

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Molecular docking in epidemics and pandemics

**Permalink**

<https://escholarship.org/uc/item/41n618z3>

**Author**

Fink, Elissa

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

Molecular docking in epidemics and pandemics

by  
Elissa Fink

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

in  
Biophysics

in the  
GRADUATE DIVISION  
of the  
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

*Brian kenton Shoichet*

C8380A8942D641D...

Brian Kenton Shoichet

Chair

DocuSigned by:

*Aashish Manglik*

DocuSigned by:

*Michael Grabe*

5B8B990D587B414...

Aashish Manglik

Michael Grabe

Committee Members

Copyright 2022

by

Elissa Fink

## **Dedication**

This work is dedicated Ukrainian people for their unwavering bravery, and to Enamine for their great impact on chemical space and drug discovery. Slava Ukraini.

## Acknowledgements

The best part of the past 4-5 years (and really, the past 8-9 years and beyond) is the support of many people, who contributed immensely to my growth as a scientist and person. I am extremely grateful for you all and the following scientific work would not have been possible without you.

First, I would like to thank my advisor, Brian Shoichet, who's mentorship, leadership, intellect, and rigor helped me become a better scientist, critical thinker, and question-asker. When I decided to pursue graduate school with an interest in computation, I thought I would be limited to modeling and predictions. However, I was able to learn so much about many more disciplines along the way because of the great ideas, projects, and collaborations Brian has established. I probably could write an entire thesis on what I have learned from Brian during my Ph.D., but here are a few of things: 1) There is only one 'formula' for writing a discussion section. 2) Always set up the question so that any outcome is publishable. 3) Do not let the data fool you, test for controls and artifacts. 4) Never doubt that novel chemistry can uncover novel biology.

I would also like to thank my thesis committee, Aashish Manglik and Michael Grabe, for their support and guidance throughout my time at UCSF. As champions of computational biophysics, drug discovery, and transmembrane proteins, their excitement for my work in thesis meetings was a consistent theme and it helped me continue to fuel my own excitement.

To the larger network of collaborators, thank you all for trusting computational predictions enough to test them in your assays, cells, and animals. It was amazing to see scientific collaboration in action. To Allan Basbaum and Joao Braz, thank you both for being fantastic scientists and friends

of the lab who cheered up every other Thursday at 1 pm and taught us so much in pain biology and pharmacology. To my  $\alpha_{2A}$  collaborators, Allan, Joao, Peter Gmeiner, Harald Hübner, Michel Bouvier, Yang Du, Jun Xu, and many others – it was amazing to see great scientific collaboration in action working on important translational research. I also want to thank Charly Craik, Conner Bardine, Danica Galonić Fujimori, Annía Rodríguez-Hernández, and Masoud Vedadi for your contributions to the work that follows. Without everyone above, I would have to stop at predictions and would not know the amazing pharmacology, structural, and therapeutic details you helped reveal.

I want to thank Yurii Moroz and everyone at Enamine and Chemspace for your innovation in expanding chemical space. This is so fundamental to the work of my thesis, our lab, and the greater community of drug discovery. Earlier in 2022, we watched in horror as the Russian invasion threatened the lives of Ukrainians. Their bravery to continue chemical synthesis amidst fighting for their freedom is indescribable and every Enamine shipment reminds me of this.

Thank you to friends I have made along the way for allowing me to learn from your own research passions and projects, and for providing a network of laughter and fun. This includes my biophysics cohort: Maru Jaime Garza, Nick Hoppe, Hayarpi Torosyan, Bryan Faust, Christina Stephens, Stephanie Wankowicz, Aji Palar, Jack Strickland, Daniel Barrero, and the rest of our larger iPQB cohort. Thanks to my CCB friends, Holly Vickery, Julian Braxton, and Matt Callahan, who understood when I wondered, “But show me a chemical structure?” and adopted me as a kind-of-chemist. To Matt, also thank you for braving CUBS, AmpC and boronic acids with me. To Holly (and Ern), thanks for providing so much fun these past few years; graduate school and the pandemic would have been so terrible without you (Ella agrees).

I also made so many friends and gained great mentors in the Shoichet lab or from the larger Shoichet alumni network. Thank you to Hayarpi for introducing me to everyone in our first year and helping me become friends with your friends in the lab. Thank you to Anat Levit Kaplan for mentoring me during my rotation, helping me transition to a fully computational scientist, and for being the best project manager anyone could ask for. To Isha Singh, thank you for being the friendliest face in the lab, always offering words of support and encouragement, and answering all of my experimental questions. Thank you, Tia Tummino, for being a great resource for really everything. I have learned so much neuroscience, Prism, and critical thinking from you. Your journey to the Shoichet lab was not designed, but I am selfishly so glad that it happened so we could become friends. To Jiankun Lyu, thank you for asking the best fundamental questions about docking, chemical databases, and any graph I have ever shown; you will be a great professor, mentor, and contributor to the field. Thank you to Stefan Gahbauer for teaching me so much about computational research and providing comic relief, both of which I am sure will continue in our next venture. Thank you to Chase Webb and Brian Bender for your constant enthusiasm for GPCRs and happy hours. To Andrii, thank you for joining us and contributing your great personality, humor, and intellect to the lab. Thank you to John Irwin, who alongside Enamine, has been instrumental in expanding chemical space, and continues to push the bounds of virtual screening. Lastly, thank you to the past and present members of the lab for being a fantastic, friendly and always supportive community.

Thank you to my Christopher Newport University mentors and support system that helped me get to graduate school and continue to support me: Kathryn Cole, Joshua Patterson, and Todd Gruber. Also, thank you to my mentor at the University of Pittsburgh, Lillian Chong, for allowing me to spend a summer in her lab learning more about computational research.

To my family, Mark, Shirley, Saskia, Dustin, Henry, Adera, Maren, Ada, Pop Pop, Granny San, Paw Paw, Dor Dor, Pam, Renate, and George, Mariner families, and many, many more, thank you for always supporting my adventures, especially as they take me to faraway places, and listening to and sharing in my excitement as I talk about my research. Ευχαριστώ πολύ, Paul, for your unwavering support, laughter, and adventures throughout the past few years. Lastly, thank you to my dog, Ella, for helping me get through school, the pandemic, and life.



## 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. Peter Gmeiner, Dr. Michel Bouvier, Dr. Yang Du, Dr. Masoud Vedadi, Dr. John Irwin, and Dr. Charles Craik.

Chapter 2 is a reprint of the material:

“Structure-based discovery of nonopioid analgesics acting through the  $\alpha_{2A}$ -adrenergic receptor” as it appears in **Fink EA**, Xu J, Hübner H, Braz JM, Seemann P, Avet C, Craik V, Weikert D, Schmidt MF, Webb CM, Tolmachova NA, Moroz YS, Huang XP, Kalyanaraman C, Gahbauer S, Chen G, Liu Z, Jacobson MP, Irwin JJ, Bouvier M, Du Y, Shoichet BK, Basbaum AI, Gmeiner P. *Science*. 2022 Sep 30; **377**(6614). DOI: 10.1126/science.abn7065. **Fink EA**, Xu J, Hübner H, Braz JM, and Seemann P contributed equally to this work. Bouvier M, Du Y, Shoichet BK, Basbaum AI, and Gmeiner P are corresponding authors who supervised the research.

Chapter 3 is a manuscript in preparation:

Singh I, Li F, **Fink E**, Chau I, Li A, Rodriguez-Hernández A, Zapatero-Belinchón FJ, Rodriguez M, Devkota K, Deng Z, White K, Wan X, Tolmachova NA, Moroz YS, Gross J, Kaniskan HU, Ott M, García-Sastre A, Jin J, Galonić Fujimori D, Irwin JJ, Vedadi M, Shoichet BK. Large library docking for novel SARS-CoV-2 main protease non-covalent and covalent inhibitors. Co-first authors, Singh I, Li F, and **Fink E**, contributed equally to this work. Corresponding authors, Vedadi M and Shoichet BK, supervised the research.

Chapter 4 is a manuscript in preparation:

**Fink EA**, Bardine C, Gahbauer S, Singh I, White K, Gu S, Wan X, Ary B, Glenn I, O’Connell J, O’Donnell H, Fajtová P, Lyu J, Vigneron S, Young NJ, Kondratov IS, O’Donoghue AJ, Moroz YS, Taunton J, Renslo AR, Irwin JJ, García-Sastre A, Shoichet BK, Craik CS. Structure-based discovery of inhibitors of the SARS-CoV-2 Nsp14 N7-methyltransferase. The following author contributed equally – **Fink EA**, Bardine C, Gahbauer S, and Singh I. Corresponding authors who supervised the research are Shoichet BK and Craik CS.

# Molecular docking in epidemics and pandemics

Elissa Fink

## Abstract

There is always a new therapy to be discovered, either by overcoming side effects or other liabilities in current drugs, or through exploring new technologies, proteins, or diseases. Focusing on small molecules, the initial stages of the drug discovery process require the identification of molecules modulating a protein of interest. In Chapter 1, one approach available for discovery is described – molecular docking to discover new small molecules of interest from virtual chemical databases. This thesis applies the approach to overcome side effects of current opioid pain drugs, to identify new antivirals for SARS-CoV-2, and to probe larger applications of the growing chemical space.

Chapter 2 describes how the computational approach identifies new nonopioid pain therapeutics acting through the  $\alpha_{2A}$ -adrenergic receptor ( $\alpha_{2A}AR$ ). Here, we show the translational impact of computational predictions as many are efficacious in mice. We argue a case for further drug discovery focused on this protein, because not only were the opioid side effects avoided, but also a characteristic  $\alpha_{2A}AR$ -caused side effect is not seen with new compounds.

Chapter 3 demonstrates application of the computational technique to an understudied viral protein, the SARS-CoV-2 nonstructural protein 14 (nsp14), using different subsets of the chemical databases for targeted results. Here, ‘traditional’ noncovalent ‘lead-like’ subsets, smaller ‘fragment’ molecules, and covalent molecules are utilized and successfully identify nsp14 inhibitors. We show in a pandemic-driven project that multiple techniques are needed to progress towards antivirals inhibiting an understudied protein.

Chapter 4 continues the themes of Chapter 3, but with the SARS-CoV-2 main protease (M<sup>Pro</sup>). Here, we use a similar approach with different types of chemistry found in the databases and identify numerous inhibitors. Structural studies demonstrate the computationally predicted and experimental ligand geometries are in high agreement. Finally, we begin to probe the antiviral efficacy of these new compounds and contribute the knowledge to the field for further antiviral development.

Chapter 5 focuses on a new application of the virtual screening of chemical databases. Here, we ask if this approach can be applied prospectively to discover molecules with polypharmacology – having designed activities on multiple proteins of interest. We simulate the effect of the growing chemical databases to understand their effect on docking for polypharmacology. This chapter also blends into the most immediate future directions of the thesis work and is described in context of pain therapeutic discovery.

## Table of Contents

<b>INTRODUCTION TO VIRTUAL SCREENING .....</b>	<b>1</b>
1.1 REFERENCES .....	4
<b>GLOSS TO CHAPTER 2 .....</b>	<b>9</b>
<b>STRUCTURE-BASED DISCOVERY OF NONOPIOID ANALGESICS ACTING THROUGH THE <math>\alpha_2</math>A-ADRENERGIC RECEPTOR.....</b>	<b>11</b>
2.1 ABSTRACT.....	13
2.2 INTRODUCTION .....	14
2.3 RESULTS .....	15
2.4 DISCUSSION .....	27
2.5 FIGURES .....	30
2.6 TABLES .....	58
2.7 MATERIALS AND METHODS .....	76
2.8 ACKNOWLEDGEMENTS.....	107
2.9 AUTHOR CONTRIBUTIONS.....	108
2.10 REFERENCES .....	109
<b>GLOSS TO CHAPTER 3.....</b>	<b>123</b>
<b>STRUCTURE-BASED DISCOVERY OF INHIBITORS OF THE SARS-COV-2 NSP14 N7- METHYLTRANSFERASE .....</b>	<b>125</b>
3.1 ABSTRACT.....	127
3.2 INTRODUCTION .....	128
3.3 RESULTS .....	131
3.4 DISCUSSION .....	141
3.5 FIGURES .....	144
3.6 TABLES .....	159
3.7 MATERIALS AND METHODS .....	170

3.8 ACKNOWLEDGEMENTS.....	181
3.9 AUTHOR CONTRIBUTIONS.....	182
3.10 REFERENCES.....	183
<b>GLOSS TO CHAPTER 4.....</b>	<b>196</b>
<b>LARGE LIBRARY DOCKING FOR NOVEL SARS-COV-2 MAIN PROTEASE NON-COVALENT INHIBITORS.....</b>	<b>197</b>
4.1 ABSTRACT.....	199
4.2 INTRODUCTION.....	200
4.3 RESULTS.....	202
4.4 DISCUSSION.....	213
4.5 FIGURES.....	216
4.6 TABLES.....	236
4.7 MATERIALS AND METHODS.....	245
4.6 ACKNOWLEDGEMENTS.....	258
4.7 AUTHOR CONTRIBUTIONS.....	259
4.8 REFERENCES.....	260
<b>GLOSS TO CHAPTER 5.....</b>	<b>281</b>
<b>POLYPHARMACOLOGY AND EXPANDING VIRTUAL CHEMICAL SPACE.....</b>	<b>283</b>
5.1 INTRODUCTION.....	284
5.2 RESULTS.....	286
5.3 DISCUSSION.....	291
5.4 FUTURE DIRECTIONS.....	293
5.5 FIGURES.....	294
5.6 TABLES.....	300
5.7 MATERIALS AND METHODS.....	303
5.8 ACKNOWLEDGEMENTS.....	305
5.9 AUTHOR CONTRIBUTIONS.....	306

