (1994).

78. Solorzano, C. *et al.* Primary Afferent and Spinal Cord Expression of Gastrin-Releasing Peptide: Message, Protein, and Antibody Concerns. *J. Neurosci.* **35**, 648–657 (2015).

Gloss to Chapter 5

The cannabinoid docking project began as my rotation project in the lab and by far was the project I worked on for the longest amount of time and faced the largest number of challenges. You'd probably call me crazy for agreeing to join a docking project that had previously been pursued in the lab and failed. A sentiment Brian held at the time was that *we always struggle with lipid receptors*, and I am glad this project and the lipid-binding GPCR docking projects that have occurred since have proven him wrong. I truly believe in the importance of lipid-binding receptors as important drug targets and that large-scale docking can find us better drug-like ligands, even if the hit rates aren't quite as spectacular as they are for other targets such as monoaminergic GPCRs.

At the time of writing this dissertation, this project had been through a gauntlet of trials and tribulations. The library wasn't particularly large in the area of chemical space we needed, the analogs were expensive and often times not in the make-on-demand database, the assays didn't work in our first collaborators' hands and had to be tested elsewhere, the scintillation counter broke and was backordered for 6 months, the compounds stuck to the vials or were tricky to dissolve and formulate, the first cryo-EM structure was missing density for the ligand, and the first draft of the paper was rejected, to name a few of the challenges. Recently, we discovered an issue with the chemistry of the lead molecule, being off by 1-methyl during some synthetic batches and not others. The work presented here is the current updated version that addresses these issues and mistakes to the best of our ability at time of submission.

Chapter 5:Structure-based discovery of cannabinoid-1 receptor

agonists with reduced side effects

Structure-based discovery of cannabinoid-1 receptor agonists with reduced side effects

Tia A. Tummino[1,2,†], Christos Iliopoulos-Tsoutsouvas[3,†], Joao M. Braz[4,†], Evan S. O'Brien[5], Reed M. Stein[1,2], Veronica Craik[4], Ngan K. Tran[3], Suthakar Ganapathy[3], Yuki Shiimura[5,6], Fei Tong[3], Thanh C. Ho[3], Dmytro S. Radchenko[7], Yurii S. Moroz[8,9], Fangyu Liu[1], Sian Rodriguez Rosado[4], Karnika Bhardwaj[4], Jorge Benitez[4], Yongfeng Liu[10], Herthana Kandasamy[11], Claire Normand[11], Meriem Semache[11], Laurent Sabbagh[11], Isabella Glenn[1], John J. Irwin[1], Kaavya Krishna Kumar[5,*], Alexandros Makriyannis[3,12*], Allan I. Basbaum[4,*], & Brian K. Shoichet[1,*]

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158, USA

[2]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, University of California, San Francisco, San Francisco, CA 94158, USA

[3]Center for Drug Discovery and Department of Pharmaceutical Sciences, Northeastern University, Boston, MA 02115, USA

[4]Department of Anatomy, University of California, San Francisco, San Francisco, CA 94158, USA

[5]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA

[6]Division of Molecular Genetics, Institute of Life Science, Kurume University, Fukuoka, Japan

[7]Enamine Ltd., 67 Chervonotkatska Street, Kyiv, 02094, Ukraine

[8]National Taras Shevchenko University of Kyiv, 60 Volodymyrska Street, Kyiv 01601, Ukraine

[9]Chemspace LLC, 85 Chervonotkatska Street, Suite 1, Kyiv, 02094, Ukraine

[10]National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), School of Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

[11]Domain Therapeutics North America Inc., Montréal, Québec, H4S 1Z9, Canada, Montréal, QC, H3T 1J4, Canada

[12]Department of Chemical and Chemical Biology, Northeastern University, Boston, MA 02115, USA

†Contributed equally.

*Correspondence: kaavyak@stanford.edu, a.makriyannis@northeastern.edu, allan.basbaum@ucsf.edu, bshoichet@gmail.com

5.1 Abstract

Docking tangible virtual libraries can reveal unexpected chemotypes that complement the structures of biological targets. Seeking new agonists for the cannabinoid-1 receptor (CB1R), we docked 74 million tangible molecules, prioritizing 46 high ranking ones for *de novo* synthesis and testing. Nine were active by radioligand competition, with > 50% radioligand displacement, a 20% hit-rate. Structure-based optimization of one of the most potent of these (K_i = 731 nM) led to **'4042**, a 1.9 nM binder and a full CB1 agonist. A cryo-EM structure of the **'4042**-CB1-G_{i1} complex confirmed its docked pose, providing a template for further optimization. The new agonist was strongly analgesic especially against thermal pain, with a 10-fold therapeutic window over sedation and catalepsy and no observable conditioned place preference or aversion. These findings suggest that new cannabinoid chemotypes may be able to disentangle the characteristic "tetrad" side-effects from its desired analgesic effect, supporting the further development of cannabinoids as pain therapeutics.

5.2 Introduction

Although the therapeutic use of cannabinoids dates back to at least the 15th century[1,2], their use in modern therapy, for instance as analgesics, has been slowed by their sedative and mood-altering effects, and by concerns over their reinforcing and addictive properties[3,4]. With changes in cannabis' legal status, an ongoing epidemic of chronic pain, as well as an effort to reduce reliance on opioids for pain management, has come a renewed interest in understanding both the endocannabinoid system and how to leverage it for therapeutic development[5]. Areas of potential application include anxiety[6], nausea[7], obesity[8], seizures[9], and pain[10], the latter of which is the focus of this study. Progress in these areas has been slowed by the physical properties of the cannabinoids themselves, which are often highly hydrophobic, by the challenges of the uncertain legal environment, and by the substantial adverse side effects often attending on cannabinoids, including sedation, psychotropic effects, and concerns about reinforcement and addiction[3]. Indeed, a characteristic defining feature of cannabinoids is their "tetrad" of effects[11]: analgesia, hypothermia, catalepsy, and hypolocomotion, the latter three of which may be considered adverse. Additionally, inconclusive results in human clinical trials[12] have led to uncertainty in the field as to the effectiveness of cannabinoids as therapeutics. Nevertheless, the strong interest in new analgesics, and the clear efficacy of cannabinoids in animal models of nociception[13], have maintained therapeutic interest in these targets.

The cannabinoid-1 and -2 receptors (CB1R and CB2R), members of the lipid family of G-protein coupled receptors (GPCRs), are the primary mediators of cannabinoid

activity[14]. The structural determination of these receptors[15–21] affords the opportunity to use structure-based methods to find ligands with new chemotypes. Recent structure-based docking of make-on-demand virtual libraries have discovered new chemotypes for a range of targets, often with new pharmacology and reduced side effects[22–28]. Thus, new CB1R chemotypes might address some of the unfavorable properties of current cannabinoids, such as their physicochemical properties or side-effect profiles. To identify such new chemotypes, we computationally docked a library of 74 million virtual but readily accessible ("tangible") molecules against CB1R, revealing a range of new scaffolds with favorable physical properties. Structure-based optimization led to agonists binding with low-nanomolar binding affinities. The lead agonist is a potent analgesic, with pain-relieving activity at doses as low as 0.1 mg/kg. It has a ten-fold separation between analgesia and both sedation and catalepsy, addressing two of the four aspects of the "tetrad" and highlighting the utility of large-scale virtual screening for identifying unique biology through new chemistry.

5.3 Results

Large-library docking against CB1R. The CB1R orthosteric site is large and lipophilic, explaining the high molecular weight and hydrophobicity of many of its ligands (**Fig. 5.S1**), which are metabolic and solubility liabilities[29]. We therefore sought molecules in a more "lead-like" physical property range. In preliminary studies, strict enforcement of such properties (i.e., MW $\leq$ 350 amu, cLogP $\leq$ 3.5) revealed no new ligands from docking. Accordingly, we created a special 74-million molecule subset of the ZINC15 database[30] composed of molecules 350 to $\geq$ 500 amu and calculated LogP (cLogP) 3 to $\leq$ 5,

reasoning that these would be more likely to complement the CB1R site, while still being more polar and smaller than typical of cannabinoid ligands (**Fig. 5.1B**). Each molecule was docked in an average of 3.04 million poses (orientations x conformations), totaling roughly 63 trillion sampled and scored complexes. Seeking a diverse set of molecules to test, the top-ranking 300,000 were clustered into 60,420 sets, and the highest scoring member of each cluster was filtered for topological dissimilarity to known CB1/CB2 receptor ligands in ChEMBL[31,32] using Tanimoto coefficient (Tc < 0.38) comparisons of ECFP4-based molecular fingerprints. High-ranking library compounds that did not resemble known ligands were filtered for potential polar interactions with S383[7.39] and H178[2.65] (superscripts denote Ballesteros-Weinstein nomenclature[33]; see **Methods, Fig. 5.1A, Table 5.S1**). The top-ranking 10,000 remaining molecules were visually evaluated in UCSF Chimera[34], and 60 were prioritized for *de novo* synthesis. Of these, 46 were successfully made and tested for CB1R activity. Consistent with the design of the library, the new molecules were smaller and more polar than most existing cannabinoid ligands, skirting the edge of property-space that is suitable for the large and hydrophobic CB1 orthosteric pocket (**Fig. 5.1B**).

In single-point radioligand displacement experiments, nine of the 46 prioritized molecules displaced over 50% of the radioligand, a 20% hit-rate (**Fig. 5.1C-D**, **Table 5.S1**). The top four of these (ZINC537551486, ZINC1341460450, ZINC749087800, and ZINC518437019, referred to as **'51486**, **'0450**, **'7800**, and **'7019**, respectively, from here on) were then tested in full concentration-response. All four displaced the radioligand ^{3}H-CP-55,940, with K_i values ranging from ~700 nM to 4 μM (**Fig. 5.1E**). Owing to coupling

to the inhibitory $G_{\alpha i}$ G-protein, functional efficacy experiments monitoring a decrease in forskolin (FSK) simulated cAMP were tested using hCB1-expressing cells, with **'51486** and **'0450** showing modest agonist activity. Limited solubility prohibited testing at high enough concentrations to obtain accurate EC_{50} measurements; fortunately, colloidal aggregation counter-screens showed no such activity below 10 μM (**Fig. 5.S2**), suggesting that activity seen in binding and functional assays is not due to this confounding phenomenon. Taken together, the nine actives explore a range of chemotypes topologically unrelated (i.e., dissimilar by Tanimoto coefficient) to known CB1 ligands (**Table 5.S1**), with relatively favorable physical properties (i.e., smaller mass with increased hydrophilicity; **Fig 5.1B,D**).

Although the new ligands are chemically and physically distinct from established cannabinoids, their docked poses recapitulate the interactions of the known ligands but do so with different scaffold and recognition elements. All of the four most potent ligands docked to adopt the "C" shaped conformation characteristic of the experimentally observed geometries of MDMB-Fubinaca[18], AM11542, and AM841[16] bound to CB1R. Similarly, all four are predicted to hydrogen-bond with S383[7.39], a potency-determinant interaction at CB1 receptors observed in all agonist-bound ligand-receptor complexes[35]. Additionally, all four ligands are predicted to make secondary hydrogen bonds to H178[2.65], a feature seen in only the most potent CB1 ligands, such as MDMB-Fubinaca. Largely, these electrostatic interactions are made using unique hydrogen-bond acceptor groups, such as an oxazole, oxathiine, or pyridazinone. Other characteristic hydrophobic and aromatic stacking interactions are found throughout the ligands, including with F268[ECL2],

W279[5.43], and F174[2.61], though again often using different aromatic groups than found in the known ligands (**Fig. 5.1F**). Similarly, all four ligands exhibit aromatic stacking and hydrophobic packing with the twin-toggle switch residues W356[6.48] and F200[3.36] which are important for receptor activation[36,37].

We sought to optimize these initial ligands. Molecules with ECFP4 Tcs $\geq$ 0.5 to the four actives were sought among a library of 12 billion tangible molecules using SmallWorld (NextMove Software, Cambridge UK), a program well-suited to ultra-large libraries. These analogs were built, docked, filtered, and selected using the same criteria as in the original docking campaign. Between 11 and 30 analogs were synthesized for each of the four scaffolds. Optimized analogs were found for three of the four initial hits, improving affinity by between 5 and 24-fold, with **'51486** improving 16-fold to a K_i of 44 nM, **'7019** improving 5-fold to 87 nM, and **'0450** improving 24-fold to 163 nM (**Table 5.S2**). In subsequent bespoke synthesis, the 44 nM analog of **'51486**, **'60154**, was further optimized to compound Z8504214042 (from here on referred to as **'4042**) with a K_i of 1.9 nM (**Fig. 5.S3**). **Figure 5.2** summarizes the structure-activity relationship (SAR) of the **'51486/'4042** series**.

Key learnings from the SAR include the importance of a hydrophobic group in the R_1 position of **'4042**, which is modeled to pack against W279[5.43] and T197[3.33] and methylation of the chiral center (R_4 position), which is predicted to increase Van der Waals interactions between the ligand and transmembrane helix 2. Finally, the terminal ester is oriented to hydrogen bond with H178[2.65] of the receptor, though the distance suggests

either a water-mediated interaction, or simply a weak hydrogen bond. As expected, the carboxylate analog of the ester which carries a formal negative charge, **'4051**, was a weak binder (K_i = 5 μM, 5,000-fold less potent)—this molecule, a very close analog to **'4042**, may provide the inactive member of a "probe pair" for future research. The lead that emerged, **'4042** at 1.9 nM, is about 2-fold more potent than the widely used CB1R probe CP-55,940 (**Fig. 5.4B**, below) and equipotent to the marketed drug nabilone (**Fig. 5.S3A, Table 5.S2**). Although its cLogP is higher than the docking hit '**51486**, its lipophilic ligand efficiency improved from 3.1 to 4.6 (**Fig. 5.2B**).

Cryo-EM structure of the '1350-CB1R-G$_{i1}$ complex. To understand the SAR of the **'4042** series at atomic resolution, and to template future optimization, we determined the structure of the agonist in complex with the activated state of the receptor. Initial efforts at single particle cryo-electron microscopy (cryo-EM) of **'4042** in complex with CB1R and the G$_{i1}$ heterotrimeric G-protein led to a structure where the ligand density seemed to reflect either multiple conformations of a single ligand, or multiple ligands. As **'4042** is a racemate, we purified it into it its component isomers, **'1350** and **'8690** using chiral chromatography (**Fig. 5.S4**) and measured CB1R binding by radioligand competition, as above. With K_i values of 0.95 nM and 90 nM, respectively, **'1350** was substantially more potent than its enantiomer, and subsequent functional studies revealed it to be the much stronger agonist (**Fig. 5.4A-B, Fig. 5.S4**; below). Accordingly, we re-determined the cryo-EM structure of the **'1350-CB1R-G$_{i1}$** complex (**Fig. 5.3**, **Fig. 5.S5,** see **Methods**) to a nominal resolution of 3.3 Å (**Table 5.S3**). Consistent with earlier structures of CB1R in its

activated state, the ligand occupies the orthosteric pocket formed by transmembrane helices (TMs) 2-3 and 5-7 and is capped by extracellular loop (ECL) 2.

The experimental structure of '**1350** superposes well on the docking-predicted pose of **'4042** in its *R*-enantiomer, which was the enantiomer with the better docking score to the receptor (-43 DOCK score versus -38 DOCK score for the *S*-enantiomer). The predicted and experimental structures superposed with an all-atom RMSD of 1.37 Å (**Fig. 5.3B**). The major interactions with CB1R predicted by the docking are preserved in the experimental structure, including the key hydrogen-bond between the amide carbonyl of the ligand and S383$^{7.39}$, though the distance between the donor and acceptors suggest there might be a water-mediated interaction that is not seen given the resolution of the current structure. The trifluoromethyl group is complemented by van der Waals and quadrupole interactions with residues W279$^{5.43}$ and T197$^{3.33}$, as anticipated by the docked structure, and consistent with the improvement in affinity by -1.7 kcal/mol (17-fold in K_i) on its replacement of the original fluorine.

Agonism and subtype selectivity of '4042. Given the potent affinity of **'4042** and of **'1350** (**Fig. 5.4A**), we next investigated their functional activity, and how they compared to that of the widely studied cannabinoid, CP-55,940[2]. We first measured $G_{i/o}$ mediated agonism via inhibition of forskolin-stimulated cAMP in the Lance Ultra cAMP assay (see **Methods**). Both **'4042**, **'1350**, and several of its analogs are agonists in human CB1R-expressing cells (hCB1R), with EC$_{50}$ values commensurate with their affinities (**Table 5.S2, 5.S4 Fig. 5.S3, 5.S6-7**) and with efficacies close to full agonism (E$_{max}$ typically >

75%). **'4042** and **'1350** had hCB1R EC_{50} (E_{max}) values of 3.3 nM (78%) and 1.6 nM (77%)

(**Fig. 5.4B**). The activity of racemic **'4042** was confirmed in several orthogonal cAMP and

ß-Arrestin assays (see **Methods**), including in the Cerep cAMP assay (**Fig. 5.S3C**), the

Glosensor assay (**Fig. 5.S3D**), the Tango ß-Arrestin translocation assay (**Fig. 5.S3E**) and

the DiscoverX ß-Arrestin-2 recruitment assay (**Fig. 5.S3F**). In summary, **'4042** and its *R*-

isomer, **'1350,** are potent agonists of hCB1R with low nM EC_{50} values.

Fortified by this potent activity, and to control for system bias[38–40] and questions of

signal amplification in the cAMP assays, we investigated both **'4042** and the more active

of its stereoisomers, **'1350**, for differential recruitment of several G-proteins and βArrestin-

2 against both CB1R and CB2R in the ebBRET bioSens-All® platform, comparing its

activity to CP-55,940 (**Fig. 5.4C-F, Fig. 5.S6, Table 5.S5-5.S6**). A good way to picture

the differential effects of **'1350** and **'4042** relative to CP-55,940 at CB1R and CB2R is via

"radar" plots (**Fig. 5.4C** and **5.4E**) depicting the relative effectiveness[38] toward each

signaling pathway ($10^{\Delta log(Emax/EC50)}$, see **Methods**). In CB1R, **'1350** was approximately 2

times more relatively efficacious at recruiting $G_{i/o}$ and G_{13} subtypes than CP-55,940,

though the pattern of effectors recruited was similar. Similar coupling profiles were seen

for **'4042**, though the effects were smaller, consistent with the latter compound being an

enantiomeric mixture. Whereas the CB1R radar plots were similar in pattern for **'1350**,

'4042 and CP-55,940, the differential activities for the highly related CB2R differed

qualitatively (**Fig. 5.4E-F; Fig. 5.S6; Table 5.S7-5.S8**). Although the affinity of **'4042** at

the two receptors is almost undistinguishable (**Fig. 5.S8**), there was a marked difference

in functional activity, with **'4042** consistently being a weaker efficacy partial agonist at

CB2R (**Fig. 5.S6C-D, 5.S8**) versus its essentially full agonism at CB1. This was true for the racemate **'4042** as well as its active enantiomer **'1350** across four separate functional assays including the bioSens-All® BRET assay, the Lance Ultra cAMP assay, TRUPATH BRET2 assay, and the Tango β-Arrestin recruitment assay (**Fig. 5.S8B-D**). Indeed, whereas against CB1R **'1350/ *R*-'4042** had greater relative efficacy against inhibitory G-proteins versus CP-55,940, in CB2R the pattern was reversed, with CP-55,940 being substantially more relatively efficacious than **'1350/ *R*-'4042** (**Fig. 5.4C-F**).

The new CB1R agonist is analgesic with reduced cannabinoid side effects.

Off-target selectivity and pharmacokinetics. Encouraged by the potency and functional selectivity, and the negligible functional differences between the racemic and enantiomeric mixture, we progressed **'4042** into *in vivo* studies for pain relief. We began by investigating the selectivity of **'4042** against potential off-targets. **'4042** was tested first for binding and functional activity against a panel of 320 GPCRs and 46 common drug targets at the PDSP (**Fig. 5.S9**). Little activity was seen except against the melatonin-1 (MT1R), ghrelin (GHSR), Sigma 1 and peripheral benzodiazepine receptors. In secondary validation assays, only weak partial agonism was observed against these receptors, with EC_{50} values greater than 1 µM (**Fig. 5.S9**), 1,000-fold weaker than CB1R. Intriguingly, no agonist activity was seen for the putative cannabinoid receptors GPR55, GPR18, or GPR119. Taken together, **'4042** appears to be selective for CB1 and CB2 receptors over many other integral membrane receptors.

To minimize locomotor effects in pharmacokinetic exposure experiments, we used a dose of 0.2 mg/kg (**Fig. 5.S10A-B**). At low dose, **'4042** was found appreciably in brain and plasma, but not CSF compartments, with higher exposure in brain tissue ($AUC_{0\rightarrow inf}$ = 3180 ng*min/mL) than plasma ($AUC_{0\rightarrow inf}$ = 1350 ng*min/mL). The molecule achieved modest total concentrations in the brain (C_{max} = 16.8 ng/g) and plasma (C_{max} = 5.14 ng/mL or 12 nM) at this dose. A similar pharmacokinetic profile was observed for the positive control CP-55,940 at 0.2 mg/kg, reaching similar maximum concentrations in the brain (C_{max} = 19.2 ng/g versus 16.8 ng/g for **'4042**), and similar half-lives ($T_{1/2}$ = 127 min versus 114 min for **'4042**). The main notable difference was seen in the plasma compartment, with a nearly 10-fold increased C_{max} for CP-55,940 compared to **'4042**. In bulk brain tissue, however, both **'4042** and the control compound CP-55,940 were found to be highly bound to brain tissue proteins, with **'4042** being 10-fold less bound ($f_{u,brain}$ = 0.008) than CP-55,940 ($f_{u,brain}$ = 0.0008; **Table 5.S9**). Correcting for free fraction, this suggests that **'4042** at 0.2 mg/kg dosing reaches a free concentration in the brain of 0.3 nM, approximately the same as its *in vitro* EC_{50} for stimulating $G_{i/o}$ protein recruitment, whereas CP-55,940 is reaching concentrations approximately 10-fold lower than its efficacy for CB1. Finally, the concentration of **'4042** needed to activate the identified off-target receptors even partially is greater than 10,000-fold higher than the observed concentrations, suggesting that activity seen *in vivo* with this ligand reflects on-target engagement.

Anti-allodynia and analgesia. Given its favorable exposure, we next tested the efficacy of **'4042** *in vivo*, in models of pain and inflammation. We first focused on acute

thermal pain. In both tail flick and Hargreaves tests of thermal hypersensitivity, **'4042** increased both tail flick and paw withdrawal latencies in a dose-dependent fashion, showing significant analgesia, namely thresholds above baseline, at as little as 0.1 mg/kg dosed intraperitoneally (i.p.) (**Fig. 5.5A-B**). A similar analgesic effect was observed for the positive control ligand CP-55,940 at slightly higher 0.2 mg/kg doses in the Hargreaves and tail flick tests. Next, we assessed the analgesic properties of **'4042** in the setting of inflammatory pain using the Complete Freud's Adjuvant (CFA) model. As illustrated in **Fig. 5.5C**, 0.2 mg/kg i.p. of **'4042** was not only anti-allodynic, but also analgesic, completely reversing the CFA-induced thermal hypersensitivity to well-above pre-CFA baseline levels.

We next tested the therapeutic potential of **'4042** in the spared nerve injury (SNI) model of neuropathic pain. In contrast to its strong anti-hyperalgesic effect in inflammatory pain models, at 0.2 mg/kg i.p. **'4042** was without effect in SNI mice (**Fig. 5.S11A-B**) but did have a modest anti-allodynic when dosed intrathecally (i.t.; up to 100 µg/kg; **Fig. 5.S11C-D**), suggesting weak effects on mechanical hypersensitivity consistent with literature reports for other CB1R agonists[41–43]. Furthermore, **'4042** did not alter the mechanical thresholds of naïve (non-SNI) animals dosed i.p. at 0.2 mg.kg (**Fig. 5.S11E**), a dose that was frankly analgesic in thermal pain assays. Conversely, relative to their respective vehicle controls, both **'4042** and CP-55,940 strongly reduced the SNI-induced cold allodynia, a hallmark of neuropathic pain, significantly decreasing the average number of acetone-induced nocifensive behaviors, particularly for the paw withdrawals (**Fig. 5.5F**). Finally, in the formalin model of nociceptive pain, an i.p. administration of 0.2

mg/kg **'4042** produced a profound decrease in the duration of both phase 1 and phase 2 nocifensive behaviors (**Fig. 5.5E**) throughout the 60-minute observation period.