5.10 REFERENCES .....307

## List of Figures

Figure 2.1 Newly discovered $\alpha_{2A}$ AR agonists from ultralarge library docking. ....	30
Figure 2.2 Docking-predicted poses of ‘9087 and ‘4622 superpose well on the cryo-EM structures of the ‘9087- $\alpha_{2A}$ AR-GoA and ‘4622- $\alpha_{2A}$ AR-GoA complexes .....	32
Figure 2.3 Structure-based optimization of ‘9087. ....	33
Figure 2.4 The docking-derived agonists are antinociceptive in neuropathic, inflammatory, and acute thermal pain through the $\alpha_{2A}$ AR but are not sedating .....	35
Figure 2.S1 Structural similarity of adrenergic receptors. ....	36
Figure 2.S2 Functional data for docking hits against $\alpha_{2A}$ AR. ....	37
Figure 2.S3 Functional data for docking hits against $\alpha_{2B}$ AR. ....	39
Figure 2.S4 $G_i$ -activation induced cAMP inhibition assay against $\alpha_{2A}$ AR. ....	40
Figure 2.S5 Functional properties of norepinephrine, selected docking agonists, and the bespoke synthesized analog PS75 are dependent on receptor density. ....	41
Figure 2.S6 EMTA coupling panel for select docking compounds against $\alpha_{2A}$ AR. ....	42
Figure 2.S7 Relative activities for select docking compounds against $\alpha_{2A}$ AR EMTA coupling panel. ....	43
Figure 2.S8 Internalization behavior of $\alpha_{2A}$ AR following compound treatment. ....	44
Figure 2.S9 Structural determination of ‘9087- $\alpha_{2A}$ AR- $G_{oA}$ . ....	45
Figure 2.S10 Structural determination of ‘4622- $\alpha_{2A}$ AR- $G_{oA}$ . ....	46
Figure 2.S11 ‘9087 formal and partial charges. ....	48
Figure 2.S12 Substitutions of binding site residues reveals distinct interaction patterns for known agonists and potent docking agonists. ....	50
Figure 2.S13 Substitution of binding site residues reveals distinct interaction patterns for known agonists and potent docking agonists. ....	52
Figure 2.S14 Analogs of ‘9087 reveal key SAR. ....	53
Figure 2.S15 Functional data for selected docking hits and references against $\alpha_{2A}$ AR. ....	54
Figure 2.S16 Off-target activity for $\alpha_{2A}$ AR agonists. ....	55
Figure 2.S17 Phase 1 metabolism of ‘9087, ‘7075, and PS75 in male rat liver microsomes. ....	56
Figure 2.S18 In vivo side effects of constipation and body weight. ....	57



Figure 3.1 Workflow for inhibitor discovery against N7-MTase domain of nsp14 using molecular docking.....	144
Figure 3.2. Ultra-large scale docking identifies three nsp14 inhibitors with novel chemical scaffolds .....	145
Figure 3.3. Hit optimization of the non-covalent compounds ‘9213, ‘4824, ‘2882 .....	146
Figure 3.4. Fragment inhibitors from 16M docking screen .....	147
Figure 3.5. Docking 25 million electrophiles reveals aldehyde and acrylamide inhibitors.....	148
Figure 3.6. MTase selectivity of docking-derived inhibitors.....	149
Figure 3.7. Assessing the antiviral activity of nsp14 inhibitors.....	150
Figure 3.S1. Assessment of reversibility of inhibition by rapid dilution.....	151
Figure 3.S2. Mechanism of action of inhibitors .....	152
Figure 3.S3. Concentration-response curves of non-covalent fragment hits .....	153
Figure 3.S4. Docked poses of non-covalent fragment hits .....	154
Supplementary Figure 3.S5. Concentration-response curves of covalent hits.....	155
Supplementary Figure 3.S6. Docked poses of covalent hits.....	156
Supplementary Figure 3.S7. Evaluating possible covalent binding of compound ‘1911, acryl42, and acryl42-10 to nsp14 by mass spectrometry.....	157
Supplementary Figure 3.S8. Antiviral and cell viability experiments for nsp14 inhibitors .....	158
Figure 4.1. Substrate design and assay development allows structure-based inhibitor discovery .....	216
Figure 4.2. Non-covalent compound optimization to low- $\mu$ M potencies .....	218
Figure 4.3. Covalent hits from 6.5 million virtual screen .....	220
Figure 4.4. Compound optimization of aldehyde ‘3620 .....	221
Figure 4.5. Antiviral activity and pan-coronaviral M <sup>Pro</sup> inhibition by covalent analogs .....	222
Figure 4.S1. Assay optimization for solvent and detergent.....	223
Figure 4.S2.1. Non-covalent docking hits or compounds with >30% inhibition from first virtual screen .....	224
Figure 4.S2.2. Non-covalent docking hits or compounds with >30% inhibition from first virtual screen .....	225
Figure 4.S2.3. Non-covalent docking hits from second virtual screen.....	226
Figure 4.S2.4. Non-covalent docking hits from second virtual screen.....	227
Figure 4.S3. Evaluating aggregation potential of initial docking hits and potent analogs .....	229
Figure 4.S4. LigPlot visualization of MPro-noncovalent inhibitor interactions in newly solved structures.....	230

Figure 4.S5. Docked poses of covalent hits .....	231
Figure 4.S6. Reversibility of compound '7021.....	232
Figure 4.S7. '3620 analogs with improved potencies.....	233
Figure 4.S8. LigPlot visualization of MPro-covalent inhibitor interactions in newly solved structures.....	234
Figure 4.S9. Pan-viral enzymatic activities .....	235
Figure 5.1. Random overlap of shared top-ranked molecules are not enriched with database growth .....	294
Figure 5.2. The number of shared top-ranked molecules increase with database size .....	295
Figure 5.3. Enrichment of shared molecules is greatest at mid-top-ranked docking results .....	296
Figure 5.4. Docking results in the top-ranked percent thresholds increases with database size.....	297
Figure 5.5. Enrichment of shared molecules over random reaches a plateau at different thresholds of percent top-ranked results.....	298
Figure 5.6. Database growth analysis with prospective polypharmacology pair SERT- $\mu$ OR.....	299

## List of Tables

Table 2.S1 Binding affinities and functional properties at $\alpha_{2A}$ AR or $\alpha_{2B}$ AR for hits identified in initial docking screen. ....	58
Table 2.S2 Binding affinity and functional properties of selected compounds in comparison to the reference compounds norepinephrine, dexmedetomidine, and clonidine to the human $\alpha_{2A}$ AR. ....	64
Table 2.S3 Functional properties in EMTA coupling panel of '9087 and its analogs in comparison to norepinephrine, dexmedetomidine, and brimonidine against the human $\alpha_{2A}$ AR. ....	65
Table 2.S4 EMTA coupling panel relative activities of '9087 and its analogs in comparison to norepinephrine, dexmedetomidine, and brimonidine against the human $\alpha_{2A}$ AR. ....	67
Table 2.S5 Summary of $G_{\text{ai}}$ BRET activation data for compounds and $\alpha_{2A}$ AR mutations. ....	69
Table 2.S6 Summary of $G_{\text{ai}}$ BRET activation and arrestin BRET recruitment data for selected compounds at the $\alpha_{2A}$ AR and selected mutants displayed as delta BRET values. ....	70
Table 2.S7 Binding affinity and functional properties of selected '9087 analogs to the human $\alpha_{2A}$ AR. ....	71
Table 2.S8 Binding affinity of '9087 to other adrenergic receptors. ....	73
Table 2.S9 Pharmacokinetic properties of docking agonists. ....	74
Table 2.S10 Cryo-EM data collection, refinement, and validation statistics. ....	75
Table 3.1. Expanded DOCKoValent electrophile databases ....	159
Table 3.2. Docking inhibitors are novel, non-SAM like chemotypes ....	160
Table 3.S1. Compound optimization for '2882 ....	165
Table 3.S2. Compound optimization for '9213 ....	166
Table 3.S3. Compound optimization for '4824 ....	167
Table 3.S4. Compound optimization for acryl42. ....	168
Table 3.S5. Permeability of selected inhibitors and reference compounds ....	169
Table 4.1. Hits from the first non-covalent docking screen. ....	236
Table 4.2. Hits from the first non-covalent docking screen. ....	237
Table 4.S1 Analogs of covalent docking hit '3620 with improved potencies ....	238
Table 4.S2. Antiviral activities ....	241
Table 4.S3. Pan-viral activities of '7021 ....	242

Table 4.S4. Crystallographic statistics.....	243
Table 5.1. Polypharmacology model systems.....	300
Table 5.S1. Enrichment of shared molecules linear regression results.....	301
Table 5.S2. Fold enrichment of top-ranked percent thresholds over random.....	302

## Chapter 1

### Introduction to Virtual Screening

If I could spend the rest of my working life doing one thing, it might be to look at 3-dimensional protein-ligand complexes. When I learn of new proteins or small molecules, one of my initial reactions is to see if an experimentally determined protein structure exists in the Protein Database (PDB; [www.rcsb.org](http://www.rcsb.org)), and then to see if there is a structure with ligand bound. Ever wondered what morphine (1) (8EF6), a hallucinogen (2) (6WGT), or ondansetron (3) (6W1M; my favorite drug) look like in 3D while at work in your body? When looking at these structures, it is so interesting to see what interactions (4) the proteins require from ligands that bind to them, especially if you can exploit them further to design or discover new small molecule therapeutics.

There are many techniques to discovery new ligands that bind, inhibit, or activate a protein of interest, both experimental and computational (5, 6). Akin to my own interest, molecular docking is a computational technique that combines a 3D protein structure with 3D structures of molecules to evaluate their complementarity (7). Historically, these predictions have successfully identified new ligands for proteins for use as chemical research probes (8, 9) or precursors to potential therapeutics (10–14). Understanding these predictions with experiments has, and at the same time, has not, evolved over time. We can test if the molecules do in fact bind, and this is often the primary metric to determine if our computation performed well. Using x-ray crystallography, and now cryo-electron microscopy (cryo-EM), we can compare the predicted and experimentally determined protein-ligand interactions (15–18). For therapeutics, some proteins require more specific modulation of a small molecule and that too can be tested experimentally in cells or even

animals (19–21). Together, we can expand the story of a molecule past its immediate protein-ligand interactions.

The computational throughput has increased with time, thanks to improving software and hardware. Previously, computing protein-ligand complexes took a long time and the number of ligands one could test had to remain small. At some point, throughput increased and now thousands and millions of molecules could be tested computationally (15, 16, 22–24). In parallel, great chemists aspired to increase “chemical space” – the predicted total number of  $10^{63}$  molecules (25) that could potentially exist on Earth. The race to improve hardware and collections of molecules in virtual chemical databases (26, 27) continued... the field passed thresholds of tens-of-millions, hundreds-of-millions, and billions of molecules available for molecular docking.

Even if we (the collective, not me) could synthesize all of the possible chemicals on Earth, why do it? Often, the bigger the better. As more and more molecules were added into the databases and molecular docking screens this held true (15, 16, 28). More and more molecules that bind and modulate a protein of interest occurred. Currently, the success rate of the computational predictions falls in 20-60% (8, 10, 12, 14, 16, 29, 30). With larger databases, some new biological outcomes were achieved sometimes with very important physiological outcomes (10–12, 14). More potent molecules are identified, new types of chemical scaffolds are discovered even for proteins that are very widely studied with lots of characterized ligands available (13, 15), new protein-ligand interactions are observed, and in turn differentiating cellular effects are discovered. In context to therapeutics, the technique has identified new  $\mu$ -opioid receptor pain therapeutics thought to

hopefully have less side effects (10), anti-depressants acting similar to LSD but without psychedelic effects (13), circadian rhythm modulators (14), to name a few studies.

It is rewarding when computational predictions a) are tested by people, b) a large proportion actually work experimentally, and c) they have interesting features that continue to reveal new things about even the most highly studied protein and signaling systems. It is also very special to apply a successful, fun technique to problems that do not just affect scientists, but the general public across the world. That is a value that guides the following research in this thesis.

In the following chapters, I apply molecular docking and chemical databases to discover new pain and antiviral therapeutics in light of on-going epidemics and pandemics. Specific to the opioid epidemic in Chapter 2, I explore nonopioid proteins that could be modulated in pain signaling circuitry. This is a larger theme in the lab, and in combining the overall work, I also begin to apply the concept of polypharmacology to pain in Chapter 5. With polypharmacology, we hope to discover new ligands that can modulate multiple protein nodes of the pain physiological network for a more efficacious therapeutic. In the midst of my Ph.D. a pandemic broke out, halting life but at the same time creating great areas of public need for new SARS-CoV-2 antiviral therapeutics. New discovery efforts were started, and in Chapters 3 and 4 I used different types of chemistry within the virtual chemical databases to identify inhibitors of two viral proteins. I will revisit these themes and motivations throughout the chapters with introductory Glosses.