On target activity: CB1R vs CB2R. Consistent with CB1R being the target of **'4042** in vivo, pre-treatment with the CB1R selective antagonist AM251 (5.0 mg/kg) completely blocked the analgesic effect of **'4042** in the tail flick assay (**Fig. 5.5F**). In contrast, neither CB2R knockout nor co-treatment with the CB2-selective antagonist SR-144528 (1.0 mg/kg) decreased analgesic effects of **'4042** in the tail flick or Hargreaves assays (**Fig. 5.S11F-H**). We conclude that both the anti-allodynic and analgesic effects of **'4042** are CB1R, but not CB2R, dependent.

Cannabinoid tetrad of behaviors. The cannabinoid "tetrad" of behaviors is commonly used to assess CNS engagement of cannabinoid receptors by novel ligands[11]. In addition to analgesia, this suite of tests measures three common cannabinoid side-effects—hypothermia, catalepsy, and hypolocomotion—as hallmarks of CB1R agonism. Given the novel chemotypes discovered here, we also examined our lead **'4042**, for this panel of potential side-effects.

Reduced "sedation" at analgesic doses. Hypolocomotion, one of the four features of the tetrad, is a commonly assessed proxy for the sedative side-effect of cannabinoids. Sedation is not only an important clinical adverse side effect of cannabinoids, but it also confounds preclinical reflex tests of analgesia, where unimpeded movement of a limb is the endpoint. Intriguingly, while mice treated with **'4042** appeared

less active than those treated with vehicle, '**4042**-injected mice were not sedated (**Fig. 5.6A-B**). Not only would the mice promptly move when slightly provoked (touched, or their housing cylinders slightly disturbed), but in two quantitative and widely-used assays of hypolocomotion and sedation, the open field and rotarod tests, we found no significant differences between '**4042**- and vehicle-treated animals at analgesic doses (**Fig. 5.6A**), although higher doses tended to decrease their overall locomotor activity. Only at the highest (1.0 mg/kg) dose did we record some motor deficits in the rotarod test (**Fig. 5.6B**). In contrast, all analgesic doses tested for the positive control CP-55,940 caused motor impairment in the rotarod test (**Fig. 5.6B**), confounding the aforementioned analgesia results, particularly in the Hargreaves test at 1 and 5 mg/kg (**Fig. 5.5B**). We conclude that '**4042** has a 10-fold therapeutic window for analgesia over sedation, in contrast with the typical cannabinoid CP-55,940, the analgesic effects of which are confounded by their concurrent motor side effects.

Reduced catalepsy at analgesic doses. To determine whether '**4042** induces a second member of the tetrad, catalepsy, we measured the latency '**4042**-injected mice to move all four paws when placed on a vertical wire mesh. As expected, mice injected with the non-cannabinoid control cataleptic, haloperidol, showed dramatic catalepsy (**Fig. 5.6C, 5.S11I**). In contrast, and consistent with its lack of locomotor effects, '**4042** did not induce any observable cataleptic behavior at analgesic doses of 0.2 or 0.5 mg/kg. However, at the high dose of 1 mg/kg, a small amount of catalepsy was observed at 30 minutes pose-dose. Meanwhile, a comparison to CP-55,940 at the same doses showed significantly longer latencies to move all four paws for all tested doses at both 30 minutes

and 1-hour post-injection (**Fig. 5.6C**), mimicking the effects seen on the rotarod (**Fig. 5.6B**).

'4042 induces hypothermia. Finally, looking at the fourth element of the tetrad, hypothermia, we measured the body temperature of mice implanted with telemetric probes for 30 minutes preceding injection, followed by 30 minutes with vehicle only, and finally for 90 minutes following a 0.2 mg/kg i.p. injection of either CP-55,940 or **'4042**. Both CP-55,940 and **'4042** induced hypothermia in compound-treated mice compared to baseline and respective vehicle treatments (**Fig. 5.6D**). Unlike tests of sedation and catalepsy, the degree of **'4042**-induced hypothermia was greater than CP-55,940, particularly at 60-90 minutes post-dose.

Pretreatment with '4042 increases the efficacy of morphine. As '**4042** can induce strong analgesia with reduced side effects, we next asked whether co-treatment of '**4042** with morphine has additional pain-relieving properties. We combined low doses of '**4042** with morphine (3.0 mg/kg, i.p.) and tested the analgesic efficacy of the combination vs morphine alone in the tail flick assay. As illustrated in **Fig. 5.6E**, mice co-injected with morphine (3.0 mg/kg, i.p.) and a non-analgesic (0.05 mg/kg) or a low (0.1 mg/kg) analgesic dose of '**4042** exhibited significantly longer tail flick latencies than did mice injected with morphine alone. This result suggests that the two molecules have at least an additive analgesic effect, consistent with previous studies on both CB1R and CB2R ligand polypharmacy with morphine[44,45].

The novel CB1R agonist is not rewarding. A major limiting factor in an analgesic's clinical utility, particularly opioids, is the potential for misuse because of their intrinsic rewarding properties. To determine whether '**4042** exhibits comparable liabilities, we turned to the conditioned place preference (CPP) test in which mice learn to associate one chamber of the apparatus with a rewarding compound. If mice show a preference for the drug-paired chamber, then the compound is considered to be intrinsically rewarding. As expected, mice injected with morphine significantly increased their preference for the chamber associated with that drug as opposed to its vehicle-associated chamber (**Fig. 5.S11J**). Encouragingly, mice injected with '**4042** spent similar amounts of time in the '**4042**-paired or vehicle-paired chambers, indicating that '**4042** does not induce CPP. Conversely, we observed that mice injected with the cannabinoid CP-55,940 spent significantly more time in the chamber that was paired with its vehicle, suggesting that CP-55,940 may actually induce some aversion, consistent with previous studies in a similar dose range[46].

5.4 Discussion

From a vast library of virtual molecules, structure-based discovery has led to new agonists that not only potently activate CB1R but are also strongly analgesic without key liabilities of classic cannabinoids. Three observations merit emphasis. **First**, from a tangible library of previously unsynthesized, new to the planet molecules, structure-based docking found new chemotypes for the CB1 receptor, physically distinct from previously known ligands. Using structural complementarity, and the wide range of analogs afforded by the new libraries, we optimized these new ligands, leading to a 1.9 nM K_i full agonist

of the CB1R. **Second**, the pose adopted by active enantiomer of **'4042** (**'1350** / **R-'4042**)

in a cryo-EM structure of its complex with CB1R-G$_i$ superposed closely on the docking

prediction, explaining the SAR at atomic resolution and supporting future optimization.

Third, while the new agonist is strongly anti-allodynic and analgesic across a panel of

nociception behavioral assays, **'4042** lacks some of the characteristic adverse drug

reactions of most cannabinoid anagesics, with a 10-fold window between analgesia and

both sedation and catalepsy. Further, we observed no apparent conditioned place

preference or aversion at the highest analgesic, non-sedating dose. These traits are

unusual for cannabinoids, where sedation often closely tracks with analgesia and where

catalepsy is among the "tetrad" of side-effects characteristic of cannabinoid agonists.

Encouragingly, combinations of low doses of **'4042** and morphine show improved

analgesia, suggesting potential for cotreatments to expand the therapeutic window of

each compound on their own.

Three of the four behaviors of the cannabinoid tetrad: hypolocomotion,

hypothermia, and catalepsy, are adverse reactions that limit therapeutic potential of the

fourth, analgesia. Our hope was that by exploring new chemotypes—afforded by the

structure-based approach—some of these adverse aspects of the cannabinoid tetrad

could be reduced. This turned out to be the case. While **'4042** does show some evidence

for hypolocomotion, the molecule is substantially less sedating at analgesic doses than

is the typical cannabinoid, CP-55,940. We also observed a separation between analgesia

and catalepsy, where a small amount of catalepsy was observed only at the highest,

sedating dose of 1 mg/kg, whereas CP-55,940 was cataleptic even at the lowest

analgesic doses. The new agonist neither induced conditioned-place preference nor avoidance at the highest analgesic dose, in contrast to many cannabinoids and to CP-55,940, to which it was compared in this study (**Fig. 5.S11J**). These results suggest that major adverse features of cannabinoids can be reduced, perhaps eliminated, without sacrificing analgesia, at least in mouse models.

Several caveats bear mentioning. The mechanistic bases for the disentanglement of sedation and catalepsy from analgesia remains uncertain. Often, clear differences in functional or subtype selectivity ("ligand bias") support phenotypic differences of different ligands[26,27,38,47]. Here, functional differences between **'4042,** which does not show two characteristic "tetrad" behaviors, and CP-55,940, which does, were modest, with only notable differences shown at CB1 for recruitment of G_{13}. The functional importance of G_{13} in the *in vivo* models is not understood but could be explored in the future. Pronounced differences were, however, seen in the functional effects between the CB1 and CB2 subtypes. Though it is possible that the described CB2 partial agonism could be the hallmark feature separating **'4042** from CP-55,940 and other cannabinoids, studies in cannabinoid receptor knockout animals suggest that catalepsy and sedation are completely ablated in CB1, and not CB2 mice[48]. Additionally, in our hands using CB2 knockout mice, at minimum the analgesic effects are not due to engagement of CB2 receptors. The role of other off-targets, such as antagonism of GPR55 or engagement of TRPV1, could however be explored in the future. At this point we can only lay the differences at the door of the new chemotypes explored. Additionally, although some of the negative cannabinoid side-effects, namely sedation, catalepsy, and aversion, were

spared, **'4042** still exhibited one classical side-effect, hypothermia. Although the initial ligands discovered against CB1R in the docking were at the far low end of the size and hydrophobicity distribution characteristic of cannabinoids, it must be admitted that both terms increased on optimization. Whereas **'4042** remains smaller and more polar than many potent cannabinoids, the distinction has diminished, as is common during small-molecule hit-to-lead optimization[49]. Still, the ability to find relatively small and polar agonists from the large libraries does hint at the ability to find CB1R ligands in this physical property region. Additionally, only the THC-like control ligand CP-55,940 was tested as a comparator to **'4042** in the tetrad tests, while other synthetic cannabinoid controls were not able to be included. However, literature reports suggest that synthetic cannabinoids do indeed induce tetrad phenotypes[50,51], suggesting differentiation of **'4042** from such molecules as well. Finally, while the ability to reduce morphine levels to sub-threshold doses by combination with **'4042** is encouraging, the mechanistic basis for this effect, too, is uncertain. Given the crucial role that opioids continue to play in chronic as well as acute pain management, and their dose-limiting side effects and dependence liabilities, addressing the mechanisms that underlie potential additive or synergistic effects of the novel cannabinoids and opioids merits further research.

Despite these caveats, the main observations of this study should be clear. Docking a library of virtual molecules against CB1 revealed new agonist chemotypes, the most promising of which was optimized to the potent full-agonist **'4042.** A cryo-EM structure of the **R-'4042**-CB1-G$_{i1}$ complex confirmed its docking-predicted pose. The new agonist was strongly analgesic, and unlike most cannabinoids had a 10-fold therapeutic

window over sedation and catalepsy. We suspect that there are still further new chemotypes to be discovered that can separate the dose-limiting side-effect aspects of the cannabinoid tetrad while maintaining analgesic potency, supporting the development of new cannabinoid medicines to treat pain.

5.5 Acknowledgements

Author contributions. T.A.T. and R.M.S. conducted the docking screens with input from B.K.S. Ligand optimization was performed by T.A.T. and C.I.-T. with input from B.K.S., and A.M.. N.G.T., S.G., F.T., and Y.L. performed binding or functional assays with input from T.A.T., C.I.-T., and A.M.. K.K. prepared the CB1-G$_i$ complex, collected cryo-EM data with help from E.S.O., and modelled the structure with help from F.L.. K.K. collected bimane data with help from YS.. T.A.T. and J.M.B. did the drug formulations for in vivo experiments. J.M.B. performed and analyzed the *in vivo* pharmacology experiments assisted by T.A.T., V.C., S.R.R., K.B., and J.B., supervised and coanalyzed by A.I.B.. E.S.O processed data and obtained the cryo-EM map. Y.S. performed the GTP-turnover assays with help from K.K. and E.S.O.. H.K. and C.N. tested select compounds in the panel of G-protein and β-arrestin subtypes with supervision from M.S. and L.S.. C.I.-T. and T.C.H. synthesized bespoke compounds with supervision from A.M.. I.G. performed the colloidal aggregation

screens. Y.S.M. and D.S.R. supervised compound synthesis of Enamine compounds purchased from the ZINC15 database and 12 billion catalog. J.J.I. built the ZINC15 ultra-large libraries. B.K.S., A.I.B., A.M., and K.K. supervised the project. T.A.T. wrote the paper with input from all other authors, and primary editing from B.K.S. **Competing interests.** B.K.S. is a founder of Epiodyne, BlueDolphin and Deep Apple Therapeutics and consults in docking and in the GPCR space. J.J.I. is a cofounder of BlueDolphin and Deep Apple Therapeutics. Y.S.M. is a CEO of Chemspace LLC and a scientific advisor at Enamine, Ltd. D.S.R. is an employee of Enamine, Ltd. H.K., C.N., M.S., and L.S. are employees of Domain Therapeutics North America Inc.. The authors declare no other competing interests.

5.6 Figures

A — ZINC15 cLogP 3 - 5, MW 350 - 500+ Da, DOCK3.7. 74 million docked, 18 million favorably scored. DOCK score = $E_{VDW} + E_{ES} + E_{LDS}$. Computational filters: Structural diversity, Chemical novelty, H-bond to Ser383. Manual filters: Visually inspect poses.

B — CB1R ligands, Top 300K Docked, Purchased. cLogP vs Molecular Weight (Da).

C — 9/46 hits, 19.6% hit rate. Compounds [10 μM] vs % [^{3}H]-CP-55,940 Displacement.

D

'51486 — K_i (nM): 731 Max TC: 0.3; MW (amu): 373.40 LE: 0.33; cLogP: 3.03 LLE: 3.1

'0450 — K_i (nM): 691 Max TC: 0.36; MW (amu): 396.83 LE: 0.31; cLogP: 3.95 LLE: 2.2

'7800 — K_i (nM): 1007 Max TC: 0.28; MW (amu): 380.47 LE: 0.31; cLogP: 3.95 LLE: 2.0

'7019 — K_i (nM): 4039 Max TC: 0.3; MW (amu): 418.25 LE: 0.29; cLogP: 4.07 LLE: 1.3

'7902 — MW (amu): 369.86; cLogP: 4.25; Max TC: 0.31

'7337 — MW (amu): 379.44; cLogP: 3.79; Max TC: 0.29

'1038 — MW (amu): 364.45; cLogP: 4.38; Max TC: 0.28

'7218 — MW (amu): 363.45; cLogP: 4.40; Max TC: 0.31

'2443 — MW (amu): 402.31; cLogP: 3.70; Max TC: 0.23

E — % [^{3}H]-CP-55,940 Specific Binding vs Log [Compound], (M)

Compound	pK$_i$ [95% CI]
CP-55,940	8.53 [8.39 – 8.69]
'51486	6.14 [6.01–6.26]
'0450	6.16 [5.99–6.34]
'7800	6.0 [5.78–6.21]
'7019	5.39 [5.27–5.52]

F — Binding poses of '51486, '0450, '7800, '7019 with residues TM1, TM2, TM3, TM6, TM7, H178$^{2.65}$, F177$^{2.64}$, F174$^{2.61}$, S173$^{2.60}$, F170$^{2.57}$, L193$^{3.29}$, F189$^{3.25}$, F200$^{3.36}$, T197$^{3.33}$, S383$^{7.39}$, F379$^{7.35}$, F268^{ECL2}, L359$^{6.51}$, W356$^{6.48}$, W279$^{5.43}$.

Figure 5.1. Large-scale docking of a 74-million molecule library against the CB1R. **A.** Workflow of the docking campaign. **B.** Overlap of physical properties of CB1R ligands versus the top docked and purchased ligands. **C.** Single-point radioligand displacement data for the 46 tested compounds. **D.** 2D structures and properties of the nine hits. **E.** Secondary binding assay for the top four hits. **F.** Docked poses of the top four hits with H-bonds and other binding pocket residues indicated. Data in panels **C.** and **E.** represent mean ± SEM from three independent experiments.
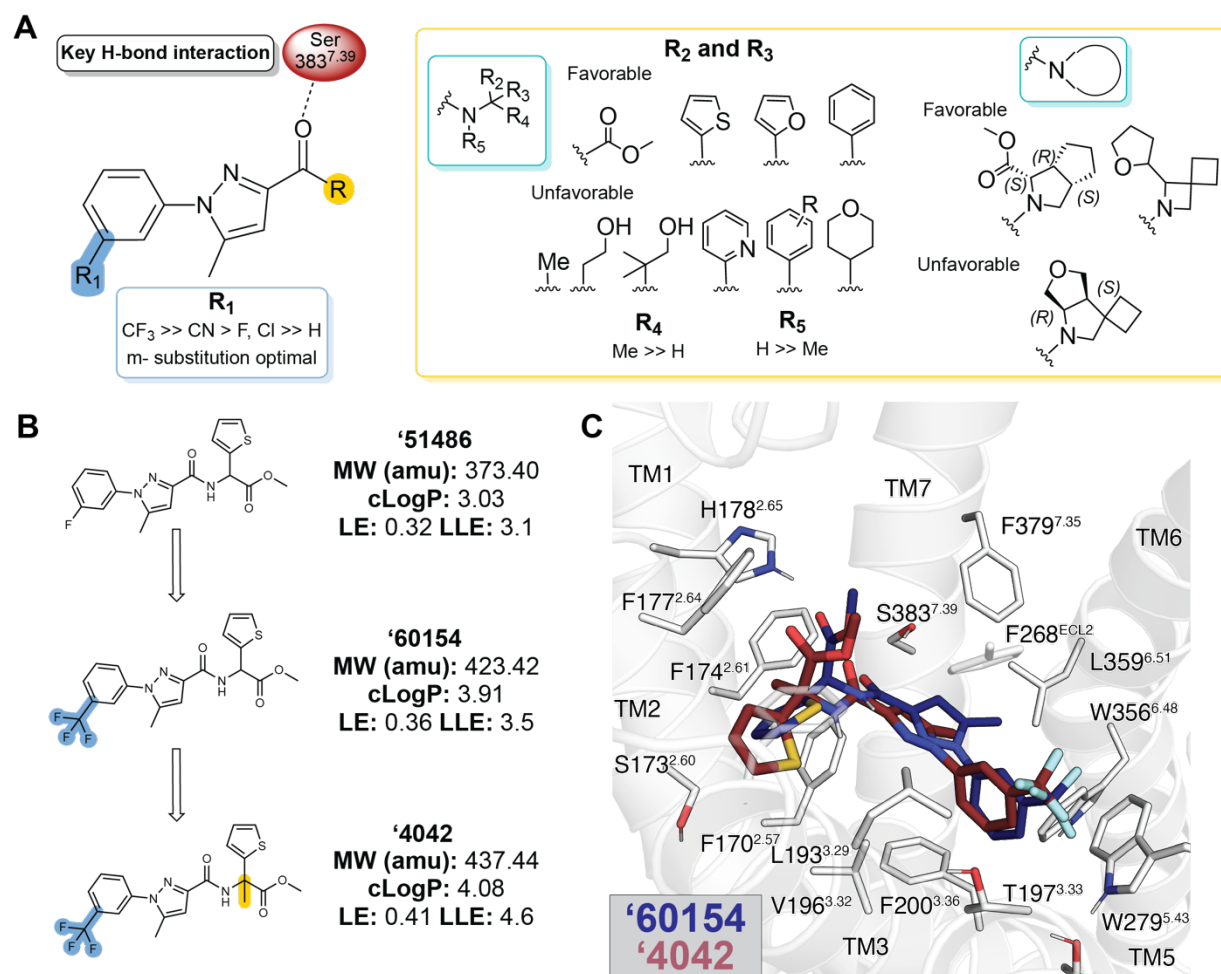
Figure 5.2. Structure-activity relationships and optimization of '51486 to '4042.
A. Pharmacophore model based on the structure-activity relationships discovered via analoging '**51486**. **B.** 2D structures of the docking hit '**51486** and analogs that lead to '**4042**. **C.** Docking predicted pose of '**60154** (navy) and '**4042** (purple).
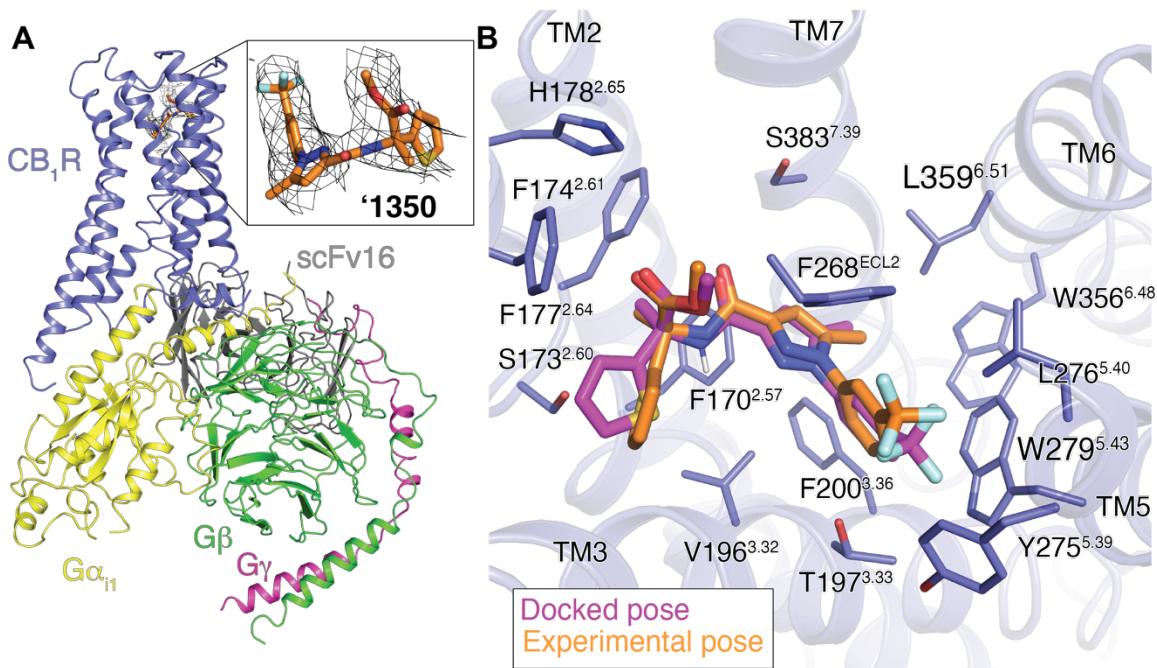
Figure 5.3. Cryo-EM structure of '1350-CB1R-Gi1 complex.
A. Cryo-EM structure of **'1350-CB1R-G$_{i1}$** highlighting the ligand density. **B.** Overlay of the docked pose (magenta) with the experimental pose (orange) of **'1350.**
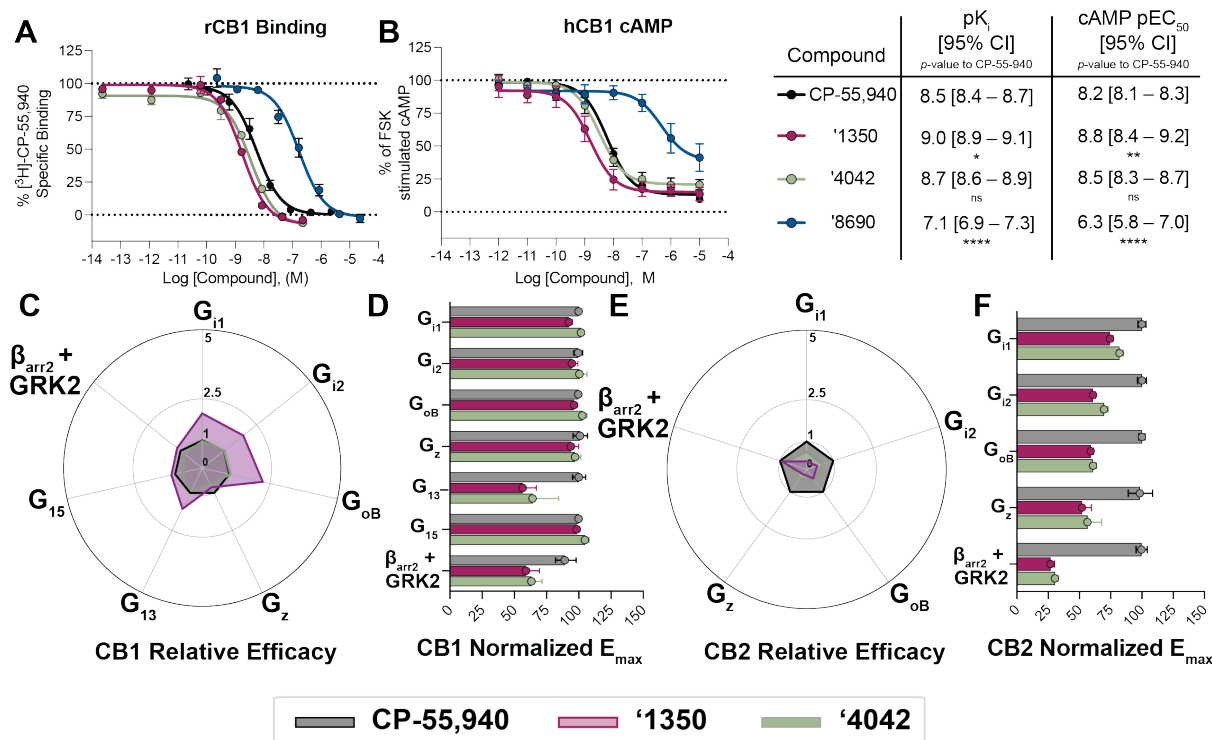
Figure 5.4. Functional activity of '4042 and its active enantiomer '1350.
A. Binding affinity or **B.** Functional cAMP inhibition of **'4042** and its enantiomers **'1350** and **'8690** compared to CP-55,940**.** One-way ANOVA statistical significance of individual pKi (**A**) or pEC50 (**B**) comparisons to CP-55,940 after correction with Dunnett's test of multiple hypotheses are depicted in the table; ns = not significant, * p<0.05, ** p<0.01, **** p<0.001. **C.** Relative efficacy of **'1350** and **'4042** compared to CP-55,940 at hCB1. **D.** Normalized E_{max} from the experiments in **C. E.** Relative efficacy of **'1350** and **'4042** compared to CP-55940 at hCB2. **F.** Normalized E_{max} from the experiments in **E.** Data in **A. & B.** represent mean ± SEM from three independent experiments. Data in **D & F.** represent mean ± 95% CI of the best-fit E_{max} value from two to four independent experiments.