## 1.1 References

1. Y. Zhuang, Y. Wang, B. He, X. He, X. E. Zhou, S. Guo, Q. Rao, J. Yang, J. Liu, Q. Zhou, X. Wang, M. Liu, W. Liu, X. Jiang, D. Yang, H. Jiang, J. Shen, K. Melcher, H. Chen, Y. Jiang, X. Cheng, M.-W. Wang, X. Xie, H. E. Xu, Molecular recognition of morphine and fentanyl by the human  $\mu$ -opioid receptor. *Cell*. **185**, 4361-4375.e19 (2022).
2. K. Kim, T. Che, O. Panova, J. F. DiBerto, J. Lyu, B. E. Krumm, D. Wacker, M. J. Robertson, A. B. Seven, D. E. Nichols, B. K. Shoichet, G. Skiniotis, B. L. Roth, Structure of a Hallucinogen-Activated Gq-Coupled 5-HT<sub>2A</sub> Serotonin Receptor. *Cell*. **182**, 1574-1588.e19 (2020).
3. S. Basak, A. Kumar, S. Ramsey, E. Gibbs, A. Kapoor, M. Filizola, S. Chakrapani, High-resolution structures of multiple 5-HT<sub>3A</sub>R-setron complexes reveal a novel mechanism of competitive inhibition. *eLife*. **9**, e57870 (2020).
4. C. Bissantz, B. Kuhn, M. Stahl, A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **53**, 5061–5084 (2010).
5. V. Blay, B. Tolani, S. P. Ho, M. R. Arkin, High-Throughput Screening: today's biochemical and cell-based approaches. *Drug Discov. Today*. **25**, 1807–1821 (2020).
6. E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, A. G. Taranto, Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* **8**, 343 (2020).
7. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).
8. A. Levit Kaplan, R. T. Strachan, J. M. Braz, V. Craik, S. Slocum, T. Mangano, V. Amabo, H. O'Donnell, P. Lak, A. I. Basbaum, B. L. Roth, B. K. Shoichet, Structure-Based Design of



- a Chemical Probe Set for the 5-HT<sub>5A</sub> Serotonin Receptor. *J. Med. Chem.* **65**, 4201–4217 (2022).
9. E. Gregori-Puigjané, V. Setola, J. Hert, B. A. Crews, J. J. Irwin, E. Lounkine, L. Marnett, B. L. Roth, B. K. Shoichet, Identifying mechanism-of-action targets for drugs and probes. *Proc. Natl. Acad. Sci.* **109**, 11178–11183 (2012).
  10. A. Manglik, H. Lin, D. K. Aryal, J. D. McCorvy, D. Dengler, G. Corder, A. Levit, R. C. Kling, V. Bernat, H. Hübner, X.-P. Huang, M. F. Sassano, P. M. Giguère, S. Löber, Da Duan, G. Scherrer, B. K. Kobilka, P. Gmeiner, B. L. Roth, B. K. Shoichet, Structure-based discovery of opioid analgesics with reduced side effects. *Nature.* **537**, 185–190 (2016).
  11. I. Singh, A. Seth, C. B. Billesbølle, J. Braz, R. M. Rodriguiz, K. Roy, B. Bekele, V. Craik, X.-P. Huang, D. Boytsov, P. Lak, H. O'Donnell, W. Sandtner, B. L. Roth, A. I. Basbaum, W. C. Wetsel, A. Manglik, B. K. Shoichet, G. Rudnick, “Structure-based Discovery of Conformationally Selective Inhibitors of the Serotonin Transporter” (preprint, Biochemistry, 2022), , doi:10.1101/2022.06.13.495991.
  12. E. A. Fink, J. Xu, H. Hübner, J. M. Braz, P. Seemann, C. Avet, V. Craik, D. Weikert, M. F. Schmidt, C. M. Webb, N. A. Tolmachova, Y. S. Moroz, X.-P. Huang, C. Kalyanaraman, S. Gahbauer, G. Chen, Z. Liu, M. P. Jacobson, J. J. Irwin, M. Bouvier, Y. Du, B. K. Shoichet, A. I. Basbaum, P. Gmeiner, Structure-based discovery of nonopioid analgesics acting through the  $\alpha$ 2A-adrenergic receptor. *Science.* **377**, eabn7065 (2022).
  13. A. L. Kaplan, D. N. Confair, K. Kim, X. Barros-Álvarez, R. M. Rodriguiz, Y. Yang, O. S. Kweon, T. Che, J. D. McCorvy, D. N. Kamber, J. P. Phelan, L. C. Martins, V. M. Pogorelov, J. F. DiBerto, S. T. Slocum, X.-P. Huang, J. M. Kumar, M. J. Robertson, O. Panova, A. B. Seven, A. Q. Wetsel, W. C. Wetsel, J. J. Irwin, G. Skiniotis, B. K. Shoichet, B. L. Roth, J. A.

- Ellman, Bespoke library docking for 5-HT<sub>2A</sub> receptor agonists with antidepressant activity. *Nature*. **610**, 582–591 (2022).
14. R. M. Stein, H. J. Kang, J. D. McCorvy, G. C. Glatfelter, A. J. Jones, T. Che, S. Slocum, X.-P. Huang, O. Savych, Y. S. Moroz, B. Stauch, L. C. Johansson, V. Cherezov, T. Kenakin, J. J. Irwin, B. K. Shoichet, B. L. Roth, M. L. Dubocovich, Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature*. **579**, 609–614 (2020).
  15. J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O’Meara, T. Che, E. Alga, K. Tolmacheva, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, Ultra-large library docking for discovering new chemotypes. *Nature*. **566**, 224–229 (2019).
  16. A. Alon, J. Lyu, J. M. Braz, T. A. Tummino, V. Craik, M. J. O’Meara, C. M. Webb, D. S. Radchenko, Y. S. Moroz, X.-P. Huang, Y. Liu, B. L. Roth, J. J. Irwin, A. I. Basbaum, B. K. Shoichet, A. C. Kruse, Structures of the  $\sigma_2$  receptor enable docking for bioactive ligand discovery. *Nature*. **600**, 759–764 (2021).
  17. S. Morandi, F. Morandi, E. Caselli, B. K. Shoichet, F. Prati, Structure-based optimization of cephalothin-analogue boronic acids as  $\beta$ -lactamase inhibitors. *Bioorg. Med. Chem.* **16**, 1195–1205 (2008).
  18. B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, K. M. Perry, Structure-Based Discovery of Inhibitors of Thymidylate Synthase. *Science*. **259**, 1445–1450 (1993).
  19. X.-P. Huang, J. Karpiak, W. K. Kroeze, H. Zhu, X. Chen, S. S. Moy, K. A. Saddoris, V. D. Nikolova, M. S. Farrell, S. Wang, T. J. Mangano, D. A. Deshpande, A. Jiang, R. B. Penn, J. Jin, B. H. Koller, T. Kenakin, B. K. Shoichet, B. L. Roth, Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature*. **527**, 477–483 (2015).

20. P. Kolb, D. M. Rosenbaum, J. J. Irwin, J. J. Fung, B. K. Kobilka, B. K. Shoichet, Structure-based discovery of beta2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6843–6848 (2009).
21. B. L. Roth, J. J. Irwin, B. K. Shoichet, Discovery of new GPCR ligands to illuminate new biology. *Nat. Chem. Biol.* **13**, 1143–1151 (2017).
22. A. A. Sadybekov, A. V. Sadybekov, Y. Liu, C. Iliopoulos-Tsoutsouvas, X.-P. Huang, J. Pickett, B. Houser, N. Patel, N. K. Tran, F. Tong, N. Zvonok, M. K. Jain, O. Savych, D. S. Radchenko, S. P. Nikas, N. A. Petasis, Y. S. Moroz, B. L. Roth, A. Makriyannis, V. Katritch, Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature.* **601**, 452–459 (2022).
23. C. Gorgulla, A. Boeszoermyeni, Z.-F. Wang, P. D. Fischer, P. W. Coote, K. M. Padmanabha Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner, H. Arthanari, An open-source drug discovery platform enables ultra-large virtual screens. *Nature.* **580**, 663–668 (2020).
24. P. Beroza, J. J. Crawford, O. Ganichkin, L. Gendeleev, S. F. Harris, R. Klein, A. Miu, S. Steinbacher, F.-M. Klingler, C. Lemmen, Chemical space docking enables large-scale structure-based virtual screening to discover ROCK1 kinase inhibitors. *Nat. Commun.* **13**, 6447 (2022).
25. R. S. Bohacek, C. McMartin, W. C. Guida, The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996).
26. T. Sterling, J. J. Irwin, ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).

27. B. Tingle, K. Tang, J. Castanon, J. Gutierrez, M. Khurelbaatar, C. Dandarchuluun, Y. Moroz, J. Irwin, “ZINC-22 - A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery” (preprint, Chemistry, 2022), , doi:10.26434/chemrxiv-2022-82czl.
28. J. Lyu, J. Irwin, B. Shoichet, “Modeling the expansion of virtual screening libraries” (preprint, Chemistry, 2022), , doi:10.26434/chemrxiv-2022-6lv34-v2.
29. M. Schuller, G. J. Correy, S. Gahbauer, D. Fearon, T. Wu, R. E. Díaz, I. D. Young, L. Carvalho Martins, D. H. Smith, U. Schulze-Gahmen, T. W. Owens, I. Deshpande, G. E. Merz, A. C. Thwin, J. T. Biel, J. K. Peters, M. Moritz, N. Herrera, H. T. Kratochvil, QCRG Structural Biology Consortium, A. Aimon, J. M. Bennett, J. Brandao Neto, A. E. Cohen, A. Dias, A. Douangamath, L. Dunnett, O. Fedorov, M. P. Ferla, M. R. Fuchs, T. J. Gorrie-Stone, J. M. Holton, M. G. Johnson, T. Krojer, G. Meigs, A. J. Powell, J. G. M. Rack, V. L. Rangel, S. Russi, R. E. Skyner, C. A. Smith, A. S. Soares, J. L. Wierman, K. Zhu, P. O’Brien, N. Jura, A. Ashworth, J. J. Irwin, M. C. Thompson, J. E. Gestwicki, F. von Delft, B. K. Shoichet, J. S. Fraser, I. Ahel, Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).
30. A. Rudling, R. Gustafsson, I. Almlöf, E. Homan, M. Scobie, U. Warpman Berglund, T. Helleday, P. Stenmark, J. Carlsson, Fragment-Based Discovery and Optimization of Enzyme Inhibitors by Docking of Commercial Chemical Space. *J. Med. Chem.* **60**, 8160–8169 (2017).

## Gloss to Chapter 2

This chapter marks the beginning of my thesis research as my first project I took on after joining the Shoichet lab. The timing coincided with the beginning of a large DARPA grant focused on finding new pain therapeutics acting on any viable protein that was not the  $\mu$ -opioid receptor ( $\mu$ OR), the protein through which opioids are efficacious. Naturally, I gained my own protein target, the  $\alpha_{2A}$ -adrenergic receptor ( $\alpha_{2A}$ AR), that was “highly recommended” by Brian. So, I set out with my first docking project.

Fortunately, we knew  $\alpha_{2A}$ ARs are involved in physiological processes that can stop the perception of pain, and that some FDA-approved  $\alpha_{2A}$ AR drugs are used in particular hospital settings for treating pain. My contributions include setting up and optimizing computational software and files leading to the ‘ultra-large’ virtual screen. While much of this is glossed over in the following chapter, this is where I learned a lot about computational control experiments. Following the 301 million molecules computationally screened against  $\alpha_{2A}$ AR, I also learned how to more rigorously ‘hit-pick’ molecules, or put another way, select 50 to 100 high quality molecules to purchase for synthesis. Then, compounds were handed off to pharmacologists (Gmeiner and Bouvier labs), structural biologists (Du lab), and *in vivo* pharmacologists and neurobiologists (Basbaum lab). We learned through the culmination of their great work, that many compounds could modulate  $\alpha_{2A}$ ARs in cells and in mice, with pain relieving effects in the latter. Lots of additional compounds were designed or experiments were performed. Perhaps the most surprising result was that the compounds did not have the traditional  $\alpha_{2A}$ AR-caused effect of sedation in mice. The mechanism to which sedation does not occur is unknown, but further investigation will be valuable knowledge for continuing the search of nonopioid pain therapeutics.

Things I learned along the way that are not in the following chapter: how to say dexmedetomidine (dex-med-uh-toe-meh-dean), how \*fun\* it is to dissolve many small compound vials of 1-10 mg, how to efficiently, although sometimes very inefficiently, ship to Germany, Canada, China, Ukraine, and India, and how to tell reviewers you cannot add a full literature review into your already very large manuscript.

Now some serious things I learned, also not in the following chapter. When I came to UCSF, I hoped to do ‘computational drug discovery,’ but I thought I would only scratch the surface of what my vision of this was. The following work demonstrated to me, and to the field along with other papers from the lab, that actually we can do so much with these computational screens and have a great impact. I could have never guessed compounds coming straight from computational predictions would get through *in vitro* analysis unscathed and be efficacious in mice with no optimization. Maybe it was luck with the right protein, maybe not.

I also learned how great and rewarding it is to bring together experts for scientific collaboration. Having others test our computational predictions in cells, animals, and other assays for their ‘in real life’ activities is immensely rewarding. As a team we can then tell the story of each molecule. Through this great network of scientists, I also had the opportunity to learn from their respective areas of expertise. It has helped me greatly improve my ability to digest data of all types and communicate it forward to other audiences of all backgrounds. And finally, this was such an enjoyable experience with fantastic, friendly people that made the project, meetings (even ones at 7 am), and process fun.

## Chapter 2

### Structure-based discovery of nonopioid analgesics acting through the $\alpha_{2A}$ -adrenergic receptor

#### Contributing Authors

Elissa A. Fink<sup>1,2,†</sup>, Jun Xu<sup>3,4,†</sup>, Harald Hübner<sup>5,†</sup>, Joao M. Braz<sup>6,†</sup>, Philipp Seemann<sup>5,†</sup>, Charlotte Avet<sup>7</sup>, Veronica Craik<sup>6</sup>, Dorothee Weikert<sup>5</sup>, Maximilian F. Schmidt<sup>5</sup>, Chase M. Webb<sup>1,8</sup>, Nataliya A. Tolmachova<sup>9,10</sup>, Yurii S. Moroz<sup>11,12</sup>, Xi-Ping Huang<sup>13</sup>, Chakrapani Kalyanaraman<sup>1</sup>, Stefan Gahbauer<sup>1</sup>, Geng Chen<sup>3</sup>, Zheng Liu<sup>3</sup>, Matthew P. Jacobson<sup>1</sup>, John J. Irwin<sup>1</sup>, Michel Bouvier<sup>7,\*</sup>, Yang Du<sup>3,\*</sup>, Brian K. Shoichet<sup>1,\*</sup>, Allan I. Basbaum<sup>6,\*</sup>, & Peter Gmeiner<sup>5,\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA.

<sup>2</sup>Graduate Program in Biophysics, University of California, San Francisco, San Francisco, CA, USA.

<sup>3</sup>Kobilka Institute of Innovative Drug Discovery, School of Life and Health Sciences, Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, China.

<sup>4</sup>Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, CA, USA.

<sup>5</sup>Department of Chemistry and Pharmacy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nikolaus-Fiebiger-Str. 10, 91058, Erlangen, Germany.

<sup>6</sup>Department of Anatomy, University of California, San Francisco, San Francisco, CA, USA.

<sup>7</sup>Department of Biochemistry and Molecular Medicine, Institute for Research in Immunology and Cancer, Université de Montréal, QC, Canada.

<sup>8</sup>Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, University of California, San Francisco, San Francisco, CA, USA.

<sup>9</sup>Enamine Ltd., Chervonotkatska St. 78, 02094 Kyiv, Ukraine.

<sup>10</sup>Institute of Bioorganic Chemistry and Petrochemistry, National Ukrainian Academy of Science, Murmanska St. 1, 02660 Kyiv, Ukraine.

<sup>11</sup>National Taras Shevchenko University of Kyiv, 60 Volodymyrska Street, Kyiv 01601, Ukraine.

<sup>12</sup>Chemspace, ilukstes iela 38-5, Riga, LV-1082, Latvia.

<sup>13</sup>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, USA.

†These authors contributed equally.

\*Corresponding authors: [peter.gmeiner@fau.de](mailto:peter.gmeiner@fau.de) (P.G.); [allan.basbaum@ucsf.edu](mailto:allan.basbaum@ucsf.edu) (A.I.B.); [bshoichet@gmail.com](mailto:bshoichet@gmail.com) (B.K.S.); [yangdu@cuhk.edu.cn](mailto:yangdu@cuhk.edu.cn) (Y.D.); [michel.bouvier@umontreal.ca](mailto:michel.bouvier@umontreal.ca) (M.B.)



## 2.1 Abstract

Because nonopioid analgesics are much sought after, we computationally docked more than 301 million virtual molecules against a validated pain target, the  $\alpha_{2A}$ -adrenergic receptor ( $\alpha_{2A}$ AR), seeking new  $\alpha_{2A}$ AR agonists chemotypes that lack the sedation conferred by known  $\alpha_{2A}$ AR drugs, such as dexmedetomidine. We identified 17 ligands with potencies as low as 12 nanomolar, many with partial agonism and preferential  $G_i$  and  $G_o$  signaling. Experimental structures of  $\alpha_{2A}$ AR complexed with two these agonists confirmed the docking predictions and templated further optimization. Several compounds, including the initial docking hit **'9087** [mean effective concentration ( $EC_{50}$ ) 52 nanomolar], and two analogs, **'7075** and **PS75** ( $EC_{50}$  4.1 and 4.8 nanomolar), exerted on-target analgesic activity in multiple *in vivo* pain models, without sedation. These newly discovered agonists are interesting as therapeutic leads that lack the liabilities of opioids and the sedation of dexmedetomidine.

## 2.2 Introduction

Epidemics in pain (1) and in opioid-use disorder (2, 3) have inspired a search for nonopioid analgesics (1, 4). The  $\alpha_{2A}$ -adrenergic receptor ( $\alpha_{2A}$ AR) is a nonopioid receptor targeted by dexmedetomidine, a sedative that also has strong analgesic activity (5). Although dexmedetomidine has many advantages in emergency room and intensive care settings, its strong sedative effects (6, 7) and its lack of an oral formulation (8) have limited its broad use as an analgesic. These properties are barriers for future therapeutics targeting this receptor.

Most  $\alpha_{2A}$ AR analgesics are chemically related, and the relationship of their sedative to their analgesic properties is unclear. To find therapeutics with new pharmacology, we sought new  $\alpha_{2A}$ AR chemotypes, topologically unrelated to known  $\alpha_{2A}$ AR agonists. The  $\alpha_{2B}$ -adrenergic receptor ( $\alpha_{2B}$ AR) active-state structure (9) became available, and its binding site is highly conserved compared with that of  $\alpha_{2A}$ AR (**Fig. 2.S1**); therefore, it should be possible to identify new  $\alpha_{2A}$ AR agonists by structure-based docking. Meanwhile, the advent of readily accessible make-on-demand (“tangible”) molecules (10–12) ranging from hundreds of millions (10, 13, 14) to more than a billion molecules (15, 16) has vastly increased the chemotypes available for ligand discovery. Docking these libraries has revealed new chemotypes with 20 to 60% hit rates (13, 14, 17–20) and sometimes nanomolar potencies for a growing range of targets (10, 13, 14, 18, 21–24), often with new pharmacology (10, 13, 17, 25). Therefore, we targeted the  $\alpha_{2B}$ AR with an ultralarge library docking screen.

## 2.3 Results

### Docking 301 million molecules versus the $\alpha_{2B}AR$

The ZINC15 and ZINC20 virtual libraries are comprised of millions to billions of tangible molecules, depending on the molecular property range targeted, and are accessed by combining hundreds of thousands of diverse building blocks through hundreds of well-characterized reactions (10–12). Most of the molecules have not previously been synthesized and range in mass, calculated LogP (cLogP) values (a measure of hydrophobicity), and formal charge. Given the small size of the  $\alpha_{2B}AR$  orthosteric site, we docked both the 20 million fragment-like [compounds with smaller masses of < 250 atomic mass units (amu)] and 281 million lead-like (compounds with larger masses of 250 to 350 amu) molecules from the ZINC15 library (both sets having cLogP  $\leq$  3.5) (11) (Fig. 2.1A). More than 233 trillion complexes, an average of 452,000 per molecule, were sampled by DOCK3.7 and scored with its physics-based energy function (26) across three separate screens (two fragment screens with different variables and one lead-like screen; see Materials and methods). For each screen, the top 300,000 docking-ranked compounds were clustered for topological similarity and then filtered to identify scaffolds dissimilar to known agonists using an extended connectivity fingerprint (ECFP4). These agonists were drawn from the International Union of Basic and Clinical Pharmacology (IUPHAR)-British Pharmacological Society (BPS) database (27) and from the literature (28–31). Ligands with internal torsional strain were deprioritized (32). An additional novelty filter removed molecules similar to annotated  $\alpha_{2A}AR$  compounds in ChEMBL29 (28). Of the remaining top-ranked cluster representatives, 5,000 from each fragment screen and 20,000 for the lead-like screen were manually evaluated in UCSF Chimera (<https://rbvi.ucsf.edu/chimera>) for key polar and nonpolar interactions with  $\alpha_{2B}AR$  (9), including with D92<sup>3.32</sup>, F412<sup>7.39</sup>, F387<sup>6.51</sup>, Y391<sup>6.55</sup>, and F388<sup>6.52</sup> [residues conserved in  $\alpha_{2A}AR$ :

D128<sup>3.32</sup>, F427<sup>7.39</sup>, F405<sup>6.51</sup>, Y409<sup>6.55</sup>, and F406<sup>6.52</sup>; superscripts use Ballesteros-Weinstein and the G Protein-Coupled Receptor database (GPCRdb) nomenclature (33)]. Most  $\alpha_{2A}$ AR agonists, and certainly the clinically used dexmedetomidine and clonidine, are fragments (27), and the docking results reflected this. The docked fragment molecules fit in the orthosteric site, making key contacts with the receptor, whereas molecules in the lead-like screen generally did not fit in the small cavity (**Fig. 2.1A**). Accordingly, most selected ligands came from the fragment docking screens.

From the 64 high-ranking docked compounds prioritized by visual inspection and purchased for *in vitro* testing, 48 were successfully synthesized – 44 fragments and 4 lead-like molecules. Compounds were first tested for binding to the human  $\alpha_{2B}$ AR receptor, the structure used in docking screens. Thirty molecules of the 48 tested had binding constant ( $K_i$ ) values less than 10  $\mu$ M (**Table 2.S1**). This 63% hit rate is among the highest for a docking campaign to date (10, 14, 21, 23, 34). In radioligand competition assays, compound **ZINC1173879087** (from here on referred to as ‘**9087**) had a  $K_i$  of 1.7 nM; the remaining 29 had  $K_i$  values ranging from 60 nM to 9.4  $\mu$ M, which is relatively potent for initial docking hits. Ten compounds (21%) had  $K_i$  values below 1  $\mu$ M (**Table 2.S1**). The compounds were then tested for binding to the murine  $\alpha_{2A}$ AR, again by radioligand competition. Of these, 17 (35%) had a  $K_i$  better than 10  $\mu$ M, with affinities ranging from 72 nM to 9.4  $\mu$ M; five compounds (10%) had  $K_i$  values below 1  $\mu$ M (**Table 2.S1**). Against human  $\alpha_{2A}$ AR, the highest affinity was 12 nM (**Table 2.S2**).

### Discovery of new $\alpha_{2A}AR$ , partial $G_i$ and $G_o$ ( $G_{i/o}$ ) agonists

In functional assays most of the potent binders were partial or full agonists for  $\alpha_{2A}AR$  and  $\alpha_{2B}AR$  (Fig. 2.1B-D, Table 2.S1, Table 2.S2, Fig. 2.S2, Fig. 2.S3, Fig. 2.S4); few antagonists were found among the more potent docking hits. This reflects the targeting of the activated state of the receptor (35, 36) and was a goal of the screen. The best four agonists from the docking screen include ‘9087 as well as ZINC1240664622, ZINC1242282998, and ZINC001242890172 (from here on referred to as ‘4622, ‘2998, and ‘0172, respectively), with the  $\alpha_{2A}AR$ -mediated  $G_i$  activation maximum effect ( $E_{max}$ ) ranging from 60 to 95% of norepinephrine response and mean effective concentration ( $EC_{50}$ ) values of 9.7 to 210 nM in  $G\alpha_i$  bioluminescence resonance energy transfer (BRET) assays (Fig. 2.1C, Table 2.S2). We tested the effect of receptor expression in cells on the functional properties of the partial agonist, ‘9087, and ultimately of two optimized analogs, ‘7075 and PS75; all three remained potent  $G_i$  partial agonists, with  $E_{max}$  decreasing with receptor expression (Fig. 2.S5). In an orthogonal cyclic adenosine monophosphate (cAMP) assay, ‘9087 was a partial agonist with an  $EC_{50}$  of 87 nM and  $E_{max}$  of 42%, which is broadly consistent with the BRET assay (from here on  $G_i$  activities are the  $G\alpha_i$  BRET assay values unless otherwise noted) (Fig. 2.1C, Fig. 2.S4, Table 2.S2).

The docking agonists had strong differential activity for  $G_i$  activation compared with recruitment of  $\beta$ -arrestin-2. Although this was also true of the known agonists, dexmedetomidine and clonidine, for the new agonists the difference was accentuated so that arrestin recruitment was almost completely eliminated at relevant concentrations. Of the four best docking agonists, only ‘0172 had a measurable efficacy for  $\beta$ -arrestin-2 recruitment, but even here only with 22% of the  $E_{max}$  of norepinephrine and with weak potency ( $EC_{50} = 1.7 \mu M$ ); for the other three,  $\beta$ -arrestin-2

recruitment was negligible (**Fig. 2.1C**, **Table 2.S2**). We note that this lack of arrestin recruitment could reflect the partial agonism of the new agonists combined with the weaker coupling of the arrestin pathway versus the well-coupled  $G_i$  pathway, as indicated by the differences in potency and efficacy of the reference agonists, dexmedetomidine and clonidine.