A **Tail Flick**

B **Hargreaves**

C **CFA (Hargreaves)**

D **Acetone Test**

E **Formalin Test**

F **Tail Flick**

235
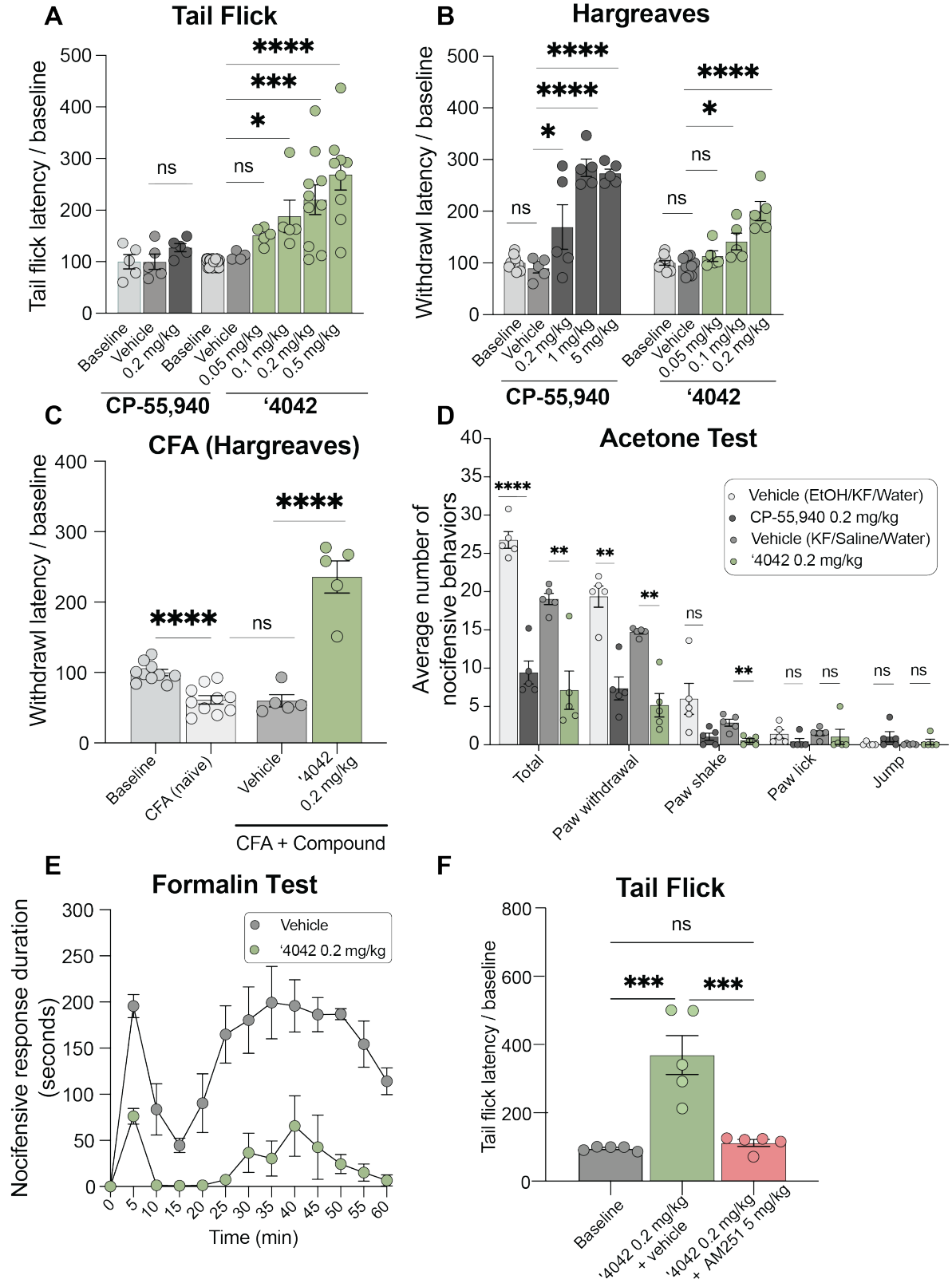
Figure 5.5. In vivo analgesic profile of '4042.
A. Dose-response activity in the tail flick assay for **'4042** (0.05 and 0.1 mg/kg, $n = 5$; 0.2 and 0.5 mg/kg $n = 10$; one-way ANOVA, $F_{(4, 54)} = 18.5$, $P < 0.0001$; asterisks define individual group differences to respective vehicle control using Dunnett's multiple comparisons post-hoc test correction) and CP-55,940 ($n = 5$; unpaired two-tailed t-test, $t_{(8)} = 1.62$, $P > 0.05$). **B.** Dose-response activity in the Hargreaves assay for **'4042** ($n = 5$; one-way ANOVA, $F_{(3, 21)} = 16.26$, $P < 0.0001$; asterisks define individual group differences to respective vehicle control using Dunnett's multiple comparisons post-hoc test correction) and CP-55,940 ($n = 5$; one-way ANOVA, $F_{(4, 25)} = 26.16$, $P < 0.0001$; asterisks define individual group differences to respective vehicle control using Dunnett's multiple comparisons post-hoc test correction). **C.** Hargreaves test of **'4042** ($n = 5 - 10$ per group; two-tailed unpaired t-test, **'4042** versus vehicle: $t_{(8)} = 7.2$, $P < 0.0001$; vehicle versus CFA: $t_{(13)} = 0.13$, $P = 0.89$) after CFA treatment (two-tailed unpaired t-test, CFA versus baseline: $t_{(18)} = 5.2$, $P < 0.0001$). **D.** Chemical hyperalgesia test after spared nerve injury (all $n = 5$; **'4042** vs. vehicle: multiple two-tailed unpaired t-tests, total: $t_{(8)} = 4.6$, $P = 0.007$; paw withdrawal: $t_{(8)} = 6.2$, $P = 0.001$; paw shake: $t_{(8)} = 4.5$, $P = 0.007$; paw lick: $t_{(8)} = 0.4$, $P > 0.05$; jump: $t_{(8)} = 0.8$, $P > 0.05$; CP-55,940 vs. vehicle: multiple two-tailed unpaired t-tests, total: $t_{(8)} = 9.3$, $P < 0.0001$; paw withdrawal: $t_{(8)} = 5.9$, $P = 0.001$; paw shake: $t_{(8)} = 2.4$, $P > 0.05$; paw lick: $t_{(8)} = 1.5$, $P > 0.05$; jump: $t_{(8)} = 1.4$, $P > 0.05$; asterisks define differences to vehicle control after the Holm-Šídák multiple comparisons post-hoc test correction). **E.** Nocifensive response duration after formalin treatment ($n = 5$; multiple two-tailed unpaired t-tests at each timepoint with the Holm-Šídák post-hoc test correction; all times *$P < 0.05 -$ ****$P < 0.0001$ except 0 min. and 15 min., not significant). **F.** Tail flick latency after co-treatment with the selective CB1 antagonist AM251 (all $n = 5$; one-way ANOVA, $F_{(2, 17)} = 29.9$, $P < 0.0001$; asterisks define individual group differences to baseline control after Tukey's multiple comparisons post-hoc test correction.
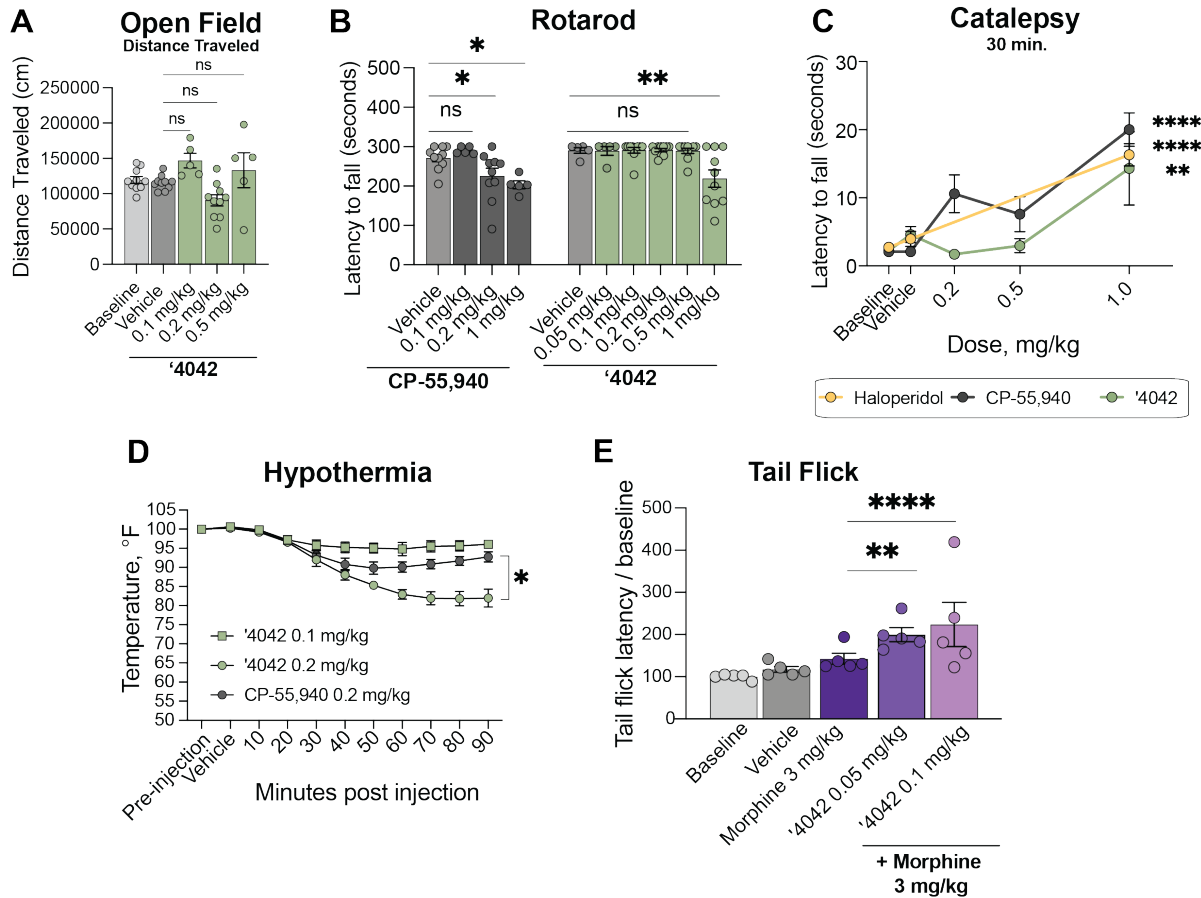
Figure 5.6. In vivo side-effect and cotreatment profile of '4042.
A. Dose-response of **'4042** in the open-field test of hypolocomotion (0.1 and 0.5 mg/kg, n = 5; 0.2 mg/kg n = 10; one-way ANOVA, $F(3, 26)$ = 4.0, P = 0.02; asterisks define individual group differences to vehicle control after Dunnett's multiple comparisons post-hoc test correction). **B.** Rotarod test of sedation comparison of CP-55,940 (all n = 5 except 0.2 mg/kg n = 10; one-way ANOVA, $F(4, 30)$ = 3.5, P = 0.02; asterisks define individual group differences to respective vehicle control after Dunnett's multiple comparisons post-hoc test correction) to **'4042** (all n = 10 except 0.05 mg/kg n = 5; one-way ANOVA, $F(5, 44)$ = 6.2, P = 0.002; asterisks define individual group differences to respective vehicle control after Dunnett's multiple comparisons post-hoc test correction). **C.** Mesh grip test of catalepsy at 30 minutes pose-dose. Comparison of CP-55,940 (n = 5-10; two-way ANOVA; time x drug treatment interaction: $F(6, 78)$ = 5.34, $P < 0.0001$; time: $F(2, 78)$ = 24.7, $P < 0.0001$; drug treatment: $F(3, 78)$ = 20.3, $P < 0.0001$; asterisks define difference to respective vehicle control), haloperidol (n = 5; two-way ANOVA; time x drug treatment interaction: $F(2, 24)$ = 8.7, P = 0.002; time: $F(2, 24)$ = 15.7, $P < 0.0001$; drug treatment: $F(1, 24)$ = 31.7, $P < 0.0001$; asterisks define difference to respective vehicle control), and **'4042** (n = 5; two-way ANOVA; time x drug treatment interaction: $F(6, 48)$ = 2.1, $P > 0.05$; time: $F(2, 48)$ = 3.9, P = 0.03; drug treatment: $F(3, 48)$ = 6.8, $P < 0.001$; asterisks define difference to respective vehicle control). One representative vehicle control shown for simplicity. Data at 1 hr timepoint are in **Fig. 5.S11**. **D.** Body temperatures of mice treated with CP-55,940 or **'4042**. Pre-injection and vehicle values are averages over 30 minutes (Continued on the next page.)

(Continued from previous page.) (n = 5 mice per group, separate vehicle groups for CP-55,940 and **'4042**; CP-55,940 vs. **'4042:** multiple two-tailed unpaired t-tests with the Holm-Šídák post-hoc test correction; all times after 60 minutes *P < 0.05). **E.** Cotreatment of subthreshold morphine with **'4042** on the tail flick (all n = 5; two-way ANOVA; single drug x polypharmacy interaction: $F_{(2, 24)}$ = 7.5, P = 0.003; single drug: $F_{(2, 24)}$ = 5.5, P = 0. 01; polypharmacy treatment: $F_{(1, 24)}$ = 104.2, P < 0.0001; asterisks define cotreatment differences to morphine alone (3 mg/kg) using Dunnett's multiple comparisons post-hoc test correction). For all statistical tests: ns, not significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. All data represent mean ± SEM of 5-10 animals.

Residue Number	Residue Type	Polarity	Kyte & Doolittle Scale [-4.5-4.5]	Eisenberg Scale [-2.5-1.38]
108	PHE	NP	2.80	1.19
170	PHE	NP	2.80	1.19
173	SER	P	-0.80	-0.18
174	PHE	NP	2.80	1.19
177	PHE	NP	2.80	1.19
178	HIS	C/P	-3.20	-0.4
189	PHE	NP	2.80	1.19
192	LYS	C	-3.90	-1.5
193	LEU	NP	3.80	1.06
196	VAL	NP	4.20	1.08
197	THR	P	-0.70	-0.05
200	PHE	NP	2.80	1.19
267	ILE	NP	4.50	1.38
268	PHE	NP	2.80	1.19
269	PRO	NP	-1.60	0.12
271	ILE	NP	4.50	1.38
275	TYR	P	-1.30	0.26
276	LEU	NP	3.80	1.06
279	TRP	NP	-0.90	0.81
359	LEU	NP	3.80	1.06
363	MET	NP	1.90	0.64
379	PHE	NP	2.80	1.19
383	SER	P	-0.80	-0.18
386	CYS	NP	2.50	0.29
		Average	1.59	0.68

Figure 5.S1. Hydrophobicity calculations for the hCB1R orthosteric pocket based on PDB: 5XR8.
Residues within 5 Å of AM841 are considered. **A.** Depiction of the hCB1 orthosteric pocket, colored by the Eisenberg Scale, where darker red colors indicate more hydrophobic residues and lighter red or gray colors indicate less hydrophobic residues. **B.** A table of the residues within 5 Å of AM841, with their polarity class, and two hydrophobicity scores indicated.

239
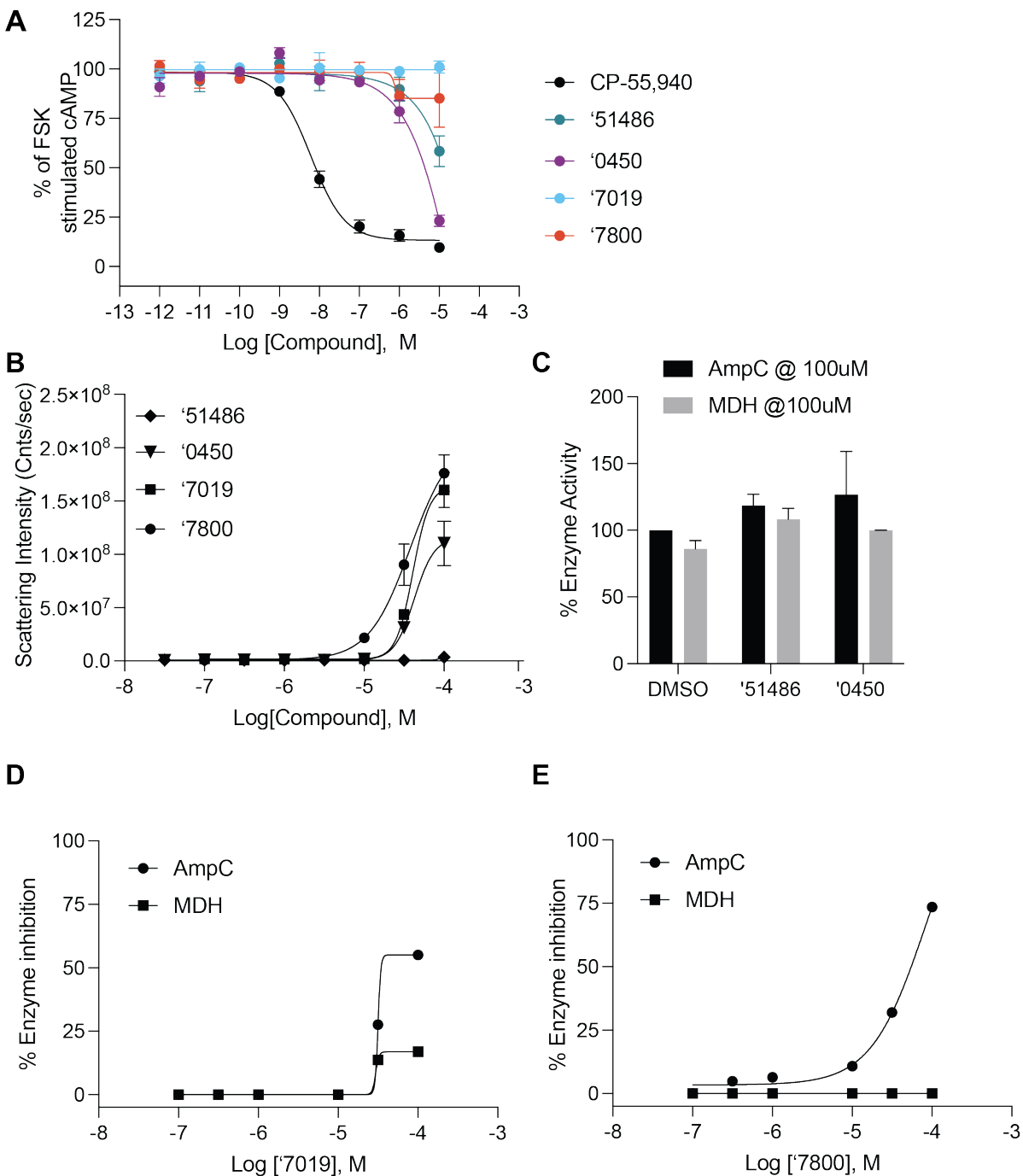
Figure 5.S2. Functional measurements for a subset of screening hits.
A. Functional cAMP inhibition at hCB1R by the four most potent docking hits. **B.** Scattering intensity in dynamic light scattering experiments of colloidal aggregation. **C.** Inhibition of the off-target enzymes MDH and AmpC Beta-lactamase at 100 uM. **D.** and **E.** Single-point inhibition of the off-target enzymes MDH and AmpC Beta-lactamase by **'7019 (D.)** and **'7800 (E.).** All data represent mean ± SEM of three independent experiments in triplicate except **B.** which represents one independent experiment in triplicate.
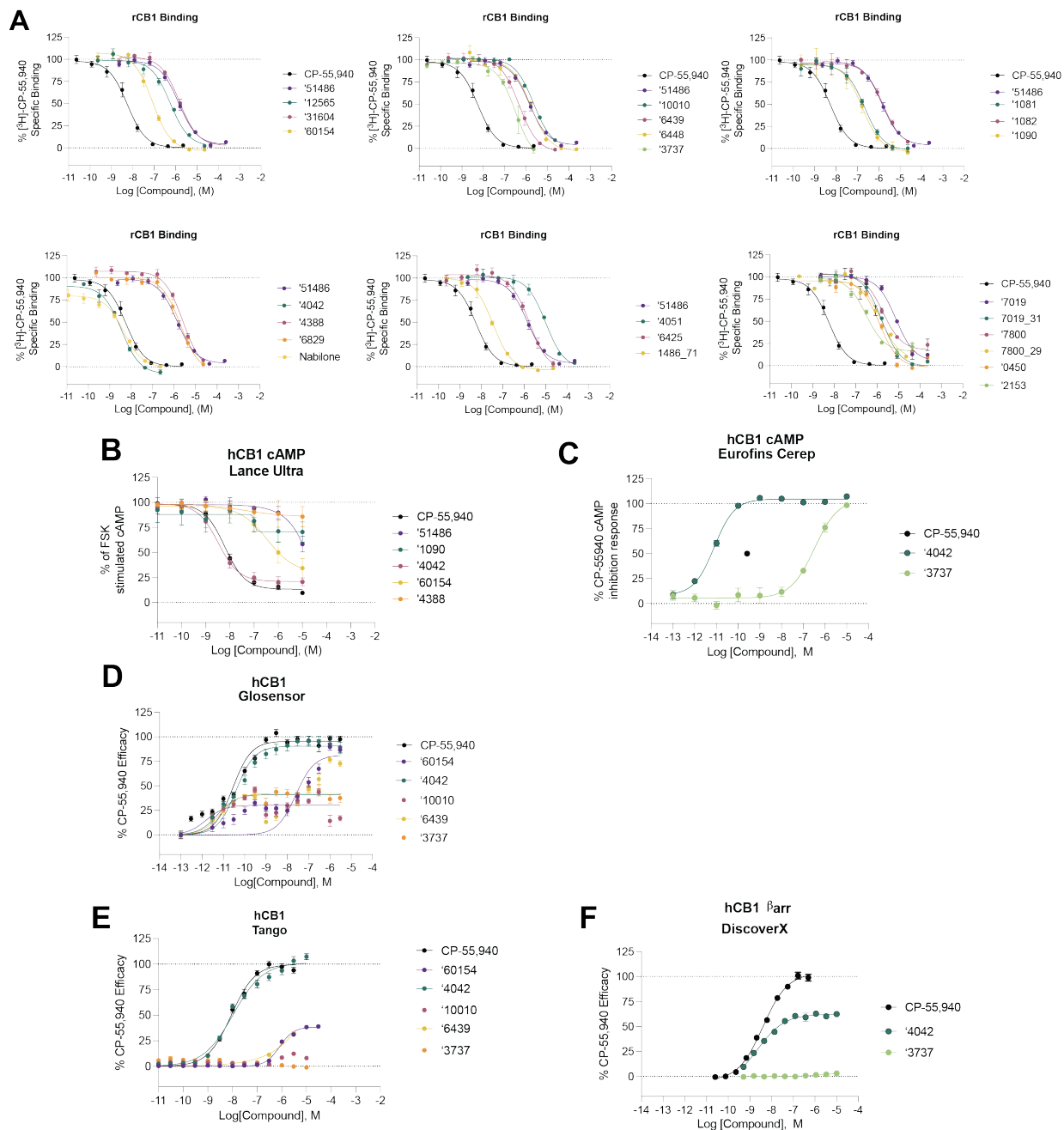
Figure 5.S3. hCB1 binding and functional data for analogs.
A. Competition binding data for primary hits and a subset of their analogs at hCB1. **B-D.**
Functional cAMP inhibition for a subset of analogs at hCB1 across three separate assays.
E-F. Functional ß$_{arr}$ recruitment for a subset of analogs. All data represent mean ± SEM
of at least three independent experiments in triplicate except **C.** and **F.** which represent
one independent experiment in triplicate. Best fit values can be found in **Table 5.S2**.
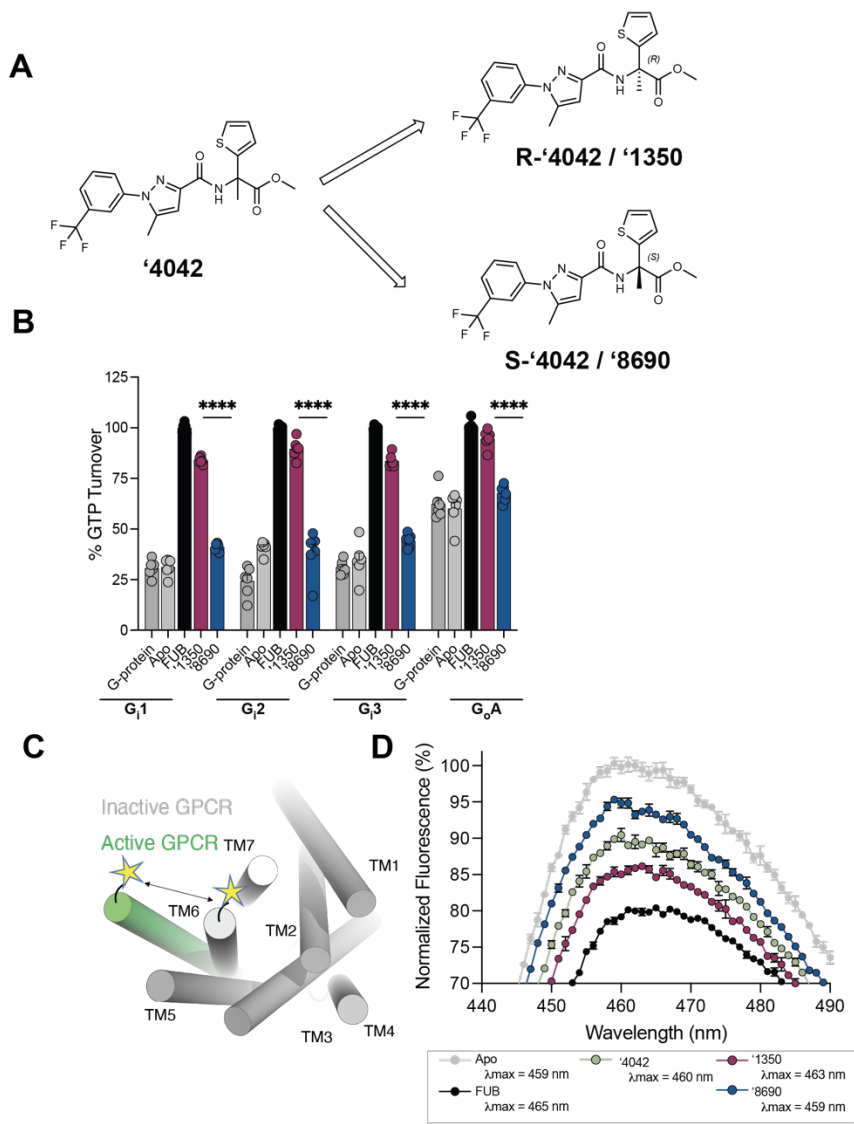
Figure 5.S4. Additional pharmacological characterization of '4042 and its enantiomers.