Agonists of  $\alpha_{2A}AR$ , including its endogenous ligand norepinephrine, also activate other G protein pathways (37). Accordingly, we used the enhanced bystander BRET (ebBRET)-based effector membrane translocation assay (EMTA) (38) to test **'9087** and its analogs, **'7075** and **PS75**, against a more expansive panel of G protein and  $\beta$ -arrestin subtypes. The docking compounds preferentially activated  $G_i$ ,  $G_o$ , and  $G_z$  ( $G_{i/o/z}$ ) signaling, whereas known agonists norepinephrine, dexmedetomidine, and brimonidine strongly activated multiple additional G proteins and  $\beta$ -arrestins (**Fig. 2.S6**, **Fig. 2.S7**, **Table 2.S3**, **Table 2.S4**). Receptor internalization after treatment with compound was also investigated by monitoring disappearance of  $\alpha_{2A}AR$ s from the plasma membrane ( $\alpha_{2A}AR$ -RlucII/rGFP-CAAX biosensor) and relocalization of the receptors in endosomes ( $\alpha_{2A}AR$ -RlucII/rGFP-FYVE biosensor) (39). Known agonists brimonidine and norepinephrine show comparable responses for both biosensors, whereas dexmedetomidine has about half of this response. Consistent with their absence of  $\beta$ -arrestin recruitment, we found no effect of **'9087**, **'7075**, and **PS75** on disappearance from the plasma membrane and marginal effect at the highest concentrations on endosomal relocalization (**Fig. 2.S8**). Although such functional selectivity was not explicitly modeled in the docking, it likely results from the new chemistry, which was explicitly required (13, 14, 17).

Comparing the new agonists with dexmedetomidine, clonidine, norepinephrine, and a previously described pharmacophore model for  $\alpha AR$  selective agonists (9), both similar and distinct features

emerge (**Fig. 2.1B**). The pharmacophore model for known agonists and the new docking compounds both have basic, nitrogen-containing rings. However, known agonists are dominated by imidazoles (unsaturated or partially saturated) whereas the docking compounds have diverse, nonimidazole rings. Both sets of compounds contain additional moieties off of a second aryl ring, typically two substituents for the known agonists; however, for the docking-derived compounds, these vary from bulky hydrophobic rings, to hydrophilic rings, to single substituents, to having no substituents off of the aryl ring at all. Not all of the docking compounds have an exocyclic linker as described in the pharmacophore model. The protonated imidazole of known agonists ion pairs with D92<sup>3.32</sup> and hydrogen bonds to the backbone of F412<sup>7.39</sup> of  $\alpha_2\text{BAR}$  (*9, 40*) (**Fig. 2.S1**). Although several of the docking-derived compounds also interacted with both D92<sup>3.32</sup> and F412<sup>7.39</sup>, they did so with different heterocyclic rings (**Fig. 2.1D**).

To test the docking model and to template structure-based optimization, we determined the structure of the ‘**9087**- $\alpha_2\text{AAR}$ -G<sub>oA</sub> and ‘**4622**- $\alpha_2\text{AAR}$ -G<sub>oA</sub> complexes at a nominal resolution of 3.47 and 3.38 Å, respectively, using single particle cryo-electron microscopy (cryo-EM) (**Fig. 2.2A-D, Fig. 2.S9, Fig. 2.S10**). The predicted docked pose superimposes on the cryo-EM result of ‘**9087** with a 1.14 Å all-atom root mean square deviation (RMSD) of the agonist; the docking-predicted interactions are recapitulated in the experimental structure (**Fig. 2.2B**). The interactions between ‘**9087** and  $\alpha_2\text{AAR}$  differ from that of norepinephrine, but resemble those of imidazoline-containing agonists (*9, 40*). ‘**9087** interacts with  $\alpha_2\text{AAR}$  mainly through van der Waals and aromatic interactions to transmembrane helices (TM) 3, 5, 6 and 7 and I205<sup>45.52</sup> of extracellular loop 2 (ECL2). It also forms an ionic interaction with the conserved D128<sup>3.32</sup>, and although this interaction is relatively distant at 3.6 Å, it is similar to those of norepinephrine (*40*) and

dexmedetomidine (9) that are 3.0 and 3.7 Å from D128<sup>3.32</sup>, respectively (**Fig. 2.2B**). As in the docking prediction, the basic, formally cationic nitrogen of ‘9087 is not oriented toward D128<sup>3.32</sup> to form a salt bridge (**Fig. 2.2B, Fig. 2.S11**), as seen in norepinephrine, but instead hydrogen bonds with the backbone carbonyl of F427<sup>7.39</sup>, as does dexmedetomidine (9). The bridging exocyclic and formally neutral amine of ‘9087 ion pairs with D128<sup>3.32</sup>. Typically for aminergic G protein-coupled receptors (GPCRs), the conserved hydrogen bond with D128<sup>3.32</sup> would be made by the stronger base (9, 17, 20, 40). In fact, the formal charge of ‘9087 after protonation of the pyridine moiety is almost equally shared between the two nitrogens, as calculated by semi-empirical quantum mechanics and as reflected in the docking model (**Fig. 2.S11**). For ‘4622, the docked pose is also in good agreement with the cryo-EM result with an all-atom RMSD of 1.14 Å; ‘4622 forms a 3.4 Å hydrogen bond to D128<sup>3.32</sup> and makes several hydrophobic interactions (**Fig. 2.2D**). Both ‘4622- and ‘9087-bound structures have similar receptor-G<sub>oA</sub> interfaces to other ligand-bound α<sub>2</sub>AR-G protein complexes (**Fig. 2.S9, Fig. 2.S10**).

The interactions observed in the ‘9087 and ‘4622 receptor complexes, and in the modeled pose of analog ‘7075, were tested by residue substitution for impacts on G<sub>i</sub> activation and β-arrestin recruitment (**Fig. 2.S12, Fig. 2.S13, Table 2.S5, Table 2.S6**). Consistent with the observed ion pair with D128<sup>3.32</sup>, norepinephrine and dexmedetomidine are highly sensitive to substitutions to D128<sup>3.32</sup>, with an almost complete loss of G<sub>i</sub> activation and β-arrestin recruitment. For dexmedetomidine, the G<sub>i</sub> EC<sub>50</sub> is 170,000-fold higher for activation of D128<sup>3.32</sup>A. Although the G<sub>i</sub> activity of ‘9087 and ‘7075 is also diminished in the D128<sup>3.32</sup> mutant receptors, potency only falls by ~200- to 1600-fold. By contrast, the G<sub>i</sub> activity of ‘9087, ‘7075 and ‘4622 is eliminated in the F427<sup>7.39</sup>A mutant. The backbone carbonyl of F427<sup>7.39</sup> hydrogen-bonds with ‘9087, while its aromatic side chain stacks with the agonist in the cryo-EM structure, perhaps indicating formation



of a cation- $\pi$  interaction between the pyridine of **'9087** and F427<sup>7,39</sup> as previously suggested for agonist-induced  $\alpha_2$ AR activation (41). Meanwhile, dexmedetomidine and especially norepinephrine, which lack these interactions, are less affected by this mutant (**Fig. 2.2B**). Mutations of Y409<sup>6,55</sup> greatly affect norepinephrine, increasing (weakening)  $G_i$   $EC_{50}$  values 500- to 10,000-fold, likely disrupting a key hydrogen bond (40, 41); the importance of position 6.55 was previously observed in agonist-induced  $\alpha_2$ AR activation (42–44). The potencies of **'9087** and dexmedetomidine are only modestly worse in Y409<sup>6,55</sup> mutants, and for **'9087**, the  $G_i$   $E_{max}$  is even slightly increased. For **'4622** and **'7075**, most substitutions diminished activity, with the exception of S215<sup>5,42</sup>A, which slightly increased the agonist activity of **'7075** and **'4622** in the  $G_i$  activation and  $\beta$ -arrestin recruitment assays and hardly influenced **'9087** and dexmedetomidine. By contrast, S215<sup>5,42</sup>A negatively affected norepinephrine-induced receptor activation, consistent with previous reports on direct interactions of full agonists and S215<sup>5,42</sup> (40, 41). The Y431<sup>7,43</sup>A and F mutations overall influenced  $\beta$ -arrestin recruitment of norepinephrine more than  $G_i$  signaling. This has been previously observed, leading to the proposal that direct hydrogen bonding between the agonist and the residue at position 7.43 could more tightly couple TM7 and thereby play a role in  $\beta$ -arrestin signaling (40). Taken together, the differential responses to the residue substitutions supports suggestions from the structures that the new agonists, although binding in the same overall site as the canonical agonists, interact in meaningfully different ways, with potential implications for differential receptor signaling.

### **Newly discovered $\alpha_{2A}$ AR agonists are analgesic with reduced side effects**

In preparation for *in vivo* studies, we investigated the selectivity and pharmacokinetic properties of our most potent agonists. **'9087** activated only a few of 320 GPCRs screened (45) (**Fig. 2.S16A**).

Only the dopamine D<sub>2</sub> receptor (D2R) had weak activity in secondary assays, with EC<sub>50</sub> values of 4.5 μM and 16 μM in G protein signaling and β-arrestin recruitment, respectively (**Fig. 2.S16B**). **'9087** did not measurably activate the μ-opioid receptor (μOR) nor did it inhibit the human ether-à-go-go-related gene (hERG) at concentrations below 10 μM (**Fig. 2.S16C-D**). In binding experiments to other adrenergic receptors, **'9087** bound to the α<sub>2C</sub>-subtype at mid-nM concentration and to other α<sub>1</sub>-subtypes in the 1-to-10-μM range (**Table 2.S8**). The molecule had no measurable binding for β-adrenergic receptors up to 10 μM. Against the imidazoline-2 receptor (I2R), a common off-target of α<sub>2A</sub>AR agonists, **'9087** bound with a K<sub>i</sub> of 300 nM, showing a modest sixfold selectivity for the α<sub>2A</sub>AR, whereas a few docking-derived compounds actually had higher affinities for I2R than for the α<sub>2A</sub>AR (**Fig. 2.S16E**).

Computational models suggested that **'9087**, **'4622**, **'7075** and **'2998** would all have good physiologic permeability, consistent with their small size, low topological polar surface area, and weakly basic character (**Table 2.S9**). Consistent with this prediction, in 10 mg/kg intraperitoneal (i.p.) injection in mice the first three compounds, especially, had high brain and cerebral spinal fluid (CSF) exposure, indicating the compounds are likely to reach centrally acting α<sub>2A</sub>ARs (**Table 2.S9**). **'9087** reached a similar maximum concentration (C<sub>max</sub>) in the CSF as did **'7075**, both of which were fourfold greater than the C<sub>max</sub> of **PS75**, and **'9087** had a 12- to 20-fold higher area under the concentration-time curve (AUC) in the CSF compared with either **'7075** and **PS75**; CSF concentrations are often used as a proxy for fraction unbound in the brain (46). Encouragingly, **'9087** reached high brain exposure after both intravenous (i.v.) and oral (p.o.) administration, with AUC values of 420,000 ng min mL<sup>-1</sup> and 2,540,000 ng min mL<sup>-1</sup> (**Table 2.S9**). The oral bioavailability was higher than 100%, which may reflect metabolic saturation at non-equal i.v. and

p.o. doses, or enterohepatic recirculation (**Table 2.S9**); this merits further investigation. **PS75** was fully bioavailable, though, as an analog of **'9087** the same caveats apply. We also investigated the metabolic stability of **'9087**, **'7075** and **PS75** in male rat liver microsomes. All three compounds remain largely unmodified after 1 hour, with **'9087** having lower clearance and a higher half-life than its two analogs (**Fig. 2.S17**).

Given their selectivity and high brain exposures, we tested the more potent agonists for pain-relief after systemic dosing (**Fig. 2.4A-G**). Initial doses were chosen to be less than the 10 mg/kg dose used in pharmacokinetics studies owing to favorable CSF and brain properties. We started with **'9087**, the initial docking hit, and evaluated analogs as they emerged from compound optimization. With naïve (uninjured) mice, **'9087** did not increase baseline mechanical withdrawal thresholds, something observed with many anti-pain medications, which often only have an antinociceptive effect in the presence of pain. We then investigated the activity of **'9087** in a mouse model of neuropathic pain, in which partial peripheral nerve injury invokes profound mechanical hypersensitivity (47). Systemic subcutaneous (s.c.) injections of **'9087** dose-dependently increased the mechanical thresholds of spared nerve injured (SNI) mice, with a sharp increase in activity from 3 mg/kg to 5 mg/kg, at which points the effects plateaued (**Fig. 2.4A**). Lower doses were antiallodynic, returning mechanical thresholds to preinjury levels, whereas the higher doses were genuinely analgesic, generating mechanical thresholds substantially higher than baseline, preinjury levels. **'9087** also increased thermal latencies in the complete Freund's adjuvant (CFA)-mediated inflammatory pain model, which indicates that the molecule is effective in both tissue and nerve injury-induced pain models (**Fig. 2.4G**). **'9087** also increased withdrawal latencies in the hot plate (55 °C) and tail flick (50 °C) assays of acute thermal (heat) pain (**Fig. 2.4E, Fig.**

**2.4F**). Consistent with its relatively high exposure on oral dosing, this molecule also conferred a dose-dependent anti-allodynic effect when delivered orally in the SNI neuropathic pain model (**Fig. 2.4A**). Doses up to 20 mg/kg of **'9087** did not reduce the ability of the mice to perform in the rotarod test, which contrasts with the complete sedation of a dexmedetomidine dose of 60  $\mu$ g/kg (**Fig. 2.4H**). This finding is an important differentiator for the new series and indicates that the analgesic effects of **'9087** are not the result of motor impairment.

We also investigated the mechanistic bases for the analgesia of the new  $\alpha_{2A}$ AR agonists, both pharmacologically and genetically. Pharmacologically, the analgesic effect of **'9087** was reversed by a systemic injection of the well-known  $\alpha_2$ AR antagonist, atipamezole (2 mg/kg; administered 15 minutes prior to **'9087**) (**Fig. 2.4C**). Because atipamezole has broad activity against the  $\alpha_2$ AR receptor subtypes and imidazoline receptors (48), we also tested **'9087** in mice that express an inactive form of the  $\alpha_{2A}$ AR (point mutation D79N) (5, 49–51). D79N mutant mice were tested in the tail flick (50 °C) assays. As previously reported, dexmedetomidine no longer induced analgesia in the mutant mice (52), and, as a control, the analgesia conferred by a 10 mg/kg dose of morphine was not significantly altered by the mutation (**Fig. 2.4D**). Consistent with activity through  $\alpha_{2A}$ AR, the analgesia conferred by **'9087** was reduced >50% back to baseline in the D79N mutant mice (**Fig. 2.4D**). Although most of the antinociceptive activity appears to derive from activity at the  $\alpha_{2A}$ AR, we cannot discount contributions from other receptors.

Five other of the new, docking-derived  $\alpha_{2A}$ AR agonists (**'2998**, **'4622**, **'0172**, **'7075**, **PS75**) also exhibited antiallodynic effects in the SNI mice (**Fig. 2.4A-B**). The **'9087** analogs, **'7075** and **PS75**, completely reversed the mechanical hypersensitivity in the neuropathic pain model, with **PS75**

being more effective than **'9087** (Fig. 2.4A). In contrast to **'9087**, **PS75** did increase the mechanical thresholds of naïve (uninjured) mice (Fig. 2.4A). The anti-allodynic effects of **'4622** and **'7075** were reversed by atipamezole (48); although this antagonist also partially reversed the anti-allodynia of **PS75**, **'2998**, and **'0172**, these effects did not reach statistical significance at the small numbers of mice tested (Fig. 2.4C). **PS75** also increased withdrawal latencies in the tail flick (50 °C) acute thermal pain assay, and when tested in the D79N mutant mice, its analgesic effect was reduced by over 50% (Fig. 2.4D-E). Compounds **'0172** and **'4622** also exhibited antihyperalgesic effects in the CFA inflammatory pain model (Fig. 2.4G); **'2998** did not, which may reflect the reduced brain penetration of this molecule (Table 2.S9). Only **'4622** caused slight motor impairment at its equianalgesic dose in the rotarod test; however, the effect did not reach the full sedation observed with dexmedetomidine (Fig. 2.4H). As with **'9087**, increased dosing up to 20 mg/kg **PS75** did not have an effect on the rotarod test (Fig. 2.4H). Taken together, these pharmacological and chemical-genetic epistasis experiments support a mechanism of action primarily through the  $\alpha_{2A}$ AR, though a lesser contribution of other  $\alpha_2$ AR subtypes cannot be ruled out.

Some  $\alpha_{2A}$ AR agonists can produce changes in feeding, weight gain, and hyperglycemia (53, 54) as side effects. Accordingly, we evaluated the effect of compound treatment on body weight over 48 hours after injection while allowing the mice to freely feed. We found no effect on body weight for **'9087** dosed at 5 mg/kg, 10 mg/kg or 20 mg/kg, nor for dexmedetomidine dosed at 30  $\mu$ g/kg (Fig. 2.S18). We also investigated whether **'9087** induced constipation, a side effect well-known for opioids and other classes of analgesics, comparing it with dexmedetomidine and with morphine tested at analgesic doses (30  $\mu$ g/kg and 10 mg/kg, respectively). The number of accumulated pellets over 6 hours following vehicle or compound injection was measured. As expected,

morphine induced constipation when compared to vehicle at the 1-, 2-, and 3-hour marks. By contrast, although pellet number did decrease modestly with ‘9087 and dexmedetomidine, neither effect differed significantly from vehicle (**Fig. 2.S18**). We recognize that other possible side effects remain untested in this study; we return to these in the Discussion section.

## 2.4 Discussion

Three key observations emerge from this study. **First**, multiple chemotypes, unrelated to known agonists, discovered directly from large-library docking are efficacious in neuropathic, inflammatory, and acute pain models through  $\alpha_{2A}$ AR agonism (**Fig. 2.4A-F**). In docking, as in other target-based screens, the initial goal is to identify molecules with *in vitro* activity; these are then optimized for *in vivo* activity through extensive structure-activity optimization (13, 14, 17, 20). Although it may be rare that direct hits from a docking screen are themselves *in vivo* active, such activity of the direct docking hits here does speak to the strengths of interrogating vast virtual libraries (11, 12). **Second**, functional assays reveal preferential  $G_{i/o/z}$  activation versus other G protein subtypes, no  $\beta$ -arrestin activity and no receptor internalization for ‘9087 and its analogs, ‘7075 and PS75, compared to the established therapeutic  $\alpha_{2A}$ AR agonists, like dexmedetomidine and brimonidine, that also activate multiple other G protein and  $\beta$ -arrestin signaling pathways (**Fig. 2.1C, Fig. 2.3B, Table 2.S2, Fig. 2.S6, Fig. 2.S7, Fig. 2.S8**). Although the high rate of agonist discovery was an intended outcome of docking against the activated state of the  $\alpha_{2B}$ AR, the functional selectivity was not designed and can be attributed to the novel chemotypes they explore and, by extension, their use of both canonical and non-canonical receptor interactions (**Fig. 2.2, Fig. 2.S12, Fig. 2.S13**). There is no logical requirement that novel chemotypes lead to new signaling pharmacology, though this has often observed for other receptors (10, 13, 17, 25). **Third**, unlike dexmedetomidine and clonidine (6, 7, 55), ‘9087 and its analogs, PS75 and ‘7075, do not cause sedation or motor impairment at analgesic doses, potentially enabling broader applications to pain treatment and attesting to the ability to differentiate these two effects with  $\alpha_{2A}$ AR agonists (**Fig. 2.4H**).

The new  $\alpha_{2A}$ AR chemotypes explored (**Fig. 2.1B**) reflect the size and diversity of the docked libraries. Most of the actives emerged from the fragment-like library in ZINC, which covers a much greater portion of the chemical universe in its size range than do the lead-like or drug-like libraries. This is akin to physical fragment libraries, which typically might include ~1,500 molecules (56) but are thought to cover more chemotypes than high-throughput screening libraries that are 1,000-fold larger. Meanwhile, the virtual fragment library in ZINC enumerates over 20 million molecules (11, 12), about 10,000-fold more than in most physical fragment libraries. Indeed, with over 800,000 Bemis-Murcko scaffolds (18), there are 500-fold more fragment scaffolds in the docked library than there are molecules in most physical fragment libraries. From this great chemotype diversity springs opportunities for ligands with new pharmacology.

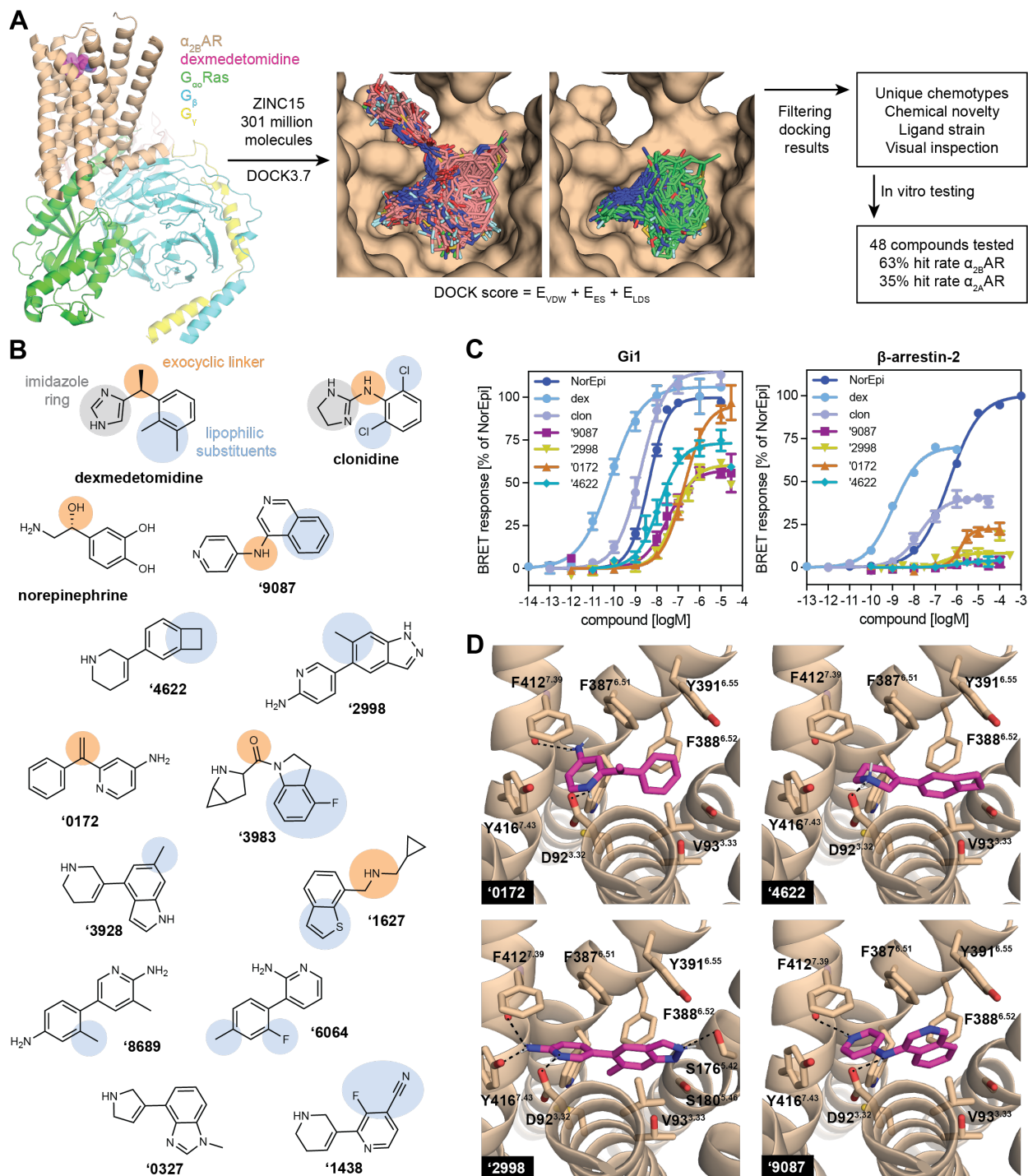
How the new chemotypes confer new pharmacology is uncertain. From the cryo-EM co-complex with '9087 and substitution of binding site residues, '9087 and '7075 appear to make weaker interactions with D128<sup>3.32</sup>, and an apparently stronger interaction with F427<sup>7.39</sup>, than do the canonical agonists (**Fig. 2.2B**, **Fig. 2.S12**, **Fig. 2.S13**). These structural differences may contribute to the unusual functional  $G_{i/o/z}$  selectivity over other G protein subtypes and  $\beta$ -arrestins as compared to known agonists, and to the lack of receptor internalization (**Fig. 2.3B**, **Fig. 2.S6**, **Fig. 2.S7**, **Fig. 2.S8**). In turn, this unusual signaling may play a role in conferring analgesia without sedation (**Fig. 2.4**). Admittedly, as the engagement of transducing G proteins and  $\beta$ -arrestins occurs 35 Å away from the orthosteric site, other mechanisms may be involved (57). Moreover, the physiological impact of the selective signaling will be entangled with the pharmacokinetics of the molecules. Regardless, what should be clear is that the analgesic potential of  $\alpha_{2A}$ AR agonists may be disentangled from their sedative effect, which is important for future drug development.



Several cautions merit re-emphasis. **'9087** and its analogs are not as potent as dexmedetomidine. The action of **'9087** *in vivo* is blocked by the  $\alpha_2$ AR antagonist atipamezole (48) and much reduced in D79N  $\alpha_{2A}$ AR mice, which indicates the  $\alpha_{2A}$ AR is the primary receptor mediating activity *in vivo* (Fig. 2.4C-D). However, especially for the mutant mice, we note that while most efficacy above baseline was reduced, it was not fully reversed (Fig. 2.4C-D), and other targets may also play a role, including the I2R and other  $\alpha_2$ AR subtypes. We also have not extensively evaluated side effects common for  $\alpha_{2A}$ AR agonists, especially cardiovascular effects mediated by  $\alpha_{2A}$ AR and  $\alpha_{2B}$ AR activation (6). Finally, we do not anticipate that the ability to translate directly from docking hits to *in vivo* activity, as we saw in this study, will be common. Here it was helped by the small size of the  $\alpha_{2B}$ AR/ $\alpha_{2A}$ AR orthosteric sites and correspondingly the high potencies and good physical properties of the docking hits (Fig. 2.S1, Table 2.S2, Table 2.S9).