A. Chiral column purification led to the separation of two independent enantiomers, **'1350** and **'8690**. '1350 was determined to be **R-'4042** from the Cryo-EM structure. **B.** GTPase Glo assay characterizing GTP turnover of G-proteins $G_{i1-3/o}$. **C.** Schematic of the environmentally sensitive fluorophore Monobromobimane (Bimane) which when site-specifically labeled (e.g. on TM6) acts as a conformational reporter. **D**. Compared to the apo (grey), the spectrum of full agonist MDMB-fubinaca (Fub)-bound CB1 (black) shows a decrease in intensity and a blue-shift in λ_{max} (Apo 459 nm to Fub 465 nm). The bimane spectrum of **'8690** (λ_{max} 459 nm, blue) is more similar to apo and the spectrum of **'1350** (λ_{max} 463 nm, magenta) is closer to that of Fub. The spectrum of the racemate, **'4042** (green) is between **'1350** (**R-'4042**) and **'8690** (**S-'4042**). All data represent mean ± SEM of three independent experiments in triplicate.

A

CB1 purification
Sf9 insect cells → Solubilization → Ni column → anti FLAG column → SEC column

scFv16 purification
T. ni insect cells → Pellet cells → super-natant → Ni column → SEC column

G$_i$ heterotrimer purification
T. ni insect cells → Solubilization → Ni column → 3C protease → reverse Ni column → MonoQ column, -P$_i$

CB1-G$_i$ complex formation
CB1 → +agonist 1 hour → agonist-CB1 → +G$_i$ 1 hour → agonist-modulator-CB1 - G$_i$ → +scFv16 +apyrase O/N → nucleotide-free complex → anti FLAG column → SEC column → cryoEM grids

SEC trace
'1350-bound CB1-G$_i$1 complex   CB1

B

8,324 micrographs → motion correction ctf correction autopicking → 4,967,593 particles → particle inspection 2x 2D classification 3D classification

Relion
CryoSPARC

3D classification

GSFSC Resolution: 3.35Å

No Mask (3.7Å)
Spherical (3.5Å)
Loose (3.4Å)
Tight (3.3Å)
Corrected (3.4Å)

Local Resolution (Å)

465,411 particles ← 2D classification Non-uniform Refinement

TM1   TM2   TM3   TM4   TM5   TM6   TM7

Figure 5.S5. Cryo-EM sample preparation and data processing.
A. Purification of hCB1, scFv16, the G_i heterotrimer, and complex formation protocols.
B. Cryo-EM data processing flow chart of CB1, including particle selection, classifications, and density map reconstruction. Details can be found in **Table 5.S3**.
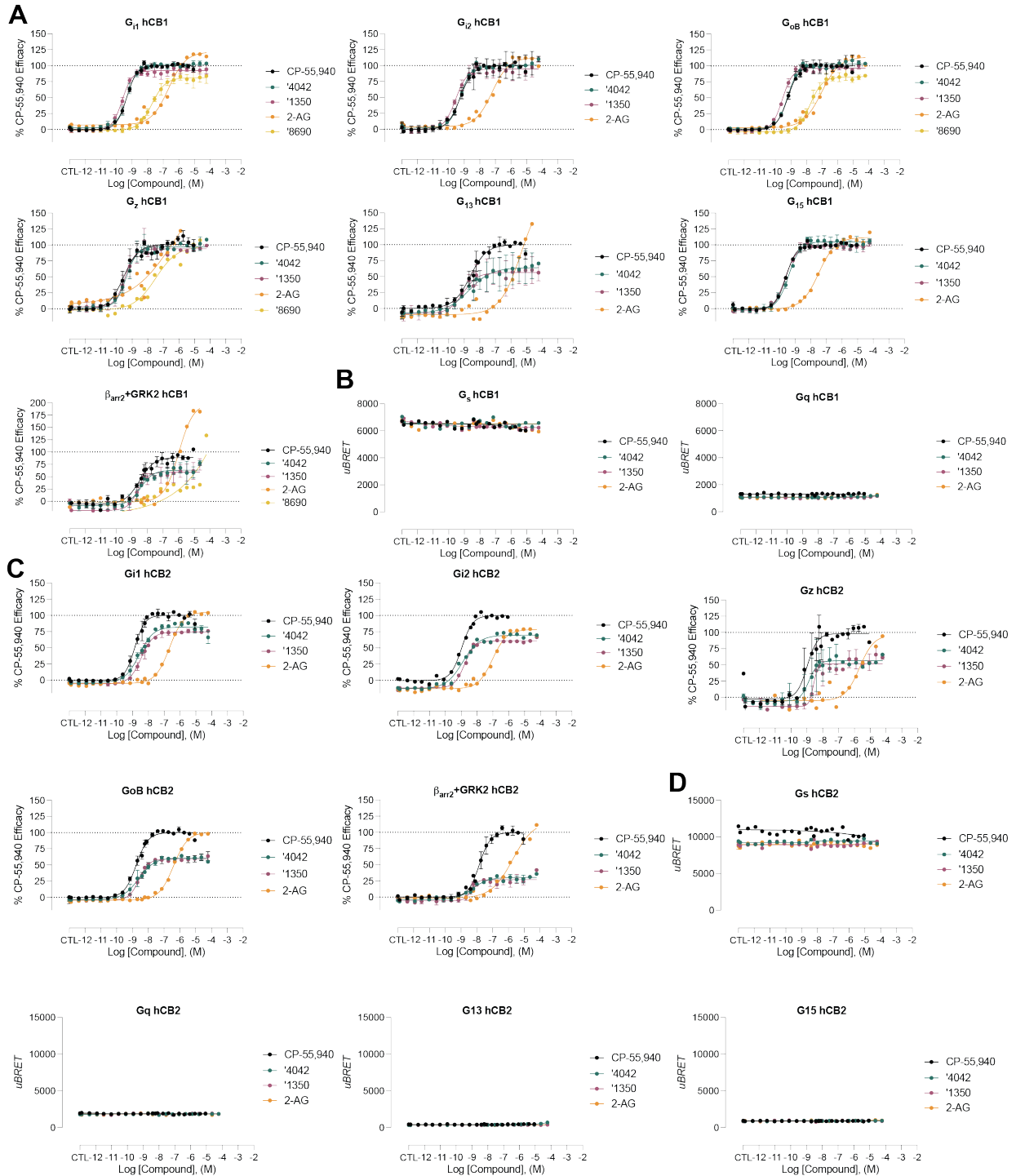
Figure 5.S6. hCB1/2 functional data for select analogs in the bioSens-All® platform.
A. Normalized activity for select analogs versus a panel of sensors in hCB1-expressing cells. **B.** Raw BRET activity for select ana- logs versus G_s and G_q in hCB1-expressing cells. **C.** Normalized activity for select analogs versus a panel of sensors in hCB2-expressing cells. **D.** Raw BRET activity for select analogs versus G_s, G_q, G_{12}, and G_{15} in hCB2-expressing cells. Best fit values can be found in **Tables 5.S5 & 5.S8**.
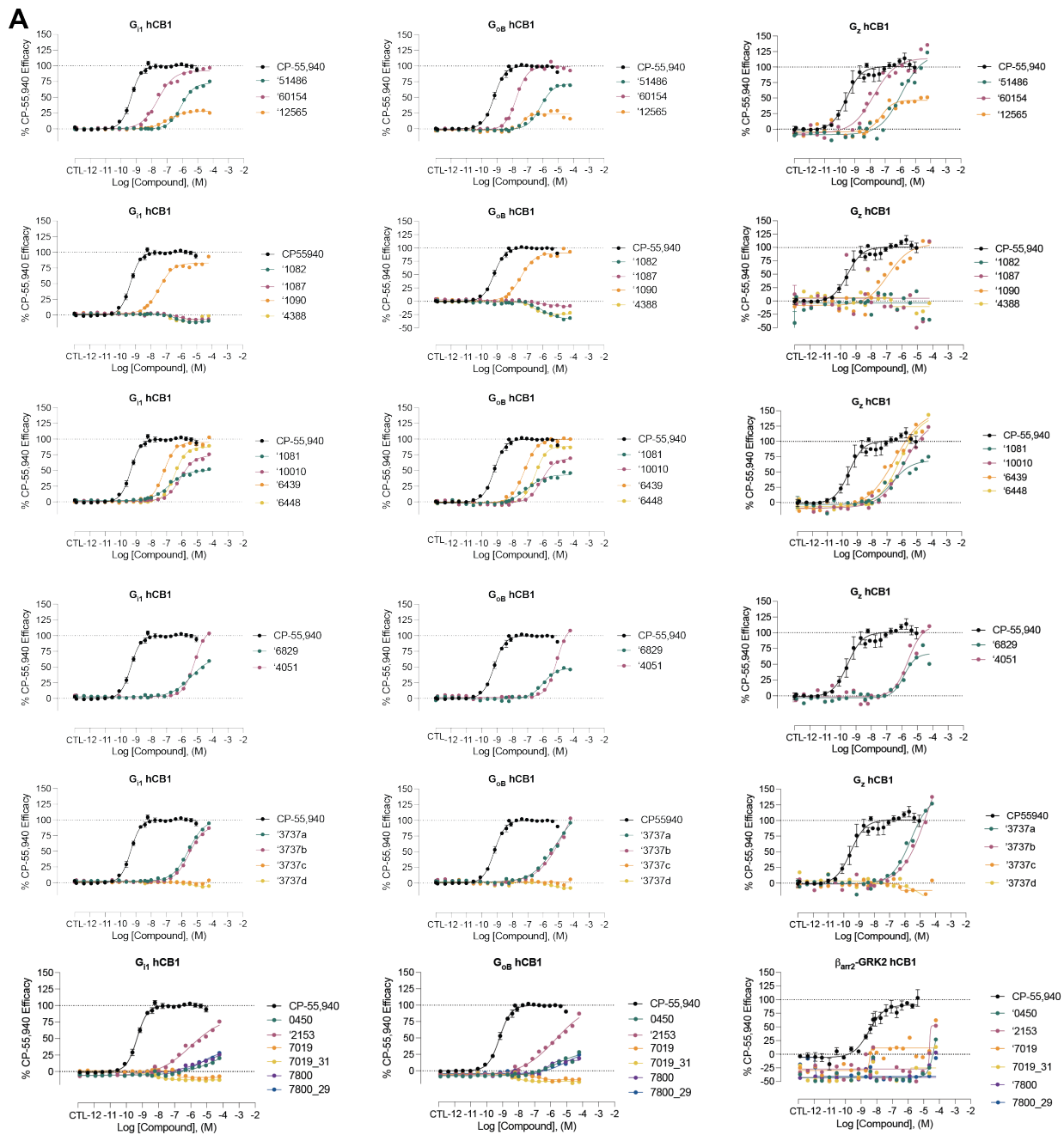
Figure 5.S7. hCB1 functional data for select analogs in the bioSens-All® platform.
A. Normalized activity for select analogs versus a panel of sensors in hCB1-expressing cells. Best fit values can be found in **Table 5.S4**.
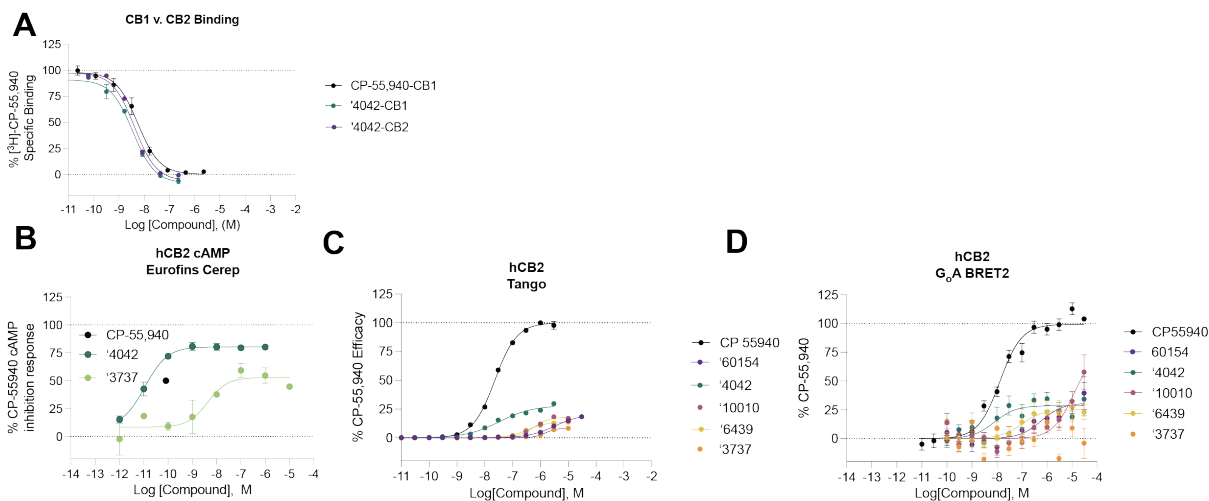
Figure 5.S8. CB2 binding and functional data for select analogs.
A. Competition binding data shows that **'4042** is modestly more potent at CB1 than CB2 (rCB1 pKi = 8.7 (95% CI 8.60 – 8.86), hCB2 pKi = 8.6 (95% CI 8.55 – 8.77); $t(4)$ =6.5, p = 0.003). **B-D.** Functional cAMP inhibition for a subset of analogs at hCB2 across three separate assays. All data represent mean ± SEM of three independent experiments in triplicate except **B.** which represents one independent experiment in triplicate. Best fit values can be found in **Table 5.S7**.
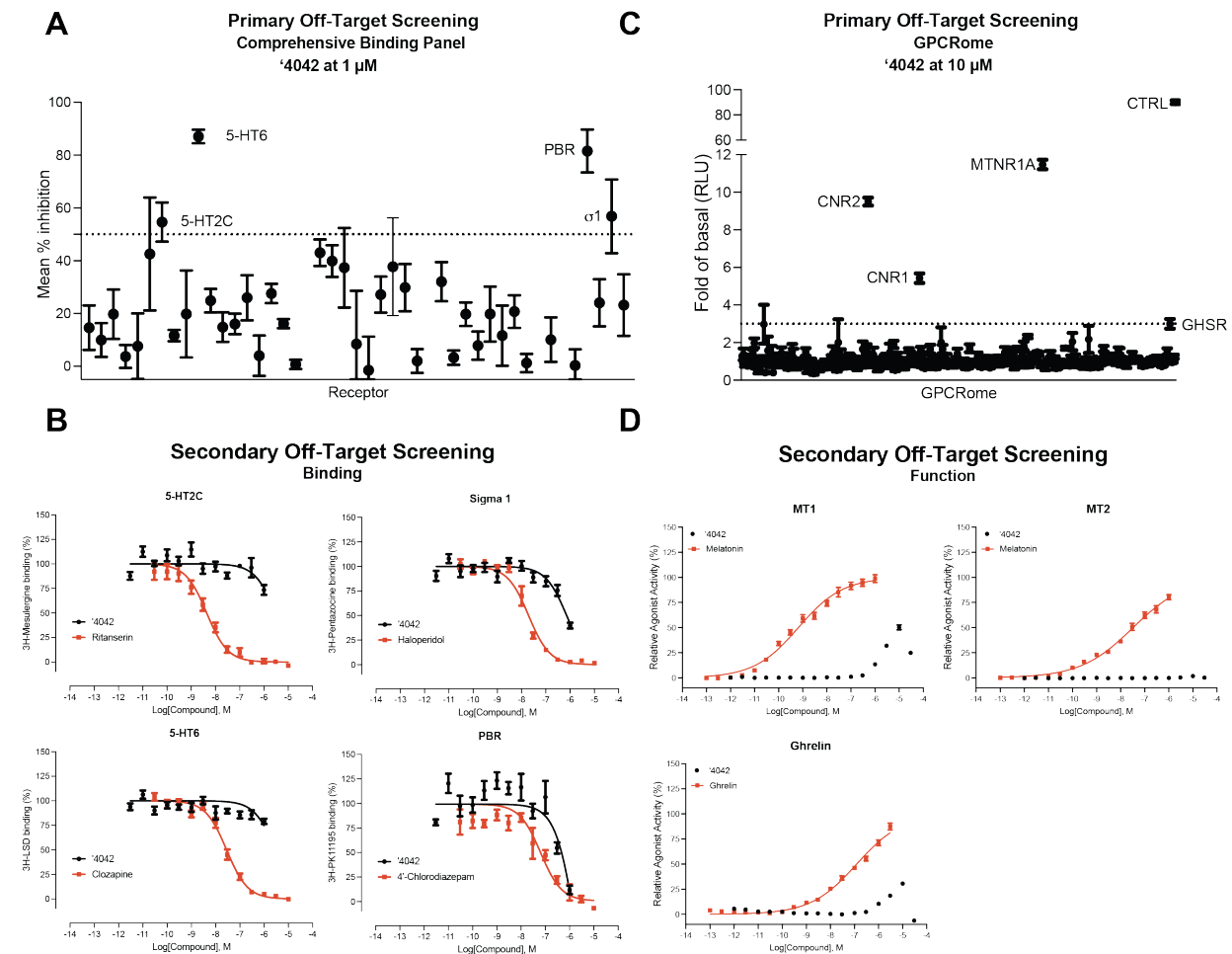
Figure 5.S9. Off-target profiling of '4042.
A. Comprehensive binding data against a panel of 45 common GPCR and non-GPCR drug targets. **B.** Follow-up dose response binding experiments for targets with > 50% inhibition in the single-point experiments. **C.** TANGO screens against a panel of 320 GPCRs for **'4042**. **D.** Follow-up dose response functional experiments for targets with > 3-fold activation in the single-point experiments. Data in **A.**, **C.,** and **D.** represent mean ± SEM of 3 independent experiments in triplicate. Data in **B.** represent mean ± SEM of 2 independent experiments in triplicate except 5-HT6 which is 3 independent experiments in triplicate.