These caveats should not obscure the key observations of this study. From an ultralarge library docking screen emerged low-nanomolar  $\alpha_{2A}$ AR partial agonists, topologically unrelated to previously known ligands, making new interactions with the receptor that appear to confer new pharmacology (Fig. 2.1, Fig. 2.2B, Table 2.S2, Fig. 2.S6, Fig. 2.S7, Fig. 2.S8, Fig. 2.S12, Fig. 2.S13). Several of the new agonists were antiallodynic and analgesic in neuropathic and inflammatory pain models, and against acute nociception in naïve animals (Fig. 2.4A-G). Among the most promising are **'9087** and **PS75**, both of which are strongly analgesic without the sedative effects of dexmedetomidine (Fig. 2.4) and are orally bioavailable (Table 2.S9). These properties make the compounds plausible therapeutic leads for new nonopioid pain therapeutics without the sedation of classic  $\alpha_{2A}$ AR drugs.

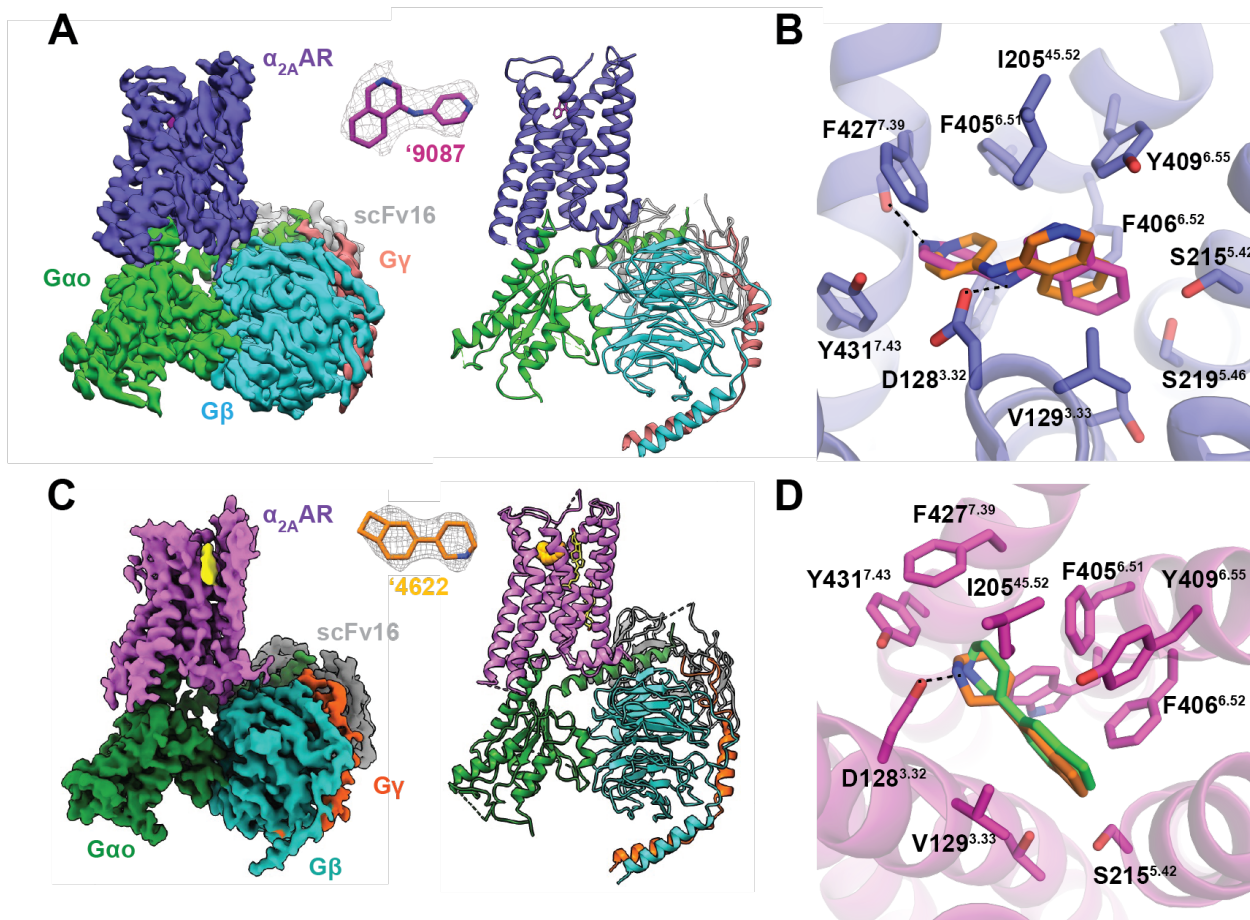
## 2.5 Figures



**Figure 2.1 Newly discovered  $\alpha_{2A}AR$  agonists from ultralarge library docking.**

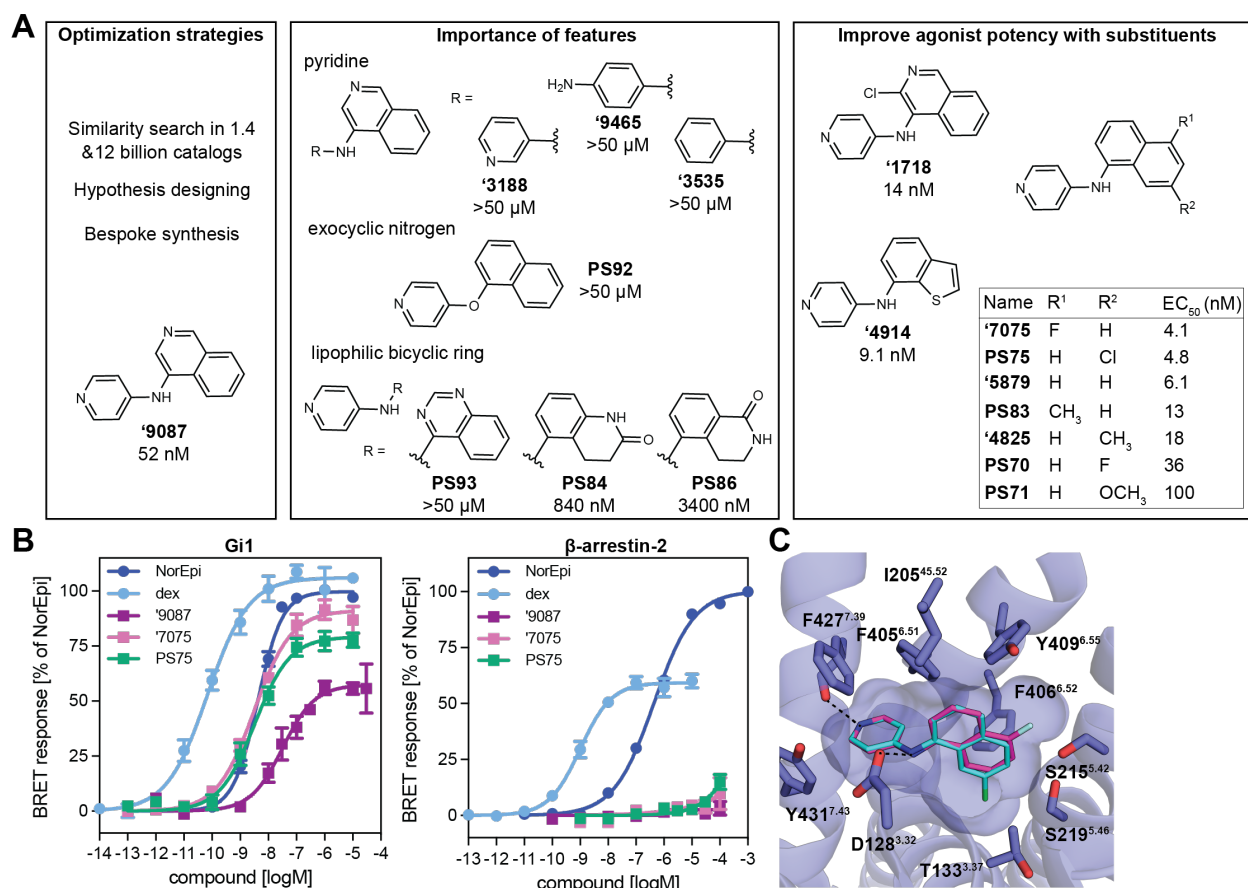
(A) 301 million molecules were docked against the active state of  $\alpha_{2B}AR$  (PDB 6K41). Lead-like molecules (pink carbons) often spilled out of the orthosteric site, whereas fragment molecules (green carbons) are well complemented by that site. Hit rates were determined with a  $K_i$  cutoff of

10  $\mu$ M.  $E_{VDW}$ , van der Waals;  $E_{ES}$ , electrostatic;  $E_{LDS}$ , ligand desolvation. **(B)** The  $\alpha$ AR pharmacophore model (9) overlaid on known  $\alpha_{2A}$ AR agonists dexmedetomidine, clonidine, and norepinephrine and new agonists from docking (colors represent the different moieties fulfilling the same role). **(C)**  $G_i$  activation and  $\beta$ -arrestin-2 recruitment for norepinephrine (NorEpi), dexmedetomidine (dex), clonidine (clon), and several of the newly discovered docking agonists. **(D)** Docked poses of these new agonists with hydrogen bonds to key recognition residues of  $\alpha_{2B}$ AR shown as black dashed lines. For (C), data are mean  $\pm$  SEMs of normalized results (n=4 to 17 measurements for  $G_i$  and n=3 to 8 measurements for  $\beta$ -arrestin-2). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.



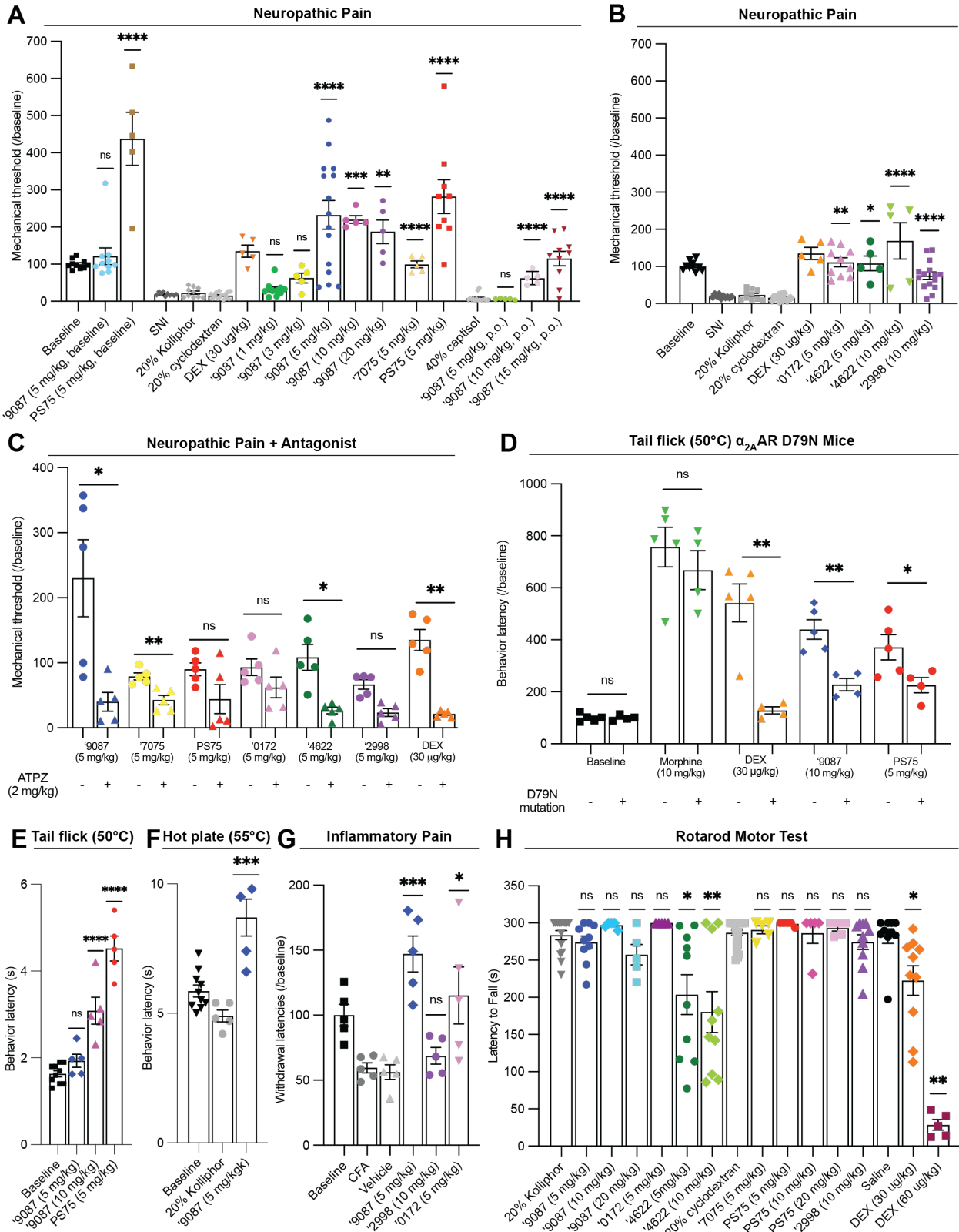
**Figure 2.2 Docking-predicted poses of '9087 and '4622 superpose well on the cryo-EM structures of the '9087- $\alpha_{2A}$ AR-GoA and '4622- $\alpha_{2A}$ AR-GoA complexes.**

(A and C) Cryo-EM structure of the '9087- $\alpha_{2A}$ AR-GoA (A) and '4622- $\alpha_{2A}$ AR-GoA (C) complexes. (B) Experimental '9087 structure (pink carbons) superposed on the docked pose (orange carbons) (PDB 7W6P; RMSD 1.14 Å). Hydrogen bonds and ion pairs are shown with dashed black lines to F427<sup>7.39</sup> and D128<sup>3.32</sup>, respectively. (D) Experimental '4622 structure (green carbons) superposed on the docked pose (orange carbons) (PDB 7W7E; RMSD 1.14 Å). Hydrogen bond shown with dashed black lines to D128<sup>3.32</sup>. For (B) and (D), sidechains of  $\alpha_{2A}$ AR residues within 4 Å of ligands are shown as sticks.