A **'4042**

B **CP-55,940**
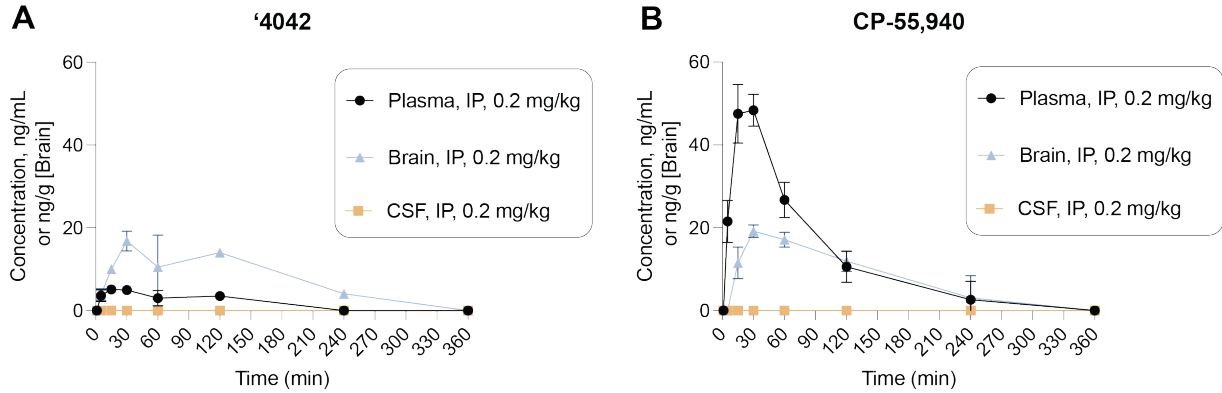
Sample	Administration	Dose, mg/kg	'4042 Pharmacokinetic Parameters					
			Tmax, Min	Cmax, ng/ml (g)	AUC_{0-t} (AUClast) ng*min/ml (g)	$AUC_{0-\infty}$ (AUCINF_obs), ng*min/ml (g)	$T_{1/2}$ (HL_Lambda_z), min	K_{el} (Lambda_z), min^{-1}
Plasma	IP	0.2	15.0	5.14	446	1350	178	0.00388
Brain	IP	0.2	30.0	16.8	2510	3180	114	0.00609
CSF	IP	0.2	ND	BQL	BQL	ND	ND	ND

Sample	Administration	Dose, mg/kg	CP-55,940 Pharmacokinetic Parameters					
			Tmax, Min	Cmax, ng/ml (g)	AUC_{0-t} (AUClast) ng*min/ml (g)	$AUC_{0-\infty}$ (AUCINF_obs), ng*min/ml (g)	$T_{1/2}$ (HL_Lambda_z), min	K_{el} (Lambda_z), min^{-1}
Plasma			30.0	48.4	3360	3990	41.5	0.0167
Brain	IP	0.2	30.0	19.2	1700	3880	127	0.098
CSF			ND	BQL	BQL	ND	ND	ND

Figure 5.S10. Pharmacokinetic profiles of '4042 compared to CP-55.940.
Pharmacokinetic profile of **'4042** (**A**.) and CP-55,940 (**B**.) after a single 0.2 mg/kg dose in brain, CSF, and plasma compartments. Data represent mean ± SEM of 3 animals per timepoint.
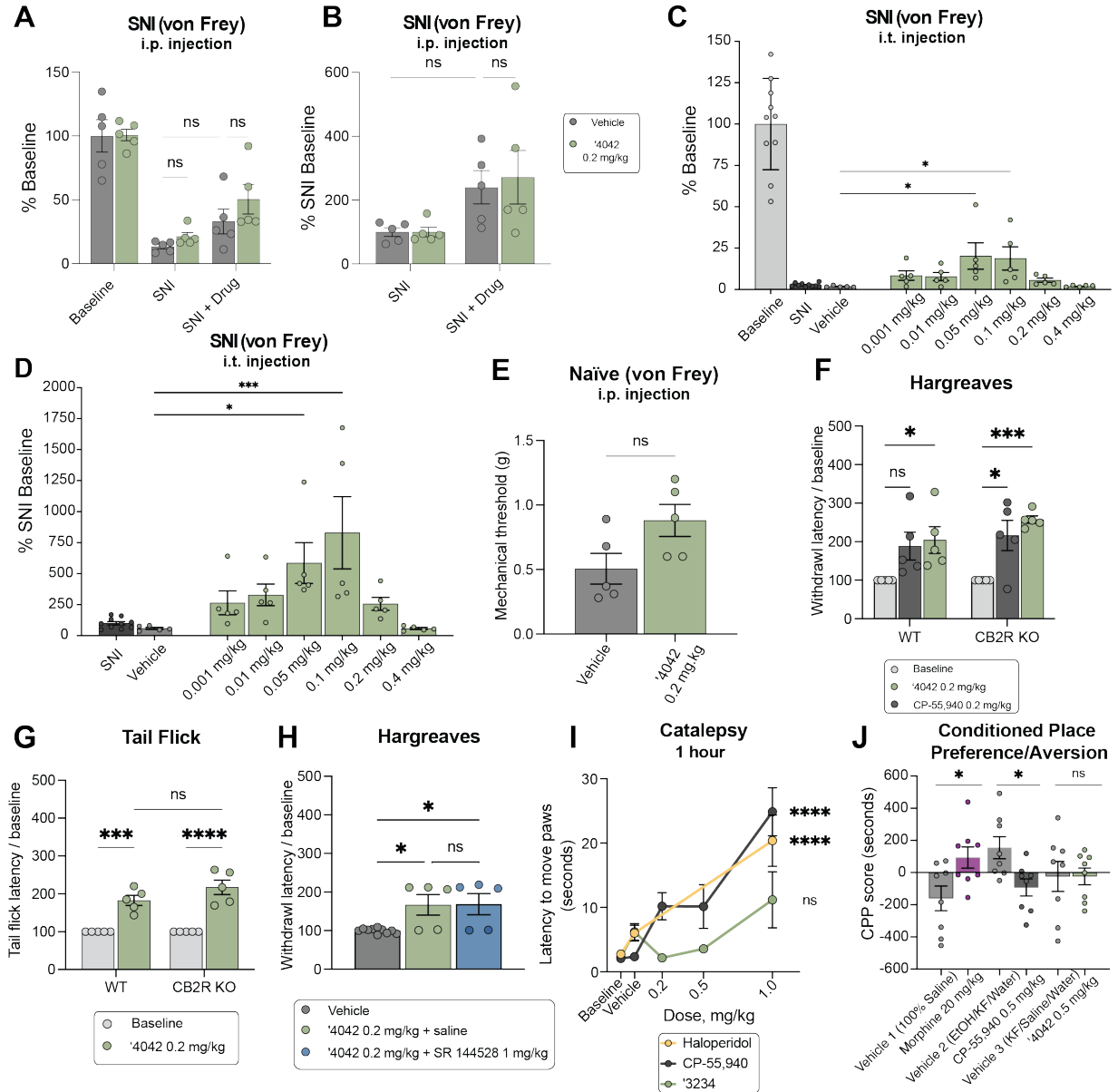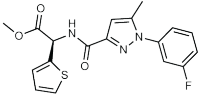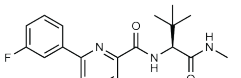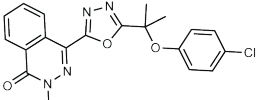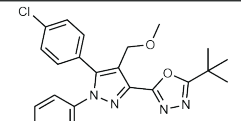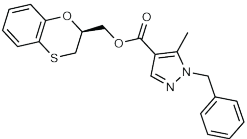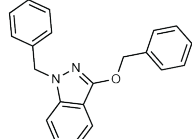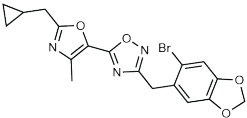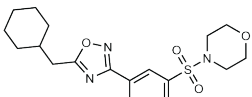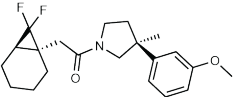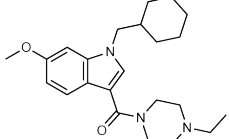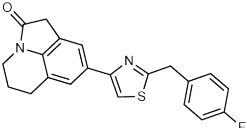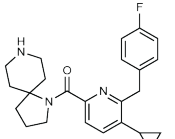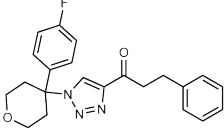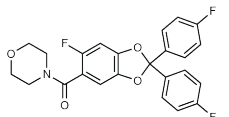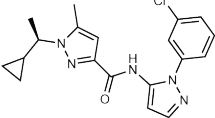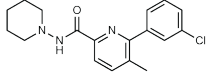
Figure 5.S11. Additional analgesic and side-effect profiles of '4042.
A. Effect of **'4042** (i.p.) in neuropathic pain model in mice after SNI with mechanical allodynia ($n = 5$; two-way ANOVA; SNI x drug treatment interaction: $F_{(2, 24)} = 0.5$, $P > 0.05$; SNI: $F_{(2, 24)} = 51.8$, $P < 0.0001$; drug treatment: $F_{(1, 24)} = 1.6$, $P > 0.05$; asterisks define individual group differences to vehicle control after Tukey's multiple comparisons post-hoc test correction). Data presented are normalized to pre-SNI baseline measurements. **B.** Effect of **'4042** (i.p.) in neuropathic pain model in mice after SNI with mechanical allodynia ($n = 5$; two-way ANOVA; SNI x drug treatment interaction: $F_{(1, 16)} = 0.1$, $P > 0.05$; SNI: $F_{(1, 16)} = 9.6$, $P = 0.007$; drug treatment: $F_{(1, 16)} = 0.1$, $P > 0.05$; asterisks define individual group differences to vehicle control after Tukey's multiple comparisons post-hoc test correction). Data presented are normalized to post-SNI baseline measurements. **C.** Effect of **'4042** (i.t.) in neuropathic pain model in mice after SNI with mechanical allodynia (Continued on the next page.)
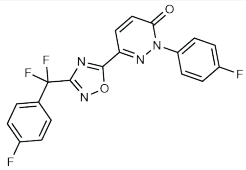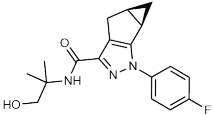
(Continued from previous page.) (n = 5; one-way ANOVA, $F(6, 28)$ = 4.2, P = 0.004; asterisks define individual group differences to vehicle control after Dunnett's multiple comparisons post-hoc test correction). Data presented are normalized to pre-SNI baseline measurements. **D.** Effect of **'4042** (i.t.) in neuropathic pain model in mice after SNI with mechanical allodynia (n = 5; one-way ANOVA, $F(7, 32)$ = 3.8, P = 0.004; asterisks define individual group differences to vehicle control after Dunnett's multiple comparisons post-hoc test correction). Data presented are normalized to post-SNI baseline measurements. **E.** Effect of **'4042** (i.p.) in naïve (non-SNI) mice in the mechanical assay (all n = 5; two-tailed unpaired t-test, $t(8)$ = 2.17, P > 0.05). **F.** Comparison of the effect of **'4042** and CP-55,940 in wildtype (WT) versus CB2R knockout (KO) mice in the Hargreaves assay (all n = 5; two-way ANOVA; genotype x drug treatment interaction: $F(2, 24)$ = 0.5, P > 0.05; genotype: $F(1, 24)$ = 1.6, P > 0.05; drug treatment: $F(2, 24)$ = 13.8, P = 0.0001; asterisks define individual group differences to baseline after Tukey's multiple comparisons post-hoc test correction). **G.** Comparison of the effect of **'4042** in wildtype (WT) versus CB2R knockout (KO) mice in the Tail Flick assay (all n = 5; two-way ANOVA; genotype x drug treatment interaction: $F(1, 16)$ = 2.2, P > 0.05; genotype: $F(1, 16)$ = 2.2, P > 0.05; drug treatment: $F(1, 16)$ = 72.3, P < 0.0001; asterisks define individual group differences to baseline after Šídák's multiple comparisons post-hoc test correction). **H.** Withdrawal latency in the Hargreaves assay after co-treatment with the selective CB2R antagonist SR 144528 (1 mg/kg) (all n = 5; one-way ANOVA, $F(2, 17)$ = 6.6, P = 0.008; asterisks define individual group differences to vehicle control after Tukey's multiple comparisons post-hoc test correction). I. Mesh grip test of catalepsy at 1 hr post-dose. Comparison of CP-55,940 (n = 5-10; two-way ANOVA; time x drug treatment interaction: $F(6, 78)$ = 5.34, P < 0.0001; time: $F(2, 78)$ = 24.7, P < 0.0001; drug treatment: $F(3, 78)$ = 20.3, P < 0.0001; asterisks define difference to respective vehicle control), haloperidol (n = 5; two-way ANOVA; time x drug treatment interaction: $F(2, 24)$ = 8.7, P = 0.002; time: $F(2, 24)$ = 15.7, P < 0.0001; drug treatment: $F(1, 24)$ = 31.7, P < 0.0001; asterisks define difference to respective vehicle control), and **'4042** (n = 5; two-way ANOVA; time x drug treatment interaction: $F(6, 48)$ = 2.1, P > 0.05; time: $F(2, 48)$ = 3.9, P = 0.03; drug treatment: $F(3, 48)$ = 6.8, P < 0.001; asterisks define difference to respective vehicle control). One representative vehicle control shown for simplicity. **J.** Comparison of morphine (n = 8; two-tailed unpaired t-test, $t(14)$ = 2.51, P = 0.03) to CP-55,940 (n = 8; two-tailed unpaired t-test, $t(14)$ = 2.9, P = 0.01) and **'4042** (n = 8; two-tailed unpaired t-test, $t(14)$ = 0.005, P > 0.05) in the Conditioned Place Preference (CPP) test. For all statistical tests: ns, not significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. All data represent mean ± SEM of 5-10 animals.

5.7 Tables

Table 5.S1. Binding affinities for hits identified in initial CB1 docking screen.

Compound	Global rank	rCB1 affinity[a] K_i [95% CI] nM pK_i [95% CI]	Tc[b]	Nearest ChEMBL ligand[c]
 '51486	117390	731 [552 – 969] 6.14 [6.01 –6.26]	0.30	 CHEMBL4110127
 '0450	6582	691 [459 – 1033] 6.16 [5.99 – 6.34]	0.36	 CHEMBL519214
 '7800	12210	1007 [615 – 1654] 6.0 [5.78 – 6.21]	0.28	 CHEMBL3116279
 '7019	20488	4039 [3027 – 5379] 5.39 [5.27 –5.52]	0.24	 CHEMBL472680
 '7218	29322	52.2% [24.79]	0.31	 CHEMBL3347301
 '1038	47606	53.6% [2.91]	0.28	 CHEMBL3890211
 '7337	24720	57.0% [3.04]	0.29	 CHEMBL259699
 '7902	139929	57.1% [0.02]	0.31	 CHEMBL3915046
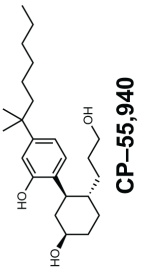
Compound	Global rank	rCB1 affinity[a] K_i [95% CI] nM pK_i [95% CI]	Tc[b]	Nearest ChEMBL ligand[c]
'2443	21964	51.1% [4.87]	0.23	CHEMBL3354970

[a]Binding affinity to rCB1 represented as Ki [95% CI] and pKi [95% CI] from three independent experiments in triplicate when measured. Otherwise, % radioligand displacement [S.D] from three replicates in a single-point competition experiment at 10 µM

[b]Tanimoto coefficient based on ECFP4 fingerprints

[c]Corresponding ChEMBL ligand with the most similar fingerprint

253

Table 5.S2. Binding affinities and functional activities for active analogs at CB1

Compound	rCB1 binding K_i [95% CI] (nM) pK_i [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Cerep cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Glosensor cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Tango ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]
CP–55,940	2.9 [2.05 – 4.2] 8.5 [8.4 – 8.7] 98% [92 – 104]	6.2 [4.7 – 8.0] 8.2 [8.1 – 8.3] 85% [86 – 85]	0.026	0.028 [0.02 – 0.04] 10.6 [10.5 – 10.7] 96% [93 – 99]	8.9 [7.5 – 10.6] 8.1 [8.0 – 8.1] 100% [96 – 104]	4.0 [3.2 –4.9] 8.4 [8.3 – 8.5] 108% [99 – 109]
'4042	1.86 [1.37 – 2.52] 8.7 [8.6 – 8.9] 91% [87 – 95] ns	3.3 [1.9 – 5.6] 8.5 [8.3 – 8.7] 78% [78 – 79] ns	0.008 [0.006 – 0.01] 11.1 [11.0 – 11.2] 96% [102 – 107]	0.039 [2.9 – 5.4] 10.4 [10.3 – 10.5] 91% [87 – 94]	10.7 [8.7 – 13.3] 8.0 [7.9 – 8.1] 102% [98 – 105]	2.3 [2.5 –4.8] 8.7 [8.3 – 9.6] 71% [60 – 65]
'1350	0.95 [.74 – 1.24] 9.02 [8.9 – 9.1] 99% [95 103] *	1.6 [0.7 – 3.6] 8.8 [8.4 – 9.2] 78% [77 – 80] **	--	--	--	--
'8690	90.2 [56.7 – 143] 7.1 [6.9 – 7.3] 98% [92 – 104] ****	473 [109 – 1822] 6.3 [5.8 – 7.0] 53% [45 – 65] ****				

254

Compound	rCB1 binding K_i [95% CI] (nM) pK_i [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Cerep cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Glosensor cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Tango ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]
'60154	44.3 [33.9 – 58.0] 7.4 [7.2 – 7.5] 107% [103 – 112]	351 [93] 6.5 [7.0] 67% [59 – 89]	--	25.2 [16 – 40] 7.6 [7.4 – 7.8] 82% [74 – 189]	819 [718 – 934] 6.1 [6.0 – 6.1] 39% [37 – 40]	--
'1081	116 [76.3 – 178] 6.9 [6.8 – 7.1] 103% [100 – 103]	--	--	--	--	--
'1082	850 [488 – 1491] 6.1 [5.8 – 6.3] 95% [89 – 100]	--	--	--	--	--
'1090	90.8 [42.7 – 192] 7.0 [6.7 – 7.4] 96% [86 – 106]	--	--	--	--	--

Compound	rCB1 binding K$_i$ [95% CI] (nM) pK$_i$ [95% CI] E$_{max}$ [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI] Significance[a]	hCB1 Cerep cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 Glosensor cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 Tango ß-arrestin recruitment EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]
'4388	1360 [998 – 1857] 5.9 [5.7 – 6.0] 108% [104 – 111]	--	--	--	--	--
'4051	5328 [3774–7507] 5.3 [5.1 – 5.4] 101% [98 – 105]	--	--	--	--	--
'6425	934 [583 – 1501] 6.0 [5.8 – 6.2] 104% [99 – 109]	--	--	--	--	--
'6829	1046 [669 – 1643] 6.0 [5.8 – 6.2] 100% [96 – 105]	--	--	--	--	--

256

Compound	rCB1 binding K_i [95% CI] (nM) pK_i [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Cerep cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Glosensor cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Tango ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]
1486_71	18.5 [13.8 – 25.0] 7.7 [7.6 – 7.9] 100% [95 – 105]	--	--	--	--	--
'12565	301 [195 – 462] 6.5 [6.3 – 6.7] 99% [95 – 104]	--	--	--	--	--
'31604	801 [596 –1,076] 6.1 [6.0 – 6.2] 103% [99 – 107]	--	--	--	--	--
'10010	1,196 [952 – 1,505] 5.9 [5.8 – 6.0] 101% [99 – 103]	--	--	N.D	> 10,000	--

257

Compound	rCB1 binding K$_i$ [95% CI] (nM) pK$_i$ [95% CI] E$_{max}$ [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI] Significance[a]	hCB1 Cerep cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 Glosensor cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 Tango ß-arrestin recruitment EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]
'6439	251 [173 – 364] 6.6 [6.4 – 6.8] 102% [97 – 107]	--	--	N.D.	> 10,000	--
'6448	866 [564 – 1,317] 6.1 [5.9 – 6.3] 100% [94 – 106]	--	--	--	--	--
'3737	173 [94 – 322] 6.76 [6.5 – 7.03] 97% [91 – 102]	--	326 [168 – 1044] 6.5 [6.0 – 6.8] 99% [91 – 135]	N.D	> 10,000	> 10,000
7019_31	876 [683 – 1123] 6.1 [6.0 – 6.2] 104% [100 – 108]	--	--	--	--	--
7800_29	825 [396 – 1,755] 6.1 [5.8 – 6.4] 96% [90 – 102]	--	--	--	--	--

258

Compound	rCB1 binding K_i [95% CI] (nM) pK_i [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Lance Ultra cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI] Significance[a]	hCB1 Cerep cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Glosensor cAMP EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 Tango ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	hCB1 DiscoverX ß-arrestin recruitment EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]
 '2153	163 [90 – 287] 6.79 [6.5 – 7.0] 99% [91 – 106]	--	--	--	--	--
 Nabilone	2.3 [1.4 – 3.8] 8.6 [8.4 – 8.9] 80% [75 – 84]	--	--	--	--	--

N.D. = best fit values were not determined due to compound inactivity or poor data quality

-- Not tested

[a]One-way ANOVA statistical significance of individual comparisons to CP-55,940 after correction with Dunnett's test of multiple hypotheses. ns = not significant, * p<0.05, ** p<0.01, *** p<0.001, **** p<0.001

259

Table 5.S3. Cryo-EM data collection, model refinement, and validation statistics.

	EMDB-29898 PDB 8GAG
Data collection and processing	
Magnification	96,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	56.6
Defocus range (μm)	-0.7- -2.0
Pixel size (Å)	0.8521
Symmetry imposed	C1
Initial particle images (no.)	4,967,593
Final particle images (no.)	465,411
Map resolution (Å)	3.3
FSC threshold	0.143
Map resolution range (Å)	2.6-4.2
Refinement	
Initial model used (PDB code)	6N4B
Model resolution (Å)	3.0
FSC threshold	0.143
Model resolution range (Å)	2.7-3.6
Map sharpening B factor (Å^2)	175.3
Model composition	
Non-hydrogen atoms	8,510
Protein residues	1116
Ligands	1
B factors (Å^2)	
Protein	33.4
Ligand	43.83
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.281
Validation	
MolProbity score	2.58
Clashscore	27.65
Poor rotamers (%)	0.11
Ramachandran plot	
Favored (%)	85.40
Allowed (%)	14.60
Disallowed (%)	0

Table 5.S4. Functional activities for select analogs versus a variety of transducers and hCB1 in the bioSens-All® platform.

Compound		hCB1 G$_{i1}$	hCB1 G$_{oB}$	hCB1 G$_z$
CP–55,940	EC$_{50}$ [95% CI] (nM)	0.46 [0.4 – 0.5]	0.63 [0.6 – 0.7]	0.28 [0.18 – 0.5]
	pEC$_{50}$ [95% CI]	9.3 [9.3 – 9.4]	9.2 [9.2 – 9.3]	9.6 [9.4 – 9.8]
	E$_{max}$ [95% CI]	100 [98 – 102]	100 [99 – 101]	102 [102 – 104]
'51486	EC$_{50}$ [95% CI] (nM)	849 [745– 947]	711 [579– 873]	1118 [363– 3447]
	pEC$_{50}$ [95% CI]	6.1 [6.0 – 6.1]	6.2 [6.1 – 6.2]	6.0 [5.5 – 6.4]
	E$_{max}$ [95% CI]	73 [70 – 75]	74 [70 – 75]	130 [996 – 163]
'60154	EC$_{50}$ [95% CI] (nM)	2.5 [1.9 – 3.4]	18.4 [15.8– 21.5]	17.5 [7.5 – 41]
	pEC$_{50}$ [95% CI]	7.6 [7.5 – 7.7]	7.7 [7.7 – 7.8]	7.8 [7.4 – 8.1]
	E$_{max}$ [95% CI]	92 [90 – 94]	100 [98 – 100]	121 [103 – 140]
'1081	EC$_{50}$ [95% CI] (nM)	150 [116 – 197]	91.7 [55 – 153]	225 [73 – 694]
	pEC$_{50}$ [95% CI]	6.8 [6.7 – 6.9]	7.0 [6.8 – 7.3]	6.7 [6.2 – 7.1]
	E$_{max}$ [95% CI]	49 [48 – 51]	45 [43 –47]	72 [55 –89]
'1082	EC$_{50}$ [95% CI] (nM)			
	pEC$_{50}$ [95% CI]	N.D.	N.D.	N.D.
	E$_{max}$ [95% CI]			
'1087	EC$_{50}$ [95% CI] (nM)			
	pEC$_{50}$ [95% CI]	N.D.	N.D.	N.D.
	E$_{max}$ [95% CI]			
'1090	EC$_{50}$ [95% CI] (nM)	35.6 [28.5 – 45]	37.3 [30 – 47]	126 [45 – 354]
	pEC$_{50}$ [95% CI]	7.5 [7.3 – 7.5]	7.4 [7.3 – 7.5]	6.9 [6.5 – 7.4]
	E$_{max}$ [95% CI]	81 [80 –82]	90 [88 –92]	115 [92 – 137]
'4388	EC$_{50}$ [95% CI] (nM)			
	pEC$_{50}$ [95% CI]	N.D.	N.D.	N.D.
	E$_{max}$ [95% CI]			
'6829	EC$_{50}$ [95% CI] (nM)	4056 [2417 – 9228]	1011 [706 – 1449]	1347 [720 – 2522]
	pEC$_{50}$ [95% CI]	5.4 [5.0 – 5.6]	6.0 [5.8 – 6.2]	5.9 [6.0 – 6.2]
	E$_{max}$ [95% CI]	68 [57 – 77]	49 [45 – 53]	70 [57 –83]
'4051	EC$_{50}$ [95% CI] (nM)	6523 [5770 – 7511]	7988 [6709 – 9511]	2431 [1069 – 5531]
	pEC$_{50}$ [95% CI]	5.2 [5.1 – 5.2]	5.1 [5.0 – 5.2]	5.6 [5.3 – 6.0]
	E$_{max}$ [95% CI]	108 [104 – 113]	115 [106 – 121]	114 [86 –141]
'12565	EC$_{50}$ [95% CI] (nM)	104 [79 – 138]	44.8 [27 – 73]	54.6 [22.1 – 135]
	pEC$_{50}$ [95% CI]	7.0 [6.9 – 7.1]	7.4 [7.1 – 7.6]	7.3 [6.9 – 7.7]
	E$_{max}$ [95% CI]	28 [27 – 28]	25 [24 –26]	51 [41 – 61]
'10010	EC$_{50}$ [95% CI] (nM)	814 [717 – 932]	801 [582– 1102]	1396 [506 – 3853]
	pEC$_{50}$ [95% CI]	6.1 [6.0 – 6.1]	6.1 [6.0 – 6.2]	5.9 [5.4 – 6.3]
	E$_{max}$ [95% CI]	74 [70 – 74]	71 [67 – 75]	144 [113 –175]
'6439	EC$_{50}$ [95% CI] (nM)	65.5 [56.8 – 755]	60.9 [52.9 – 70]	339 [96 – 1195]
	pEC$_{50}$ [95% CI]	7.2 [7.1 – 7.2]	7.2 [7.2 – 7.3]	6.5 [5.9 – 7.0]
	E$_{max}$ [95% CI]	95 [93 – 96]	100 [99 – 101]	162 [125 –199]
'6448	EC$_{50}$ [95% CI] (nM)	345 [321 – 372]	310 [267 – 360]	728 [381 – 1393]
	pEC$_{50}$ [95% CI]	6.5 [6.4 – 6.5]	6.5 [6.4 – 6.6]	6.1 [5.9 – 6.4]
	E$_{max}$ [95% CI]	87 [86 – 88]	90 [87 – 91]	151 [130 –172]
'3737a	EC$_{50}$ [95% CI] (nM)	2804 [2436 – 3285]	6729 [2678 – 16910]	2470 [998 – 6108]
	pEC$_{50}$ [95% CI]	5.6 [5.5 – 5.6]	5.2 [4.8 – 5.6]	5.6 [5.2 – 6.1]
	E$_{max}$ [95% CI]	98 [94 – 102]	118 [94 – 143]	140 [9109 – 172]
'3737b	EC$_{50}$ [95% CI] (nM)	3026 [2486 – 3822]	19780 [4343 – 90080]	30840 [1133 – 839000]
	pEC$_{50}$ [95% CI]	5.5 [5.4 – 5.6]	4.7 [4.1 – 5.4]	4.5 [3.1– 6.0]
	E$_{max}$ [95% CI]	89 [84 – 94]	152 [94 – 208]	231 [34 – 428]
'3737c	EC$_{50}$ [95% CI] (nM)			
	pEC$_{50}$ [95% CI]	N.D.	N.D.	N.D.
	E$_{max}$ [95% CI]			
'3737d	EC$_{50}$ [95% CI] (nM)			
	pEC$_{50}$ [95% CI]	N.D.	N.D.	N.D.
	E$_{max}$ [95% CI]			

Compound		hCB1 G_{i1}	hCB1 G_{oB}	hCB1 G_z
'7019	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	N.D.	N.D.	--
7019_31	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	N.D.	N,D.	--
'7800	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	6793 [983] 5.2 [6.0] 34 [24]	597 [398 – 1017] 6.2 [5.9 – 6.4] 21 [20 – 23]	--
7800_29	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	4941 [1763 – 26860000] 5.3 [3.4 – 5.8] 31 [24 – 88]	7472 [1125] 5.1 [6.0] 32 [22]	--
'0450	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	56310 [2177] 4.3 [5.7] 49 [26]	2509 [680 – 11020000] 5.6 [3.0 – 6.2] 36 [30 – 117]	--
'2153	EC_{50} [95% CI] (nM) pEC_{50} [95% CI] E_{max} [95% CI]	1011 [371 – 6576] 6.0 [5.2 – 6.4] 89 [78 – 118]	2061 [620 – 25390] 5.7 4.6 – 6.2] 109 [91 – 162]	--

N.D. = best fit values were not able to be determined due to compound inactivity or poor data quality
-- Not tested

262

Table 5.S5. Detailed functional activities for select analogs and controls versus a variety of transducers and hCB1 in the bioSens-All® platform.

Compound		hCB1 G$_{i1}$	hCB1 G$_{i2}$	hCB1 G$_{oB}$	hCB1 G$_z$	hCB1 G$_{13}$	hCB1 G$_{15}$	hCB1 Barr2 + GRK2
CP–55,940	EC$_{50}$ [95% CI] (nM)	0.46 [0.4 – 0.5]	0.55 [0.4 – 0.7]	0.63 [0.6 – 0.7]	0.28 [0.18 – 0.5]	2.4 [1.65 – 3.4]	0.26 [0.22 – 2.9]	3.1 [1.97 – 4.8]
	pEC$_{50}$ [95% CI]	9.34 [9.3 – 9.4]	9.26 [9.1 – 9.4]	9.20 [9.2 – 9.3]	9.55 [9.4 – 9.8]	8.62 [8.5 – 8.8]	9.59 [9.5 – 9.7]	8.51 [8.3 – 8.7]
	E$_{max}$ [95% CI]	100 [98 – 102]	100 [96 – 103]	100 [98 – 101]	101 [96 – 107]	100 [95 – 105]	100 [98 – 102]	89 [82 – 98]
‘4042	EC$_{50}$ [95% CI] (nM)	0.48 [0.4 – 0.6]	0.56 [0.4 – 0.9]	0.64 [0.5 – 0.8]	0.43 [0.3 – 0.6]	2.1 [0.5 – 9.6]	0.37 [0.28 – 0.49]	3.6 [2.1 – 6.8]
	pEC$_{50}$ [95% CI]	9.32 [9.2 – 9.4]	9.25 [9.0 – 9.5]	9.20 [9.1 – 9.3]	9.37 [9.3 – 9.4]	8.69 [8.0 – 9.3]	9.43 [9.3 – 9.6]	8.44 [8.2 – 8.7]
	E$_{max}$ [95% CI]	102 [100 – 104]	101 [95 – 106]	103 [99 – 106]	97 [93 – 102]	64 [54 – 84]	105 [102 – 108]	72 [71 – 74]
‘1350	EC$_{50}$ [95% CI] (nM)	0.23 [0.18 – 0.3]	0.28 [0.19 – 0.4]	0.29 [0.24 – 0.34]	0.35 [0.22 – 0.53]	0.83 [0.23 – 2.7]	0.22 [0.18 – 0.27]	2.2 [1.1 – 4.4]
	pEC$_{50}$ [95% CI]	9.63 [9.5 – 9.7]	9.54 [9.4 – 9.7]	9.54 [9.5 – 9.6]	9.46 [9.3 – 9.7]	9.08 [8.6 – 9.6]	9.66 [9.6 – 9.8]	8.66 [8.4 – 9.0]
	E$_{max}$ [95% CI]	92 [90 – 95]	95 [90 – 99]	98 [94 – 98]	94 [88 – 100]	56 [48 – 67]	99 [96 – 101]	63 [60 – 72]
‘8690	EC$_{50}$ [95% CI] (nM)	19.1 [14.4 – 25.8]	--	18 [14 – 23]	33 [15 – 117]	--	--	N.D.
	pEC$_{50}$ [95% CI]	7.72 [7.6 – 7.8]		7.8 [7.7 – 7.9]	7.5 [6.9 – 7.8]			
	E$_{max}$ [95% CI]	82 [77 – 85]		83 [80 – 87]	98 [86 – 121]			
2-AG	EC$_{50}$ [95% CI] (nM)	224 [147 – 353]	62 [42 – 918]	93 [65 – 135]	493 [15]	4470 [1618 – 398200]	32 [25 – 40]	1025 [774 – 1446]
	pEC$_{50}$ [95% CI]	6.65 [6.5 – 6.8]	7.2 [7.0 – 7.4]	7.0 [6.9 – 7.2]	6.3 [7.8]	5.4 [5.1 – 5.8]	7.5 [7.4 – 7.6]	6.0 [5.8 – 6.1]
	E$_{max}$ [95% CI]	122 [113 – 133]	112 [106 – 119]	114 [108 – 121]	205 [112]	183 [133 – 607]	112 [108 – 116]	200 [184 – 220]

N.D. = best fit values were not able to be determined due to compound inactivity or poor data quality

-- Not tested

263

Table 5.S6. Relative efficacy calculations for '4042 and '1350 versus CP-55,940.
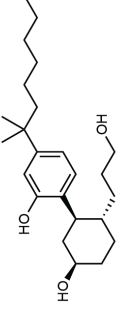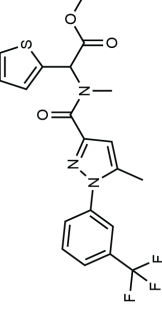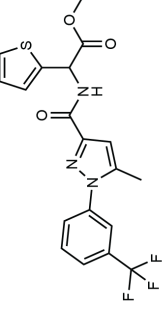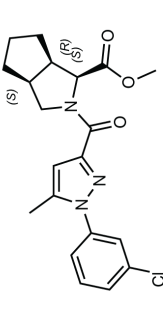
Target	Sensor	Compound	Mean log (E_{max}/EC_{50})	SEM log (E_{max}/EC_{50})	Mean Δlog (E_{max}/EC_{50})	SEM Δlog (E_{max}/EC_{50})	t-test to CP-55,940[1]	RE
hCB1	G_{i1}	CP-55,940	9.34	0.07	0.00	0.09		1.00
		'4042	9.33	0.14	-0.01	0.15	ns	0.97
		'1350	9.59	0.06	0.30	0.09	****	1.98
	G_{i2}	CP-55,940	9.24	0.13	0.00	0.18		1.00
		'4042	9.25	0.03	0.01	0.13	ns	1.01
		'1350	9.52	0.03	0.28	0.13	**	1.89
	G_{oB}	CP-55,940	9.19	0.07	0.00	0.10		1.00
		'4042	9.21	0.09	0.02	0.11	ns	1.04
		'1350	9.09	0.44	0.35	0.44	****	2.24
	G_z	CP-55,940	9.53	0.23	0.00	0.33		1.00
		'4042	9.32	0.19	-0.21	0.30	ns	0.62
		'1350	9.42	0.22	-0.11	0.32	ns	0.77
	G_{13}	CP-55,940	8.63	0.23	0.00	0.32		1.00
		'4042	8.59	0.25	-0.04	0.33	ns	0.91
		'1350	8.85	0.05	0.22	0.23	*	1.64
	G_{15}	CP-55,940	9.59	0.02	0.00	0.02		1.00
		'4042	9.46	0.12	-0.13	0.12	ns	0.74
		'1350	9.65	0.06	0.06	0.06	ns	1.14
	$ß_{arr2}$+ GRK2	CP-55,940	8.28	0.25	0.00	0.35		1.00
		'4042	8.19	0.03	-0.09	0.03	*	0.81
		'1350	8.34	0.08	0.06	0.08	ns	1.16
hCB2	G_{i1}	CP-55,940	8.86	0.12	0.00	0.17		1.00
		'4042	8.49	0.03	-0.20	0.13	ns	0.63
		'1350	8.32	0.10	-0.54	0.16	*	0.29
	G_{i2}	CP-55,940	8.97	0.00	0.00	0.00		1.00
		'4042	8.74	0.00	-0.23	0.00	--	0.59
		'1350	8.57	0.00	-0.40	0.00	--	0.40
	G_{oB}	CP-55,940	8.76	0.09	0.00	0.13		1.00
		'4042	8.43	0.19	-0.33	0.21	n.s.	0.47
		'1350	8.36	0.01	-0.40	0.09	***	0.40
	G_z	CP-55,940	8.90	0.39	0.00	0.55		1.00
		'4042	8.48	0.51	-0.42	0.64	n.s.	0.38
		'1350	8.24	0.23	-0.65	0.45	n.s.	0.22
	$ß_{arr2}$+ GRK2	CP-55,940	7.83	0.13	0.00	0.18		1.00
		'4042	7.74	0.00	-0.09	0.13	***	0.82
		'1350	7.79	0.09	-0.04	0.16	n.s.	0.92

[1]Statistical significance of all comparisons of compound activities (Mean Δlog (E_{max}/EC_{50})) to CP-55,940 control by unpaired t-test. ns = not significant, * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001

RE, relative efficacy = $10^{\Delta log\ (E_{max}/EC50)}$

--, not determined

Table 5.S7. Binding affinities and functional activities for select active analogs at CB2.

Compound	rCB2 binding EC50 [95% CI] (nM) pEC50 [95% CI] Emax [95% CI]	hCB2 Cerep cAMP EC50 [95% CI] (nM) pEC50 [95% CI] Emax [95% CI]	hCB2 BRET2 + GoA EC50 [95% CI] (nM) pEC50 [95% CI] Emax [95% CI]	hCB2 Tango ß-arrestin recruitment EC50 [95% CI] (nM) pEC50 [95% CI] Emax [95% CI]
CP–55,940	--	0.082	13.1 [9.4 – 18.5] 7.9 [7.7 – 8.0] 99% [94 – 105]	22.9 [21.3 – 24.8] 7.6 [7.60 – 7.67] 100% [98 – 102]
'4042	2.18 [1.7 – 2.8] 8.7 [8.6 – 8.8] 103% [100 – 103]	0.011 [0.002 – 0.02] 10.95 [10.7 – 11.6] 72% [77 – 84]	7.9 [1.6 – 40.6] 8.1 [7.4 – 8.8] 29% [22 – 36]	33.9 [21.7 – 61.5] 7.5 [7.2 – 7.7] 28% [26 – 32]
'60154	--	--	594 [125 – 2813] 6.2 [5.6 – 6.9] 31% [20 – 42]	3341 [2707 – 4371] 5.5 [5.4 – 5.6] 21% [19 – 23]
10010	--	--	N.D.	556 [506 – 611] 6.3 [6.2 – 6.3] 18% [17 – 19]

Compound	rCB2 binding EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB2 Cerep cAMP EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB2 BRET2 + GoA EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]	hCB2 Tango ß-arrestin recruitment EC$_{50}$ [95% CI] (nM) pEC$_{50}$ [95% CI] E$_{max}$ [95% CI]
'6439	--	--	97.1 [12.6 – 749] 7.0 [6.13 – 67.9] 81% [34 – 128]	338 [282 – 415] 6.5 [6.4 – 6.6] 14% [13 – 15]
'6448	--	--	--	--
'3737	--	4.3 [.0004 – 22.6] 8.4 [7.7 – 12.5] 45% [42 – 80%]	N.D.	1500 [1274 – 1805] 5.8 [5.7 – 5.9] 9% [8 – 9]

N.D. = best fit values were not determined due to compound inactivity or poor data quality

-- Not tested

266

Table 5.S8. Functional activities for select analogs and controls versus a variety of transducers and hCB2 in the bioSensAll platform.

Compound		hCB2 G_{i1}	hCB2 G_{i2}	hCB2 G_{oB}	hCB2 G_z	hCB2 Barr2 + GRK2
CP–55,940	EC_{50} [95% CI] (nM)	1.4 [1.1 – 1.7]	1.1 [0.9 – 1.3]	1.7 [1.5 – 2.1]	1.05 [0.6 – 1.9]	14.3 [11.6 – 17.9]
	pEC_{50} [95% CI]	8.87 [8.7 – 8.9]	8.97 [8.9 – 9.1]	8.76 [8.7 – 8.8]	8.98 [8.7 – 9.2]	7.84 [7.8 – 7.9]
	E_{max} [95% CI]	100 [97 – 103]	100 [97 – 104]	100 [98 – 103]	98 [89 – 109]	100 [95 – 104]
'4042	EC_{50} [95% CI] (nM)	2.7 [2.0 – 3.4]	1.3 [0.9 – 1.7]	2.5 [1.9 – 3.3]	1.36 [0.5 – 3.1]	5.5 [3.5 – 8.7]
	pEC_{50} [95% CI]	8.58 [8.5 – 8.7]	8.90 [8.8 – 9.0]	8.60 [8.5 – 8.7]	8.87 [8.5 – 9.3]	8.26 [8.1 – 8.5]
	E_{max} [95% CI]	82 [79 – 85]	70 [67 – 73]	61 [59 – 63]	56 [47 – 68]	33 [33 –33]
'1350	EC_{50} [95% CI] (nM)	3.55 [2.98 – 4.2]	1.6 [1.4 – 1.9]	2.6 [2.15 – 3.1]	2.6 [1.6 – 3.9]	4.2 [2.8 – 6.3]
	pEC_{50} [95% CI]	8.45 [8.4 – 8.5]	8.79 [8.7 – 8.9]	8.58 [8.5 – 8.7]	8.58 [8.4 – 8.8]	8.38 [8.2 – 8.6]
	E_{max} [95% CI]	74 [72 – 77]	61 [59 –63]	59 [57 – 61]	52 [46 – 60]	30 [28 –33]
2-AG	EC_{50} [95% CI] (nM)	217 [186 – 256]	96 [70 – 129]	394 [349 – 447]	2123 [931 – 87070]	1854 [1110–4061]
	pEC_{50} [95% CI]	6.66 [6.6 – 6.7]	7.0 [6.9 – 7.2]	6.4 [6.3 – 6.5]	5.7 [4.1 – 6.0]	5.7 [5.4 – 6.0]
	E_{max} [95% CI]	105 [102 – 109]	78 [74 – 84]	101 [98 – 104]	96 [76 – 233]	118 [104 – 140]

-- Not tested

Table 5.S9. Fraction unbound levels of CP-55,940 and '4042 in mouse brain tissue.

Compound ID	Compound concentration [μM]	Fraction unbound in brain ($f_{u,brain}$)	% Recovery	% Stability
Caffeine	2	0.6525	100	101
Midazolam	2	0.0442	108	102
CP-55,940	2	0.0008	105	104
'4042	2	0.0077	111	96

5.8 Materials and Methods

Molecular docking. A crystal structure of the active-state CB1 receptor (PDB: 5XR8)[16] was used for docking calculations. As the goal was to find small-molecule, non-phytocannabinoid ligands, we used ligand coordinates from the cryogenic ligand MDMB-Fubinaca (PDB: 6N4B)[18], after overlaying the two receptor structures. The coordinates of Met363[6.55] were modified slightly, while maintaining the residue within the electron density to reduce a clash with the overlaid ligand indole group. The combined coordinates were minimized with Schrödinger's Maestro prior to calculation of the docking energy potential grids. These grids were precalculated using CHEMGRID[52] for AMBER[53] van der Waals potential, QNIFFT[54] for Poisson-Boltzmann-based electrostatic potentials, and SOLVMAP[55] for context-dependent ligand desolvation. Atoms of the ligand determined in the cryo-EM structure (PDB: 6N4B), MDMB-Fubinaca, were used to seed the matching sphere calculation in the orthosteric site, with 45 total spheres used (these spheres act as pseudo-atoms defining favorable sub-sites on to which library molecules may be superposed[56]. The receptor structure was protonated using REDUCE[57] and AMBER united atom charges were assigned[53]. Control calculations[58] using 324 known ligands extracted from the IUPHAR database[59], CHEMBL24[32], and ZINC15, and 14,929 property-matched decoys[60] were used to optimize docking parameters based on enrichment measured by logAUC[58], prioritization of neutral over charged molecules, and by the reproduction of expected and known binding modes of CB$_1$ ligands. SPHGEN[56] was used to generate pseudo-atoms to define the extended low protein dielectric and desolvation region[22,61]. The protein low dielectric and desolvation regions were extended as previously described[62], based on control calculations, by a radius of 1.5 Å and 1.9 Å,

respectively. The desolvation volume was removed around S383$^{7.39}$ and H178$^{2.65}$ to decrease the desolvation penalty near these residues and to increase the number of molecules that would form polar contacts with them.

A subset of 74 million large, relatively hydrophobic molecules from the ZINC15 database (http://zinc15.docking.org), with calculated octanol-water partition coefficients (cLogP, calculated using JChem-15.11.23.0, ChemAxon; https://www.chemaxon.com) between 3 ≤ 5 and with molecular mass from 350 Da to ≥ 500 Da, was docked against the CB$_1$ orthosteric site using DOCK3.7[63]. Of these, more than 18 million successfully fit. An average of 4,706 orientations, and for each orientation, an average of 645 conformations was sampled. Overall, about 64 trillion complexes were sampled and scored. The total time was about 25,432 core hours, or less than 18 wall-clock hours on 1,500 cores.

To reduce redundancy of the top scoring docked molecules, the top 300,000 ranked molecules were clustered by ECFP4-based Tanimoto coefficient (Tc) of 0.5, and the best scoring member was chosen as the cluster representative molecule. These 60,420 clusters were filtered for novelty by calculating the Tc against >7,000 CB1 and CB2 receptor ligands from the CHEMBL24[32] database. Molecules with Tc ≥ 0.38 to known CB1R/CB2R ligands were not pursued further.

After filtering for novelty, the docked poses of the best-scoring members of each cluster were filtered by the proximity of their polar moieties to Ser383$^{7.39}$, Thr201$^{3.37}$, or

His178$^{2.65}$, and visually inspected for favorable geometry and interactions. For the most favorable molecules, all members of its cluster were also inspected, and one of these was chosen to replace the cluster representative if they exhibited more favorable poses or chemical properties. Ultimately, 60 compounds were chosen for synthesis and testing.

Make-on-demand synthesis and purity information. Of these 60, 52 were successfully synthesized by Enamine (an 87% fulfilment), but only 46 were ultimately screened due to poor DMSO solubility of six of the ordered ligands. The purities of active molecules and analogs synthesized by Enamine were at least 90% and typically above 95%. For bespoke compound synthesized in house purities were at least 95% and typically above 98%.

Ligand optimization. Analogs with ECFP4 Tcs ≥ 0.5 to the four most potent docking hits ('**51486**, '**0450**, '**7800**, and '**7019**) were queried in Arthor and SmallWorld (https://sw.docking.org, https://arthor.docking.org; NextMove Software, Cambridge UK) against 1.4 and 12 Billion tangible libraries, respectively, the latter primarily containing Enamine REAL Space compounds (https://enamine.net/compound-collections/real-compounds/real-space-navigator). Results were pooled, docked into the CB1R site, and filtered using the same criteria as the original screen. Between 11 and 30 analogs were synthesized for each of the four scaffolds. Second- and third-round analogs were designed in 2D space based on specific hypotheses and were synthesized at Enamine or at Northeastern University.

Radioligand Binding Experiments. The binding affinities of the compounds were obtained by competition binding using membrane preparations from rat brain (source of CB1) or HEK293 cells stably expressing human CB2 receptors and $[^3H]$-CP-55,940 as the radioligand, as described[64]. The results were analyzed using nonlinear regression to determine the IC_{50} and K_i values for each ligand (Prism by GraphPad Software, Inc., San Diego, CA). The K_i values are expressed as the mean of two to three experiments each performed in triplicate.

Functional assays

Lance Ultra cAMP Accumulation Assay. The inhibition of forskolin-stimulated cAMP accumulation assays was carried out using PerkinElmer's Lance Ultra cAMP kit following the manufacturer's protocol. In brief, CHO cells stably expressing human CB1 were harvested by incubation with Versene (ThermoFisher Scientific, Waltham, MA) for 10 min, washed once with Hank's Balanced Salt Solution, and resuspended in stimulation buffer at ~200 cells/µL density. The ligands at eight different concentrations (0.001-10,000 nM) in stimulation buffer (5 µL) containing forskolin (2 µM final concentration) were added to a 384-well plate followed by the cell suspension (5 µL; ~1000 cells/well). The plate was incubated for 30 min at room temperature. Eu-cAMP tracer (5 µL) and Ulight-anti-cAMP (5 µL) working solutions were then added to each well, and the plate was incubated at room temperature for an additional 60 min. Results were measured on a Perkin-Elmer EnVision plate reader. The EC_{50} values were determined by nonlinear regression analysis using Prism software (GraphPad Software, Inc., San Diego, CA) and are expressed as the mean of three experiments, each performed in triplicate.

Cerep cAMP Inhibition Assay. Compounds **'4042** and **'3737** were run through the Cerep HTRF cAMP assay for functional activity as agonists (catalog number 1744; Cerep, Eurofins Discovery Services; France). The hCB1 CHO-K1 cells are suspended in HBSS buffer (Invitrogen) complemented with 20 mM HEPES (pH 7.4), then distributed in microplates at a density of 5.103 cells/well in the presence of either of the following: HBSS (basal control), the reference agonist at 30 nM (stimulated control) or the test compounds. Thereafter, the adenylyl cyclase activator forskolin is added at a final concentration of 25 µM. Following 30 min incubation at 37°C, the cells are lysed, and the fluorescence acceptor (D2-labeled cAMP) and fluorescence donor (anti-cAMP antibody labeled with europium cryptate) are added. After 60 min at room temperature, the fluorescence transfer is measured at λex=337 nm and λem=620 and 665 nm using a microplate reader (Envison, Perkin Elmer). The cAMP concentration is determined by dividing the signal measured at 665 nm by that measured at 620 nm (ratio). The results are expressed as a percent of the control response to 10 nM CP-55,940. Each measurement was done in triplicate.

Glosensor cAMP Accumulation Assay. The GloSensor cAMP accumulation assay was performed as secondary validation assays (dose-response setup) as described in detail on the NIMH PDSP website at https://pdsp.unc.edu/pdspweb/content/PDSP%20Protocols%20II%202013-03-28.pdf. The results were analysed using GraphPad Prism 9. Each experiment was performed in triplicate and functional IC_{50} values were determined from the mean of three independent experiments**.**

TRUPATH BRET2 G_{oA} recruitment for CB2R. CB2 receptor was co-expressed with. G_{oA} dissociation BRET2 assays were performed as previously described with minor modifications[65]. In brief, HEK293T cells were co-transfected overnight with human CB2 receptor, $G_{\alpha o}$A-Rluc, $G_{\beta 3}$, and $G_{\gamma 9}$-GFP2 constructs. After 18–24 hours, the transfected cells were seeded into poly-L-lysine-coated 384-well white clear-bottom cell culture plates at a density of 15,000–20,000 cells and incubated with DMEM containing 1% dialyzed FBS, 100 U mL−1 of penicillin and 100 µg ml−1 of streptomycin for another 24 hours. The next day, the medium was aspirated and washed once with 20 µL of assay buffer (1× HBSS, 20 mM HEPES, 0.1% BSA, pH 7.4). Then, 20 µL of drug buffer containing coelenterazine 400a (Nanolight Technology) at 5 µM final concentration was added to each well and incubated for 5 minutes, followed by the addition of 10 µL of 3X designated drug buffer for 5 minutes. Then, 10 µL of 4X final concentrations of ligands were added for 5 minutes. Finally, the plates were read in PHERAstar FSX (BMG Labtech) with a 410-nm (RLuc8-coelenterazine 400a) and a 515-nm (GFP2) emission filter, at 0.6-second integration times. BRET ratio was computed as the ratio of the GFP2 emission to RLuc8 emission. Data were normalized to percentage of CP-55,940 and analyzed in GraphPad Prism 9.1. Each experiment was performed in triplicate and functional IC_{50} values were determined from the mean of four independent experiments**.**

Tango β-Arrestin-2 Recruitment Assay. The Tango β-Arrestin-2 recruitment assays were performed as described[66]. In brief, HTLA cells were transiently transfected with human CB1 or CB2 Tango DNA construct overnight in DMEM supplemented with 10 % FBS, 100 µg ml−1 streptomycin and 100 U ml−1 penicillin. The transfected cells were

then plated into poly-L-lysine-coated 384-well white clear-bottom cell culture plates in DMEM containing 1% dialysed FBS at a density of 10,000–15,000 cells per well. After incubation for 6 h, the plates were added with drug solutions prepared in DMEM containing 1% dialysed FBS for overnight incubation. On the day of assay, medium and drug solutions were removed and 20 μl per well of BrightGlo reagent (Promega) was added. The plates were further incubated for 20 min at room temperature and counted using the Wallac TriLux Microbeta counter (PerkinElmer). The results were analysed using GraphPad Prism 9. Each experiment was performed in triplicate and functional IC_{50} values were determined from the mean of three independent experiments.

DiscoverX PathHunter® β-Arrestin-2 Recruitment Assay. '4042 and **'3737** were run through the PathHunter® β-Arrestin-2 assay (catalog number 86-0001P-2070AG; DiscoverX, Eurofins Discovery Services; CA, USA). PathHunter cell lines (CHO-K1 lineage expressing hCB1) were expanded from freezer stocks according to standard procedures. Cells were seeded in a total volume of 20 μL into white walled, 384-well microplates and incubated at 37°C for the appropriate time prior to testing. For agonist determination, cells were incubated with sample to induce response. Intermediate dilution of sample stocks was performed to generate 5X sample in assay buffer. 5 μL of 5X sample was added to cells and incubated at 37°C or room temperature for 90 to 180 minutes. Vehicle concentration was 1%. Assay signal was generated through a single addition of 12.5 or 15 μL (50% v/v) of PathHunter Detection reagent cocktail, followed by a 1-hour incubation at room temperature. Microplates were read following signal generation with a PerkinElmer EnvisionTM instrument for chemiluminescent signal detection. Compound

activity was analyzed using CBIS data analysis suite (ChemInnovation, CA). Percentage activity was calculated using the following equation:

$$\% \, CP-55{,}940 \, activity = 100 \, \times \, \frac{(mean \, RLU_{test \, sample} - mean \, RLU_{vehicle})}{(mean \, max_{CP-55{,}940} - mean \, RLU_{CP-55{,}940})}$$

The data were analyzed in GraphPad Prism 9.1 using "dose–response-stimulation log(agonist) versus response (four parameters)" and data were presented as EC_{50} or $pEC_{50} \pm$ CIs of one independent experiment in duplicate.

Signaling profiling of hCB1 and hCB2 using bioSensAll®. ebBRET-based effector membrane translocation biosensor assays were conducted at Domain Therapeutics NA Inc. (Montreal, QC, Canada) as previously described[39]. CP-55,940, 2-AG and 25 test compounds were assayed for their effect on the signaling signature of the human cannabinoid receptor type 1 or 2 (hCB1 or hCB2) using the following bioSensAll® sensors: the heterotrimeric G protein activation sensors ($G_{\alpha s}$, $G_{\alpha i1}$, $G_{\alpha i2}$, $G_{\alpha oB}$, $G_{\alpha z}$, $G_{\alpha 13}$, $G_{\alpha q}$, $G_{\alpha 15}$) and the ßarrestin-2 plasma membrane (PM) recruitment sensor (in the presence of GRK2 overexpression). HEK293 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Wisent) supplemented with 1% penicillin- streptomycin (Wisent) and 10% (or 2 % for transfection) fetal bovine serum (Wisent) at 37°C with 5% CO2. All biosensor-coding plasmids and related information are the property of Domain Therapeutics NA Inc. The total amount of transfected DNA was adjusted and kept constant at 1 µg per mL of cell culture to be transfected using salmon sperm DNA (Invitrogen) as 'carrier' DNA, PEI (polyethylenimine 25 kDa linear, PolyScience) and DNA

(3:1 ml PEI:mg DNA ratio) were first diluted separately in 150 mM NaCl then mixed and incubated for at least 20 minutes at room temperature to allow for the formation of DNA/PEI complexes. During the incubation, HEK293 cells were detached, counted, and re-suspended in maintenance medium to a 350,000 cells per mL density. At the end of the incubation period, the DNA/PEI mixture was added to the cells. Cells were finally distributed in 96-well plates (White Opaque 96-well /Microplates, Greiner) at a density of 35,000 cells per well. Forty-eight hours post-transfection, medium was aspirated and replaced with 100 µl of Hank's Balanced Salt Solution buffer (HBSS) (Wisent) per well using 450-Select TS Biotek plate washer. After 60 min incubation in this medium, 10 µL of 10 µM e-Coelenterazine Prolume Purple (Methoxy e-CTZ) (Nanolight) was added to each well for a final concentration of 1 µM immediately followed by addition of increasing concentrations of the test compounds to each well using the HP D300 digital dispenser (Tecan). All compounds were assayed at 22 concentrations with each biosensor after a 10-minute room temperature incubation period. BRET readings were collected with a 0.4 sec integration time on a Synergy NEO plate reader (BioTek Instruments, Inc., USA; filters: 400nm/70nm, 515nm/20nm). BRET signals were determined by calculating the ratio of light emitted by GFP-acceptor (515nm) over light emitted by luciferase-donor (400nm). All BRET ratios were standardized using the universal BRET (uBRET) equation:

$$uBRET = \left(\frac{BRET\ ratio - A}{B - A}\right) \times 10{,}000$$

where A is the BRET ratio obtained from transfection of negative control and B is the BRET ratio obtained from transfection of positive control. Data were normalized to the best fit values of CP-55,940 from each individual experiment before being pooled across replicates. If CP-55,940 had no response, data were left unnormalized and *uBRET* was

used for plotting. The data were analyzed using the four-parameter logistic non-linear regression model in GraphPad Prism 9.1 and data were presented as means ± CIs of 1-4 independent experiments.

For relative efficacy calculations for **'1350** and **'4042** versus CP-55940, first E_{max} and EC_{50} values were determined from dose-response curves to calculate the $log(E_{max}/EC_{50})$ value for each pathway and each compound. Then, the difference between the $log(E_{max}/EC_{50})$ values was calculated using the following equation:

$$\Delta log \left(\frac{E_{max}}{EC_{50}} \right) = log \left(\frac{E_{max}}{EC_{50}} \right)_{compound} - log \left(\frac{E_{max}}{EC_{50}} \right)_{CP-55,940}$$

The SEM was calculated for the $log(E_{max}/EC_{50})$ ratios using the following equation:

$$SEM = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation, and n is the number of experiments.

The SEM was calculated for the $\Delta log(E_{max}/EC_{50})$ ratios using the following equation:

$$SEM_{\left[\Delta log \left(\frac{E_{max}}{EC_{50}} \right) \right]} = \sqrt{(SEM_{Compound})^2 + (SEM_{CP-55,940})^2}$$

The compounds' efficacy toward each pathway, relative to CP-55,940, were finally calculated using the following equation:

$$Relative\ Efficacy\ (RE) = 10^{\Delta log \left(\frac{E_{max}}{EC_{50}} \right)}$$

The relative efficacies were used in radar plots to demonstrate the relative compound effectiveness compared to CP-55,940.

Statistical analysis was performed using a two-tailed unpaired t-test on the $\Delta log(E_{max}/EC_{50})$ ratios to make pairwise comparisons between tested compounds and CP-55,940 for a given pathway, where P < 0.05 was considered statistically significant.

Bimane Fluoroscence. A minimal cysteine version of CB1 was generated[67] where all the cysteine residues (except C256 and C264) were mutated to alanine. A cysteine residue was engineered at residue 336 (L6.28) on TM6, which was labeled with monobromobimane (bimane) by incubating 10 µM receptor with 10-molar excess of bimane at room temperature for one hour. Excess label was removed using size exclusion chromatography on a Superdex 200 10/300 Increase column in 20 mM HEPES pH 7.5, 100 mM NaCl and 0.01% MNG/0.001% CHS. Bimane-labeled CB1 at 0.1 mM was incubated with ligands (10 µM) for one hour at room temperature. Fluorescence data was collected at room temperature in a 150 µL cuvette with a *FluorEssence v3.8 software* on a Fluorolog instrument (*Horiba*) in photon-counting mode. Bimane fluorescence was measured by excitation at 370 nm with excitation and emission bandwidth passes of 4 nm. The emission spectra were recorded from 410 to 510 nm with 1 nm increment and 0.1 s integration time.

GTP turnover assay. Analysis of GTP turnover was performed by using a modified protocol of the GTPase-Glo™ assay (Promega) described previously[68]. Ligand-

bound (10 µM ligand incubated for one hour at room temperature) or apo CB1 (1 µM) was mixed with G-protein (1 µM) in 20 mM HEPES, pH 7.5, 50 mM NaCl, 0.01% L-MNG/0.001% CHS, 100 µM TCEP, 10 µM GDP and 10 µM GTP and incubated at room temperature. GTPase-Glo-reagent was added to the sample after incubation for 60 minutes (G_{i1-3}) and 20 minutes for (G_o). Luminescence was measured after the addition of detection reagent and incubation for 10 min at room temperature using a *SpectraMax Paradigm* plate reader.

Colloidal Aggregation Counter-Screens

Dynamic Light Scattering (DLS). Samples were prepared as 8-point half-log dilutions in filtered 50 mM KPi buffer, pH 7.0 with final DMSO concentration at 1% (v/v). Colloidal particle formation was measured using DynaPro Plate Reader II (Wyatt Technologies). All compounds were screened in triplicate.

Enzyme Inhibition Counter-Screening Assays. Enzyme inhibition assays to test for colloidal inhibition were performed at room temperature using CLARIOstar Plate Reader (BMG Labtech). Samples were prepared in 50 mM KPi buffer, pH 7.0 with final DMSO concentration at 1% (v/v). Compounds were incubated with 2 nM AmpC β-lactamase (AmpC) or Malate dehydrogenase (MDH) for 5 minutes. AmpC reactions were initiated by the addition of 50 µM CENTA chromogenic substrate (219475, Calbiochem). The change in absorbance was monitored at 405 nm for CENTA (219475, Calbiochem) or 490 for Nitrocefin (484400, Sigma Aldrich) for 60 sec. MDH reactions were initiated by the addition of 200 µM nicotinamide adenine dinucleotide (NADH) (54839, Sigma Aldrich)

and 200 μM oxaloacetic acid (324427, Sigma Aldrich). The change in absorbance was monitored at 340 nm for 60 sec. Initial rates were divided by the DMSO control rate to determine % enzyme activity. Each compound was screened at 100μM in triplicate for three independent experiments, if enzyme inhibition greater than 30% was observed, 8-point half-log concentrations were performed in triplicate for three independent experiments. Data was analyzed using GraphPad Prism software version 9.1 (San Diego, CA).

Brain Tissue Binding Experiments

The study was performed using brain homogenate by equilibrium dialysis method according to previously published protocol[69]. Mouse brain homogenate was prepared from pooled brains (Balb/c, females, *n* = 25). Caffeine and Midazolam were used as reference compounds. Four volumes of DPBS pH 7.4 were added to the pre-weighted tissues (1:4, w/v), fragmented into small pieces, and homogenized using SPEX SamplePrep 1600 MiniG. Samples were centrifuged at 1500 g for 10 minutes. Supernatants were decanted, and obtained brain homogenate was flash-frozen in liquid nitrogen. Aliquots were stored at -70 °C until use.

Equilibrium Dialysis. The assay was performed in a multiple-use 96-well dialysis unit (HTD96b dialyzer). Each individual well unit consisted of 2 chambers separated by a vertically aligned dialysis membrane of predetermined pore size (MWCO 12-14 kDa). 125 μL aliquot of brain homogenate spiked with 2 μM test compound (final DMSO concentration 1%) was added to one dialysis chamber and the same volume of DPBS

buffer pH 7.4 to the other chamber. HTD96b dialyzer was covered with adhesive sealing film and incubated in a humidified (75% RH) incubator at 37°C, shaking at 250 rpm for 5 hours. To define non-specific loss of the compound during this assay, standard solution was created by mixing an aliquot of brain homogenate with blank buffer without dialysis. To collect recovery samples, aliquots of the standard solution were incubated at 37°C, shaking at 250 rpm for 5 hours. To collect stability samples, two aliquots were immediately diluted with acetonitrile and stored at 4°C until LC-MS/MS analysis. All samples were diluted 4-fold with 90% acetonitrile with internal standard with subsequent proteins sedimentation by centrifuging at 6000 rpm for 5 minutes. Supernatants were analyzed using HPLC system coupled with tandem mass spectrometer.

Fraction unbound and recovery were calculated using following equations, where D is the dilution of the brain samples (D = 5):

$$Diluted\ f_{u,drug}\ =\ \frac{peak\ area\ in\ buffer}{peak\ area\ in\ brain\ homogenate}$$

$$Undiluted\ f_u\ =\ \frac{1/_D}{\left(\left(\frac{1}{f_{u,d}}\right)-1\right)+\frac{1}{D}}$$

$$Recovery = \frac{peak\ area\ in\ brain\ homogenate + peak\ area\ in\ buffer}{peak\ area\ in\ standard\ solution}\ \times\ 100$$

$$Stability = \left(\frac{peak\ area\ in\ recovery\ sample}{peak\ area\ in\ stability\ sample}\right)\ \times\ 100$$

Cryo-EM sample preparation and structure determination

Purification of hCB1. hCB1R was expressed and purified as described previously[18]. An N-terminal FLAG tag and C-terminal histidine tag was added to human full-length CB1. This CB1 construct was expressed in *Spodoptera frugiperda Sf9* insect cells with the baculovirus method (Expression Systems). Insect cell pellets expressing CB1 was solubilized with buffer containing 1% lauryl maltose neopentyl glycol (L-MNG) and 0.1% cholesterol hemisuccinate (CHS) and purified by nickel-chelating Sepharose chromatography. The Ni column eluant was applied to a M1 anti-FLAG immunoaffinity resin. After washing to progressively decreasing concentration of L-MNG, the receptor was eluted in a buffer consisting of 20 mM HEPES pH 7.5, 150 mM NaCl, 0.05% L-MNG, 0.005% CHS, FLAG peptide and 5 mM EDTA. As the final purification step, CB1 was applied to a Superdex 200 10/300 gel filtration column (GE) in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.02% L-MNG, 0.002% CHS. Ligand-free CB1 was concentrated to ~500 μM and stored in -80 °C.

Expression and purification of $G_{i/o}$ heterotrimer. Expression and purification of all heterotrimeric G-protein ($G_{i/o}$) follow similar protocols. Heterotrimeric G_i was expressed and purified as previously described[70]. Wild-type human $G\alpha_{i1}$ subunit virus and wild-type human $\beta_1\gamma_2$ (with histidine tagged β subunit) virus were used to co-infect Insect (*Trichuplusia ni, Hi5*) cells. Cells expressing the heterotrimetric, $G_i\beta_1\gamma_2$ G-protein were lysed in hypotonic buffer and G-protein was extracted in a buffer containing 1% sodium cholate and 0.05% n-dodecyl-β-D-maltoside (DDM, Anatrace). Detergent was exchanged from cholate/DDM to DDM on Ni Sepharose column. The eluant from the Ni column was

dialyzed overnight into 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 0.1% DDM, 1 mM magnesium chloride, 100 µM TCEP and 10 µM GDP together with Human rhinovirus 3C protease (3C protease) to cleave off the His tag in the β subunit. 3C protease was removed by Ni-chelating sepharose and the heterotrimetric G-protein was further purified with MonoQ 10/100 GL column (GE Healthcare). Protein was bound to the column and washed in buffer A (20 mM HEPES, pH 7.5, 50 mM sodium chloride, 1 mM magnesium chloride, 0.05% DDM, 100 µM TCEP, and 10 µM GDP). The protein was eluted with a linear gradient of 0–50% buffer B (buffer A with 1 M NaCl). The collected G protein was dialyzed into 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride, 0.02% DDM, 100 µM TCEP, and 10 µM GDP. Protein was concentrated to about 200 µM and flash frozen until further use.

Purification of scFv16. scFv16 was purified with a hexahistidine-tag in the secreted form from *Trichuplusia ni Hi5* insect cells using the baculoviral method. The supernatant from baculoviral infected cells was pH balanced and quenched with chelating agents and loaded onto Ni resin. After washing with 20 mM HEPES pH 7.5, 500 mM NaCl, and 20 mM imidazole, protein was eluted with 250 mM imidazole. Following dialysis with 3C protease into a buffer consisting of 20 mM HEPES pH 7.5 and 100 mM NaCl, scFv16 was further purified by reloading over Ni a column. The collected flow-through was applied onto a Superdex 200 16/60 column and the peak fraction was collected, concentrated and flash frozen.

CB1-G$_{i1}$ complex formation and purification. CB1 in L-MNG was incubated with excess '**1350** for ~ 1 hour at room temperature. Simultaneously, G$_{i1}$ heterotrimer in DDM was incubated with 1% L-MNG/0.1% CHS at 4 ºC. The '**1350**-bound CB1 was incubated with a 1.25 molar excess of detergent exchanged G$_i$ heterotrimer at room temperature for ~ 3 hour. The complex sample was further incubated with apyrase for 1.5 hour at 4 °C to stabilize a nucleotide-free complex. 2 mM CaCl$_2$ was added to the sample and purified by M1 anti-FLAG affinity chromatography. After washing to remove excess G protein and reduce detergents, the complex was eluted in 20mM HEPES pH 7.5, 100mM NaCl, 0.01% L-MNG/0.001% CHS, 0.0033% GDN/0.00033% CHS, 10 µM '**1350**, 5 mM EDTA, and FLAG peptide. The complex was supplemented with 100 µM TCEP and incubated with 2 molar excess of scFv16 overnight at 4 °C. Size exclusion chromatography (Superdex 200 10/300 Increase) was used to further purify the CB1-G$_i$-scFv16 complex. The complex in 20mM HEPES pH 7.5, 100mM NaCl, 10 µM '**1350**, 0.00075% L-MNG/0.000075% CHS and 0.00025% GDN/0.000025% CHS was concentrated to ~12 mg/mL for electron microscopy studies.

Cryo-EM data acquisition. Grids were prepared by applying 3 µL of purified CB1-G$_i$ complex at 12 mg/ml to glow-discharged holey carbon gold grids (Quantifoil R1.2/1.3, 200 mesh). The grids were blotted using a Vitrobot Mark IV (FEI) with 3 s blotting time and blot force 3 at 100% humidity at room temperature and plunge-frozen in liquid ethane. A total of 8324 movies were recorded on a Titan Krios electron microscope (Thermo Fisher Scientific- FEI) operating at 300 kV at a calibrated magnification of 96,000x corresponding to a pixel size of 0.8521 Å. Micrographs were recorded using a K3 Summit

direct electron camera (Gatan Inc.) with a dose rate of 16.4 electrons/pixel/s. The total exposure time was 2.5 s with an accumulated dose of ~ 56.6 electrons per $Å^2$ and a total of 50 frames per micrograph. Automatic data acquisition was done using *SerialEM*.

Image processing and 3D reconstructions. Micrographs were subjected to beam-induced motion correction using *MotionCor2*[71] implemented in Relion 2.1.0[72]. CTF parameters for each micrograph were determined by *CTFFIND4*[73]. An initial set of 4,967,593 particle projections were extracted using semi-automated procedures and subjected to reference-free two-dimensional and multiple rounds of three-dimensional classification in *Relion 2.1.0*[72] to remove low-resolution and otherwise poor-quality particles. From this step, 750,496 particle projections were selected for further processing in *CryoSPARC*[74]. A final two-dimensional classification step in order to select for the highest-resolution particles resulted in a particle set containing 465,411 particles. These particles were reconstructed to a global nominal resolution of 3.3 Å (**Fig. 5.S5**) at FSC of 0.143 using non-uniform refinement. Local resolution was estimated within *CryoSPARC*[74].

Model building and refinement. The initial template of CB1 was the MDMB-Fubinaca-bound CB1-G_i complex structure (PDB: 6N4B). *Phenix.elbow was used to generate* Agonist coordinates and geometry restrains. Models were docked into the EM density map using *UCSF Chimera*. *Coot* was used for iterative model building and the final model was subjected to global refinement and minimization in real space using *phenix.real_space_refine* in *Phenix*. Model geometry was evaluated using *Molprobity*.

FSC curves were calculated between the resulting model and the half map used for refinement as well as between the resulting model and the other half map for cross-validation (**Fig. 5.S5**). The final refinement parameters are provided in **Supplementary Table 3**. The ligand symmetry accounted RMSD between the docked pose and cryo-EM pose of '**1350** was calculated by the Hungarian algorithm in DOCK6[75].

Off-target activity

GPCRome and Comprehensive Binding Panel. Compound '**4042** was tested at 10 μM for off-target activity against a panel of 320 non-olfactory GPCRs using PRESTO-Tango GPCRome arrestin-recruitment assay, as described[66]. Receptors with at least three-fold increased relative luminescence over corresponding basal activity are potential positive hits, and were tested in dose response follow-up studies. Compound '**4042** was further tested at 1 μM for off-target activity at a panel of 45 common GPCR and non-GPCR drug targets. Receptors with at least 50% displaced radioligand are potential positive hits and were tested in dose response follow-up studies. Screening was performed by the National Institutes of Mental Health Psychoactive Drug Screen Program (PDSP)[76]. Detailed experimental protocols are available on the NIMH PDSP website at https://pdsp.unc.edu/pdspweb/content/PDSP%20Protocols%20II%202013-03-28.pdf.

***In vivo* methods**

Animals and ethical compliance. Animal experiments were approved by the UCSF Institutional Animal Care and Use Committee and were conducted in accordance with the NIH Guide for the Care and Use of Laboratory animals (protocol #AN195657). Adult (8-10 weeks old) male C56BL/6 (strain # 664) and CB2R knockout (strain #5786)

mice were purchased from the Jackson Laboratory. Mice were housed in cages on a standard 12:12 hour light/dark cycle with food and water ad libitum. Sample sizes were modelled on our previous studies and on studies using a similar approach, which were able to detect significant changes[77,78]. The animals were randomly assigned to treatment and control groups. Animals were initially placed into one cage and allowed to freely run for a few minutes. Then each animal was randomly picked up, injected with compound treatment or vehicle, and placed into a separate cylinder before the behavioral test.

In vivo compound preparation. Ligands were sourced from Enamine (**'4042**) or Sigma-Aldrich (CP-55,940, Cat No. C1112; Haloperidol, Cat. No. H1512; AM251, Cat. No. A6226; SR 144528, Cat. No. SML1899) and dissolved 30 min before injections. **'4042** was resuspended in a 20% Kolliphor HS-15 (Sigma-Aldrich, Cat. No. 42966) / 40% saline / 40% water for injections (v/v/v) vehicle for i.p. injections. CP-55,940, SR 144528, and AM251 for i.p. injections and **'4042** for i.t. injections were resuspended in a 5% EtOH /5% Kolliphor-EL (Sigma-Aldrich Cat. No. C5135) / 90% water for injections vehicle. Morphine (provided by the NIH) was resuspended in 100% saline. Haloperidol was resuspended in 20% cyclodextrin (Sigma-Aldrich, Cat. No. H107). All cannabinoid formulations were prepared in silanized glass vials.

Pharmacokinetics. Pharmacokinetic experiments were performed by Bienta (Enamine Biology Services) in accordance with Enamine pharmacokinetic study protocols and Institutional Animal Care and Use Guidelines (protocol number 1-2/2020). Plasma, brain, and CSF concentrations were measured for **'4042** and CP-55,940

288

following a 0.2 mg/kg intraperitoneal (i.p.) dose. The batches of working formulations were prepared 5-10 minutes prior to the *in vivo* study. In each compound study, up to nine time points (5, 15, 30, 60, 120, 240, 360, 480 and 1440 min) were collected; each of the time point treatment groups included 3 male CD-1 mice. There was also a one mouse control group. All animals were fasted for 4 h before dosing. Mice were injected i.p. with 2,2,2-tribromoethanol at the dose of 150 mg/kg prior to drawing CSF and blood. Blood collection was performed from the orbital sinus in microtainers containing K_3EDTA. CSF was collected under a stereomicroscope from cisterna magna using 1 mL syringes. Animals were sacrificed by cervical dislocation after the blood samples collection. After this, right lobe brain samples were collected and weighted. All samples were immediately processed, flash-frozen and stored at -70°C until subsequent analysis.

Plasma samples (40 µL) were mixed with 200 µL of internal standard solution. After mixing by pipetting and centrifuging for 4 min at 6,000 rpm, supernatant was injected into LC-MS/MS system. Solution of Difenoconazole (50 ng/ml in water-methanol mixture 1:9, v/v) was used as the internal standard (IS) for quantification of **'4042** and mefenamic acid (100 ng/mL in water- acetonitrile mixture 1:9, v/v) was used as the IS for the quantification of CP-55,940. Brain samples (weight 59 mg – 201 mg) were homogenized with 5 volumes of IS(80) solution using zirconium oxide beads (115 mg ± 5 mg) in The Bullet Blender® homogenizer for 30 seconds at speed 8. After this, the samples were centrifuged for 4 min at 14,000 rpm, and supernatant was injected into LC-MS/MS system. CSF samples (4 µL) were mixed with 100 µL of IS(80) solution. After mixing by pipetting and centrifuging for 4 min at 6,000 rpm, 1-6 µL of each supernatant was injected into LC-MS/MS system.

Analyses of plasma, brain and CSF samples were conducted at Enamine/Bienta. The concentrations of compounds in samples were determined using high performance liquid chromatography/tandem mass spectrometry (HPLC-MS/MS) method. Data acquisition and system control was performed using Analyst 1.6.3 software (AB Sciex, Canada). The concentrations of the test compound below the lower limit of quantitation (LLOQ: 2-5 ng/mL for plasma and CSF, 1-5 ng/g for brain) were designated as zero. The pharmacokinetic data analysis was performed using noncompartmental, bolus injection or extravascular input analysis models in WinNonlin 5.2 (PharSight). Data below LLOQ were presented as missing to improve validity of T½ calculations.

Behavioral analyses. For all behavioral tests, the experimenter was always blind to treatment. Animals were first habituated for 30-60 minutes in Plexiglas cylinders and then tested 30 minutes after i.p. or i.t. injection of the compounds. The mechanical (von Frey), thermal (Hargreaves, and tail flick) and ambulatory (rotarod) tests were conducted as described[79]. Hindpaw mechanical thresholds were determined with von Frey filaments using the up-down method[80]. Hindpaw thermal sensitivity was measured with a radiant heat source (Hargreaves). For the tail flick assay, sensitivity was measured by immersing the tail into a 50°C water bath. For the ambulatory (rotarod) test, mice were first trained on an accelerating rotating rod, three times for 5 min, before testing with any compound. Therapeutic index was calculated as the ratio of the minimum dose of side effect phenotype and the minimum dose of analgesic phenotype.

SNI model of neuropathic pain. Under isoflurane anesthesia, two of the three branches of the sciatic nerve were ligated and transected distally[81], leaving the sural nerve intact. Behavior was tested 7 to 14 days after injury.

CFA. The CFA model of chronic inflammation was induced as described previously[82]. Briefly, CFA (Sigma) was diluted 1:1 with saline and vortexed for 30 min. When fully suspended, we injected 20 µL of CFA into one hindpaw. Heat thresholds were measured before the injection (baseline) and 3 days after the injection using the Hargreaves test.

Open Field Test. Thirty minutes after IP injection, mice were placed in the center of a round open-field (2 feet diameter) and their exploratory behavior recorded over the next 15 minutes. Distance traveled was used to represent open field behavior.

Conditioned Place Preference. To determine if **'4042** was inherently rewarding or aversive we used the conditioned place paradigm as described[83]. Briefly, mice were first habituated to the test apparatus, twice, and their preference for each chamber recorded for 30 minutes (Pretest). Two conditioning days followed in which mice received the vehicle control or the compound, and 30 minutes later restricted for 30 minutes in the preferred or non-preferred chamber, respectively. On day 5 (Test day), mice were allowed to roam freely between the 3 chambers of the apparatus and their preference for each chamber recorded for 30 minutes. To calculate the CPP score, we subtracted the time

spent in each chamber of the box on the Pretest day from that of the Test day (CPP score = Test - Pretest).

Acetone Test. Mice were placed on a wire mesh and thirty min after an IP injection of the compounds we applied a drop (50 μL) of acetone on the ventral aspect of the hindpaw, 5 times every 30 sec. We recorded the number of nocifensive behaviors (paw lifts/licks/shakes/bites) over the 5 applications.

Formalin Test. Thirty minutes after an IP injection of the compounds, mice received an intraplantar injection of a 20μl solution containing 2% formalin (Acros Organics) and we recorded the time mice spent licking/biting/guarding (nocifensive behaviors) the injected hindpaw over the next 60 min.

Catalepsy Test. Thirty and 60 minutes after an IP injection of the compounds, mice were placed on a vertical wire mesh and the latency to move all four paws was recorded.

Body temperature measurements. Body temperature (BT) was measured using a telemetric probe device (HD-X10; Data Science International). Briefly, under anesthesia, the probe device was placed in the mouse abdomen and a subcutaneous tunnel was created from the neck to the abdominal skin, through which a catheter (connected to the probe) was pulled and then inserted into the left carotid artery. Three weeks later, the implanted mice were singly housed in a cage that was placed on top of

the DSI receiver (for probe signal detection). We monitored the BT continuously over 2h, in the following manner: 30 minutes (for baseline), 30 minutes after injection of the vehicle and then for 1h after injection of the compound. Data was acquired using the Ponemah Telemetry acquisition software (DSI) and percent changes were presented relative to each mouse's baseline.

Statistical analyses. All statistical tests were run with GraphPad Prism 9.0 (GraphPad Software Inc., San Diego). A two-tailed unpaired *t*-test was used to compare the pKi ± 95% CI for **'4042** at CB1 versus CB2 (**Fig. 5.S8** legend). Experiments of the compounds in the in vivo assays were analyzed by unpaired two-tailed t-tests, one-way ANOVA, or two-way ANOVA, depending on the experimental design. All statistical calculations were controlled for multiple hypothesis testing using a post-hoc test as described in the **Fig. 5.5**, **Fig. 5.6**, or **Fig. 5.S11** legends. Details of the analyses, including groups compared in post-hoc sets, number of animals per group, *t* or *F* statistics, and *P* values, can be found in the figure legends.

Data availability. The structure described in this manuscript were deposited to the Protein Data Bank under accession code 8GAG, and the map coordinates to EMDB under accession code EMD-29898. Additional data provided in the main text or supplemental materials. Additional requests can be made to bshoichet@gmail.com.

Code availability. DOCK3.7 is freely available for non-commercial research in both executable and code form (http://dock.compbio.ucsf.edu/DOCK3.7/). A web-based

version is freely available to all (http://blaster.docking.org/). The ultra-large library used here is freely available (http://zinc15.docking.org, http://zinc20.docking.org).

5.9 References

1. Lipman, A. G. Medical cannabis for pain: anecdote or evidence. *J. Pain Palliat. Care Pharmacother.* **31**, 96–97 (2017).

2. Iliopoulos-Tsoutsouvas, C., Georgiadis, M.-O., Ji, L., Nikas, S. P. & Makriyannis, A. Natural Compounds and Synthetic Drugs to Target Type-1 Cannabinoid (CB1) Receptor. in *New tools to interrogate endocannabinoid signalling: from natural compounds to synthetic drugs* (ed. Maccarrone, M.) 48–88 (The Royal Society of Chemistry, 2020). doi:10.1039/9781839160752-00048.

3. Sachs, J., McGlade, E. & Yurgelun-Todd, D. Safety and toxicology of cannabinoids. *Neurotherapeutics* **12**, 735–746 (2015).

4. Cohen, K., Weizman, A. & Weinstein, A. Positive and negative effects of cannabis and cannabinoids on health. *Clin. Pharmacol. Ther.* **105**, 1139–1147 (2019).

5. Donvito, G. *et al.* The Endogenous Cannabinoid System: A Budding Source of Targets for Treating Inflammatory and Neuropathic Pain. *Neuropsychopharmacol.* **43**, 52–79 (2018).

6. Papagianni, E. P. & Stevenson, C. W. Cannabinoid Regulation of Fear and Anxiety: an Update. *Curr. Psychiat. Rep.* **21**, 38 (2019).

7. Rock, E. M. & Parker, L. A. Cannabinoids As Potential Treatment for Chemotherapy-Induced Nausea and Vomiting. *Front. Pharmacol.* **7**, 221 (2016).

8. Rossi, F., Punzo, F., Umano, G. R., Argenziano, M. & Giudice, E. M. D. Role of Cannabinoids in Obesity. *Int. J. Mol. Sci.* **19**, 2690 (2018).

9. Perucca, E. Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last? *J. Epilepsy Res.* **7**, 61–76 (2017).

10. Banister, S. D., Kumar, K. K., Kumar, V., Kobilka, B. K. & Malhotra, S. V. Selective modulation of the cannabinoid type 1 (CB1) receptor as an emerging platform for the treatment of neuropathic pain. *MedChemComm* **10**, 647–659 (2019).

11. Martin, B. R. *et al.* Behavioral, biochemical, and molecular modeling evaluations of cannabinoid analogs. *Pharmacol. Biochem. Be.* **40**, 471–478 (1991).

12. Fisher, E. *et al.* Cannabinoids, cannabis, and cannabis-based medicine for pain management: a systematic review of randomised controlled trials. *Pain* **4**, S45–S66 (2020).

13. Finn, D. P. *et al.* Cannabinoids, the endocannabinoid system, and pain: a review of preclinical studies. *Pain* **162**, S5–S25 (2021).

14. Pertwee, R. G. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ9-tetrahydrocannabinol, cannabidiol and Δ9-tetrahydrocannabivarin. *Brit. J. Pharmacol.* **153**, 199–215 (2008).

15. Hua, T. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **167**, 750-762.e14 (2016).

16. Hua, T. *et al.* Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **547**, 468–471 (2017).

17. Shao, Z. *et al.* High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* **540**, 602–606 (2016).

18. Kumar, K. K. *et al.* Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell* (2019) doi:10.1016/j.cell.2018.11.040.

19. Hua, T. *et al.* Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures. *Cell* **180**, 655-665.e18 (2020).

20. Xing, C. *et al.* Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex. *Cell* **180**, 645-654.e13 (2020).

21. Li, X. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* **176**, 459-467.e13 (2019).

22. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

23. Alon, A. *et al.* Structures of the σ2 receptor enable docking for bioactive ligand discovery. *Nature* **600**, 759–764 (2021).

24. Sadybekov, A. A. *et al.* Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022).

25. Stein, R. M. *et al.* Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 1–8 (2020).

26. Fink, E. A. *et al.* Structure-based discovery of nonopioid analgesics acting through the α2A-adrenergic receptor. *Science* **377**, eabn7065 (2022).

27. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185 (2016).

28. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).

29. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliver. Rev.* **23**, 3–25 (1997).

30. Sterling, T. & Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model* **55**, 2324–2337 (2015).

31. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).

32. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 (2014).

33. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).

34. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

35. Kapur, A. *et al.* Mutation Studies of Ser7.39 and Ser2.60 in the Human CB1 Cannabinoid Receptor: Evidence for a Serine-Induced Bend in CB1 Transmembrane Helix 7. *Mol. Pharmacol.* **71**, 1512–1524 (2007).

36. McAllister, S. D. *et al.* Structural Mimicry in Class A G Protein-coupled Receptor Rotamer Toggle Switches: The Importance of the F3.36(201)/W6.48(357) Interaction in Cannabinoid CB1 Receptor Activation. *J. Biol. Chem.* **279**, 48024–48037 (2004).

37. Ramesh, K. & Rosenbaum, D. M. Molecular basis for ligand modulation of the cannabinoid receptor CB1. *Brit. J. Pharmacol.* (2021) doi:10.1111/bph.15627.

38. Kolb, P. *et al.* Community Guidelines for GPCR Ligand Bias: IUPHAR Review 32. *Brit. J. Pharmacol.* (2022) doi:10.1111/bph.15811.

39. Avet, C. *et al.* Effector membrane translocation biosensors reveal G protein and βarrestin coupling profiles of 100 therapeutically relevant GPCRs. *Elife* **11**, e74101 (2022).

40. Manning, J. J., Green, H. M., Glass, M. & Finlay, D. B. Pharmacological selection of cannabinoid receptor effectors: Signalling, allosteric modulation and bias. *Neuropharmacology* **193**, 108611 (2021).

41. Martin, W. J., Loo, C. M. & Basbaum, A. I. Spinal cannabinoids are anti-allodynic in rats with persistent inflammation. *Pain* **82**, 199–205 (1999).

42. Fox, A. *et al.* The role of central and peripheral Cannabinoid1 receptors in the antihyperalgesic activity of cannabinoids in a model of neuropathic pain. *Pain* **92**, 91–100 (2001).

43. Sain, N. M. H., Liang, A., Kane, S. A. & Urban, M. O. Antinociceptive effects of the non-selective cannabinoid receptor agonist CP 55,940 are absent in CB1−/− and not CB2−/− mice in models of acute and persistent pain. *Neuropharmacology* **57**, 235–241 (2009).

44. Tham, S. M., Angus, J. A., Tudor, E. M. & Wright, C. E. Synergistic and additive interactions of the cannabinoid agonist CP55,940 with μ opioid receptor and α2-adrenoceptor agonists in acute pain models in mice. *Brit. J. Pharmacol.* **144**, 875–884 (2005).

45. Grenald, S. A. *et al.* Synergistic attenuation of chronic pain using mu opioid and cannabinoid receptor 2 agonists. *Neuropharmacology* **116**, 59–70 (2017).

46. Vlachou, S. & Panagis, G. Regulation of Brain Reward by the Endocannabinoid System: A Critical Review of Behavioral Studies in Animals. *Curr. Pharm. Design* **20**, 2072–2088 (2014).

47. Schmid, C. L. *et al.* Bias Factor and Therapeutic Window Correlate to Predict Safer Opioid Analgesics. *Cell* **171**, 1165-1175.e13 (2017).

48. Wang, X. *et al.* Different receptor mechanisms underlying phytocannabinoid- versus synthetic cannabinoid-induced tetrad effects: Opposite roles of CB1/CB2 versus GPR55 receptors. *Brit. J. Pharmacol.* **177**, 1865–1880 (2020).

49. Beckers, M., Fechner, N. & Stiefl, N. 25 Years of Small-Molecule Optimization at Novartis: A Retrospective Analysis of Chemical Series Evolution. *J. Chem. Inf. Model* **62**, 6002–6021 (2022).

50. Gatch, M. B. & Forster, M. J. Cannabinoid-like effects of five novel carboxamide synthetic cannabinoids. *Neurotoxicology* **70**, 72–79 (2019).

51. Marusich, J. A., Gamage, T. F., Zhang, Y., Akinfiresoye, L. R. & Wiley, J. L. In vitro and in vivo pharmacology of nine novel synthetic cannabinoid receptor agonists. *Pharmacol. Biochem. Be.* **220**, 173467 (2022).

52. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **13**, 505–524 (1992).

53. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).

54. Gallagher, K. & Sharp, K. Electrostatic Contributions to Heat Capacity Changes of DNA-Ligand Binding. *Biophys. J.* **75**, 769–776 (1998).

55. Mysinger, M. M. & Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model* **50**, 1561–1573 (2010).

56. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).

57. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton. *J. Mol. Biol.* **285**, 1735–1747 (1999).

58. Bender, B. J. *et al.* A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799–4832 (2021).

59. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res.* **44**, D1054–D1068 (2016).

60. Stein, R. M. *et al.* Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model* **61**, 699–714 (2021).

61. Shoichet, B. K., Leach, A. R. & Kuntz, I. D. Ligand solvation in molecular docking. *Proteins Struct. Funct. Bioinform.* **34**, 4–16 (1999).

62. Mysinger, M. M. *et al.* Structure-based ligand discovery for the protein–protein interface of chemokine receptor CXCR4. *Proc. National Acad. Sci.* **109**, 5517–5522 (2012).

63. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *Plos One* **8**, e75992 (2013).

64. Janero, D. R. *et al.* Molecular-Interaction and Signaling Profiles of AM3677, a Novel Covalent Agonist Selective for the Cannabinoid 1 Receptor. *Acs Chem. Neurosci.* **6**, 1400–1410 (2015).

65. Pryce, K. D. *et al.* A promising chemical series of positive allosteric modulators of the μ-opioid receptor that enhance the antinociceptive efficacy of opioids but not their adverse effects. *Neuropharmacology* **195**, 108673 (2021).

66. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biology* **22**, nsmb.3014 (2015).

67. Fay, J. F. & Farrens, D. L. The Membrane Proximal Region of the Cannabinoid Receptor CB1 N-Terminus Can Allosterically Modulate Ligand Affinity. *Biochem.* **52**, 8286–8294 (2013).

68. Gregorio, G. G. *et al.* Single-molecule analysis of ligand efficacy in β2AR–G-protein activation. *Nature* **547**, 68–73 (2017).

69. Di, L. *et al.* Species Independence in Brain Tissue Binding Using Brain Homogenates. *Drug Metab. Dispos.* **39**, 1270–1277 (2011).

70. Dror, R. O. *et al.* Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* **348**, 1361–1365 (2015).

71. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

72. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. Sect. D* **73**, 496–502 (2017).

73. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

74. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

75. Allen, W. J. *et al.* DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **36**, 1132–1156 (2015).

76. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215 (2012).

77. Scherrer, G. *et al.* Dissociation of the Opioid Receptor Mechanisms that Control Mechanical and Heat Pain. *Cell* **137**, 1148–1159 (2009).

78. Muralidharan, A. *et al.* Identification and characterization of novel candidate compounds targeting 6- and 7-transmembrane μ-opioid receptor isoforms. *Brit. J. Pharmacol.* **178**, 2709–2726 (2021).

79. Tran, M. *et al.* Ablation of spinal cord estrogen receptor α-expressing interneurons reduces chemically induced modalities of pain and itch. *J. Comp. Neurol.* **528**, 1629–1643 (2020).

80. Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M. & Yaksh, T. L. Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Meth.* **53**, 55–63 (1994).

81. Shields, S. D., Eckert, W. A. & Basbaum, A. I. Spared nerve injury model of neuropathic pain in the mouse: a behavioral and anatomic analysis. *J. Pain* **4**, 465–470 (2003).

82. Cao, Y. Q. *et al.* Primary afferent tachykinins are required to experience moderate to intense pain. *Nature* **392**, 32897 (1998).

83. Juarez-Salinas, D. L., Braz, J. M., Hamel, K. A. & Basbaum, A. I. Pain relief by supraspinal gabapentin requires descending noradrenergic inhibitory controls. *Pain Reports* **3**, e659 (2018).

Chapter 6: Conclusions and Future Perspectives

Conclusions and Future Perspectives

Tia A. Tummino[1,2]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco (UCSF), San Francisco, CA, USA.

[2]Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA.

6.1 Conclusions and Future Perspectives.

As the virtual make-on-demand libraries continue to grow into the 10+ billion molecule range[1], virtual screening capabilities will continue to be put to the test. Importantly, the focus will be less on *can we do it*, but instead will focus on how we can advance the technology to screen these much larger libraries in a reasonable amount of time, likely without needing to virtually enumerate the library in its entirely. It is exciting to think about what types of novel ligands with unique pharmacology and *in vivo* effects could be identified using even bigger libraries than those that were available when I started my projects. In fact, there are already exciting advancements in this area— from machine learning algorithms to predict a subsection of chemical space[2] for docking to fragment approaches that screen building blocks rather than individual ligands and enumerate them into larger molecules after the time-limited calculations are done[3–5]. Time will tell which direction the field moves in to meet these challenges, but it will be exciting to watch as more biological systems take advantage of docking and the virtual libraries for hit identification.

The importance of rigorous controls and checking for artifacts, both experimentally and using computational tools, will become even more important to avoid wasting resources on false positives. Follow-up studies form the lab using the σ_2 receptor as a model system have demonstrated that the molecules rising to the very top of the docking hit lists are often "cheating" molecules, and this becomes more problematic as the libraries grow.[1] Fortunately, we can identify how some of these molecules cheat the DOCK scoring function using computational tools. However, it is not a perfect system, and the newer technologies will need to address ways molecules cheat their algorithm as well. Finally, as demonstrated in **Chapters 2 & 3**, it is important to remain vigilant against confounded molecules that appear to be hits in experimental systems as well. Computational algorithms to predict potential aggregators[6] and phospholipidosis inducers[7] exist, but more work should be done to improve their functionality and to bring awareness to biologists who are testing repurposing libraries and novel docking hits in their own assays.

6.2 Advice

As noted throughout the accompanying **Gloss'** to each chapter, many challenges were faced during these projects that ultimately affected their overall outcomes, and serve as lessons learned for not only myself, but for future graduate students and the larger scientific community. Everyone is taught to "use appropriate controls", but the lessons go beyond testing controls. It is important to question the data, to not ignore data that doesn't fit your hypothesis, to sometimes start over at the beginning, and to *look* for reasons why the hypothesis is wrong. If you can't find a structure-activity relationship— something is

wrong. If you compound is weak and being tested at high concentrations— look for other ways it may be giving signal. Scrutinize the concentration-response curves. Test things in different systems, with different hands. Don't get overly invested in preliminary data and always get an $N = 3$. Look for confounding variables at every step and make an effort to mitigate them. Lastly, share this knowledge with those around you so they may avoid artifacts in their own work.

6.3 References

1. Lyu, J., Irwin, J. J. & Shoichet, B. K. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **19**, 712–718 (2023).

2. Yang, Y. *et al.* Efficient Exploration of Chemical Space with Docking and Deep Learning. *J. Chem. Theory Comput.* **17**, 7106–7119 (2021).

3. Gahbauer, S. *et al.* Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2. *Proc. National Acad. Sci.* **120**, e2212931120 (2023).

4. Schuller, M. *et al.* Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).

5. Sadybekov, A. A. *et al.* Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022).

6. Irwin, J. J. *et al.* An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).

7. Lowe, R., Glen, R. C. & Mitchell, J. B. O. Predicting Phospholipidosis Using Machine Learning. *Mol. Pharmaceut.* **7**, 1708–1714 (2010).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

Author Signature

8/15/2023

Date