**Figure 2.3 Structure-based optimization of '9087.**

(A) Strategies for analoging '9087 (left). Analogs of the pyridine, exocyclic nitrogen, and lipophilic nature of the bicyclic ring revealed their importance for '9087 activity (middle). Sampling alternate lipophilic bicyclic rings and modifying their substituents identified eight more potent agonists (right). EC<sub>50</sub> values shown for G<sub>i</sub> activation. (B) G<sub>i</sub> and β-arrestin-2 recruitment for '9087 and its two most potent analogs, '7075 and PS75. (C) Modeled poses of '7075 (pink carbons) and PS75 (blue carbons) based on the '9087-α<sub>2A</sub>AR cryo-EM structure with substituents oriented towards open space in the orthosteric site. Hydrogen bonds and ionic interactions are shown with dashed black lines to F427<sup>7.39</sup> and D128<sup>3.32</sup>, respectively. For (A), G<sub>i</sub> and β-arrestin-2 recruitment data for analogs are shown in Fig. 2.S14, Fig. 2.S15 and Table 2.S7. For (B), data are mean ± SEMs of normalized results (n=7 to 17 measurements for G<sub>i</sub> and n=4 to 8 measurements for β-arrestin-2).



**Figure 2.4 The docking-derived agonists are antinociceptive in neuropathic, inflammatory, and acute thermal pain through the  $\alpha_{2A}$ AR but are not sedating.**

(A-C) Effect of new  $\alpha_{2A}$ AR agonists in neuropathic pain model in mice after SNI with mechanical allodynia.

(A) The new agonists **'9087** and **PS75** administered in naïve mice (baseline versus **'9087**, 5 mg/kg; baseline versus **PS75**, 5 mg/kg; one-way ANOVA; ns, not significant; \*\*\*\* $P < 0.0001$ ), dose response of **'9087** in SNI mice and analogs **'7075** and **PS75** compared with their vehicles (20% kolliphor versus all **'9087** doses; 20% cyclodextran versus **'7075** and **PS75**; one-way ANOVA; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ ) with positive control dexmedetomidine (DEX), and **'9087** administered orally (p.o.) compared with its vehicle (40% captisol vs **'9087** doses; one-way ANOVA; ns, not significant, \*\*\*\* $P < 0.0001$ ).

(B) Effect of additional agonists **'4622**, **'0172**, **'2998** compared to their vehicles (20% kolliphor versus **'4622**, 5 mg/kg; **'4622**, 10 mg/kg; and **'0172**, 5 mg/kg; one-way ANOVA; 20% cyclodextran versus **'2998**; two-tailed t-test; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ ) and positive control DEX.

(C) Administration of  $\alpha_{2A}$ AR antagonist atipamezole (ATPZ, 2 mg/kg i.p.) to block agonist efficacy in neuropathic pain model (**'9087** versus **'9087** with ATPZ; **'7075** versus **'7075** with ATPZ; **PS75** versus **PS75** with ATPZ; **'0172** versus **'0172** with ATPZ; **'4622** versus **'4622** with ATPZ; **'2998** versus **'2998** with ATPZ; DEX versus DEX with ATPZ; two-tailed t-test; \* $P < 0.05$ , \*\* $P < 0.01$ ).

(D) Diminished analgesia in  $\alpha_{2A}$ AR D79N mice in the 50 °C tail flick assay for acute thermal (heat) pain. The mutation does not affect morphine analgesia but substantially decreases the analgesia by DEX, **'9087**, and **PS75** (baseline WT versus D79N; morphine WT versus D79N; DEX WT versus D79N; **'9087** WT versus D79N; **PS75** WT versus D79N; two-tailed t-test; \* $P < 0.05$ , \*\* $P < 0.01$ ).

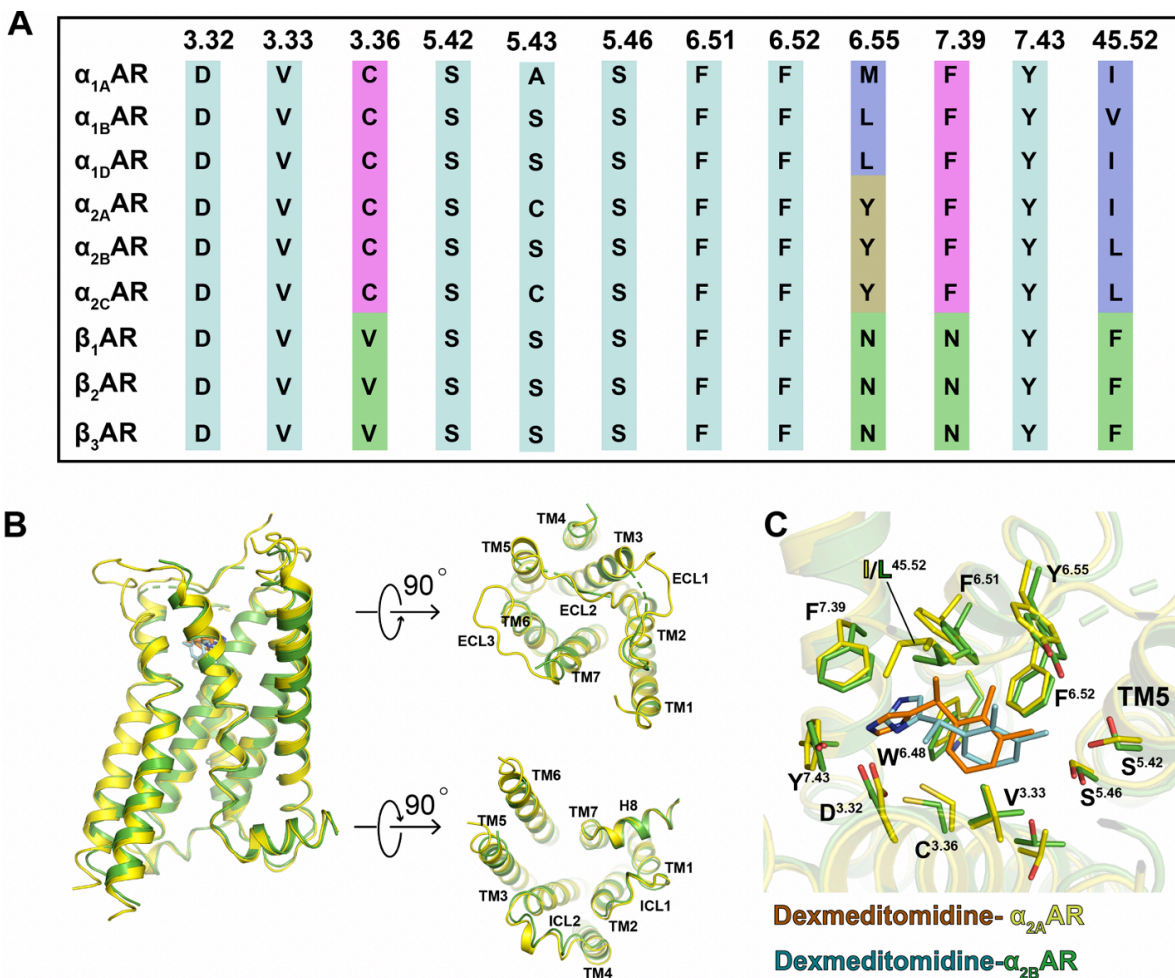
(E) Analgesia of **'9087** and **PS75** in 50 °C tail flick assay for acute thermal (heat) pain compared to its vehicle (20% Kolliphor versus **'9087** and **PS75**; one-way ANOVA; \*\*\*\* $P < 0.0001$ ).

(F) Analgesia of **'9087** in 55 °C hot plate assay for acute thermal (heat) pain compared to its vehicle (20% Kolliphor versus **'9087**; two-tailed t-test; \*\*\*  $p < 0.001$ ).

(G) Efficacy of newly characterized agonists in CFA-induced hyperalgesia compared to the vehicle (vehicle versus **'9087**, **'2998**, and **'0172**; one-way ANOVA; \* $P < 0.05$ , \*\*\* $P < 0.001$ ).

(H) Evaluating motor impairment and sedation of newly characterized agonists in the rotarod motor test. Only **'4622** causes slight motor impairment while other agonists do not. DEX causes significant impairment and complete sedation at higher doses. All compounds compared to their vehicles (20% Kolliphor versus **'9087**, **'0172**, **'4622**; 20% cyclodextran versus **'2998**, **'7075** and **PS75**; saline versus DEX; one-way ANOVA; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ ).

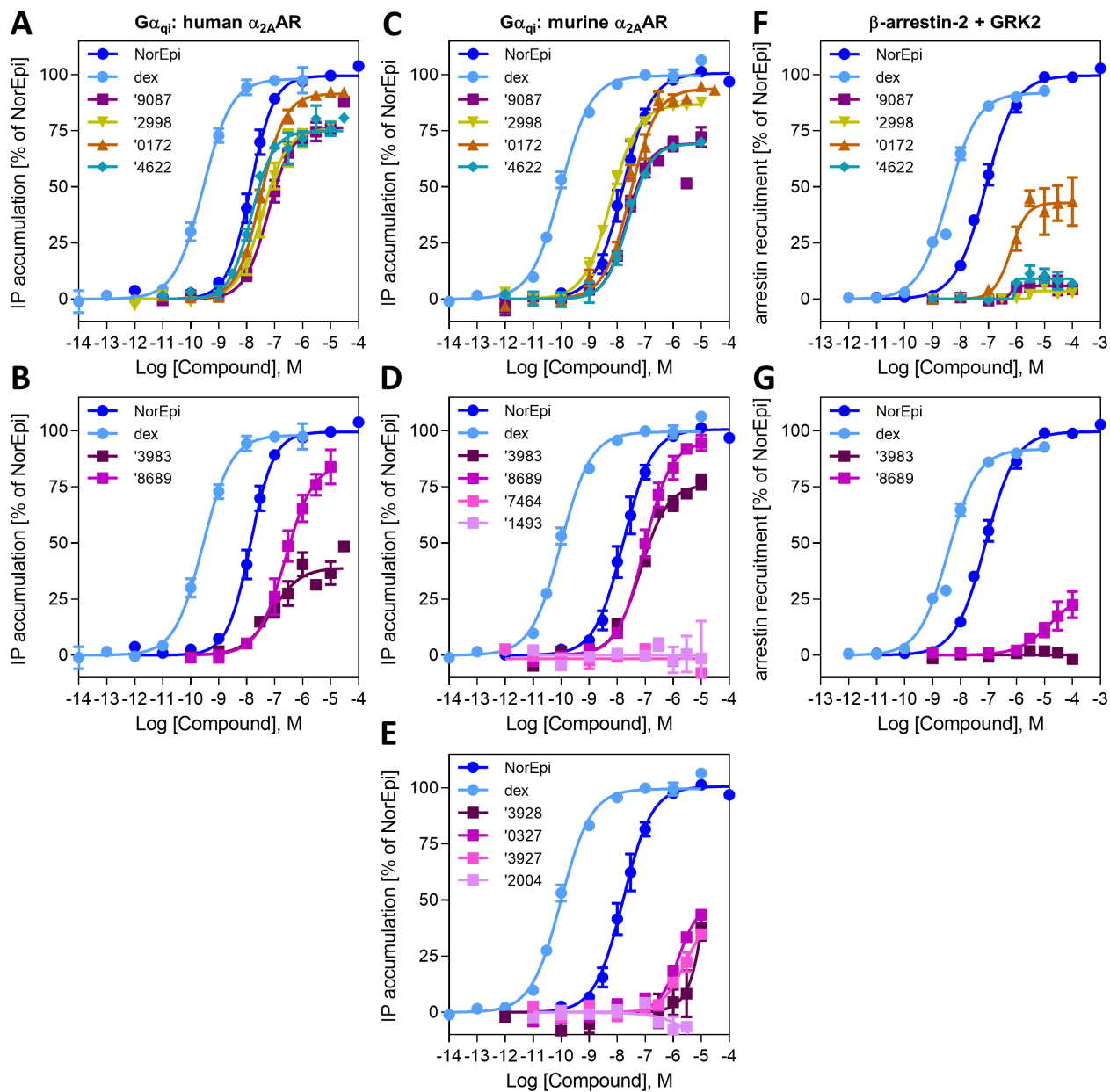
For (A-G), all compounds were administered s.c. unless otherwise indicated. Data are shown as individual data points and mean  $\pm$  SEMs (n=5 to 25 measurements).



**Figure 2.S1 Structural similarity of adrenergic receptors.**

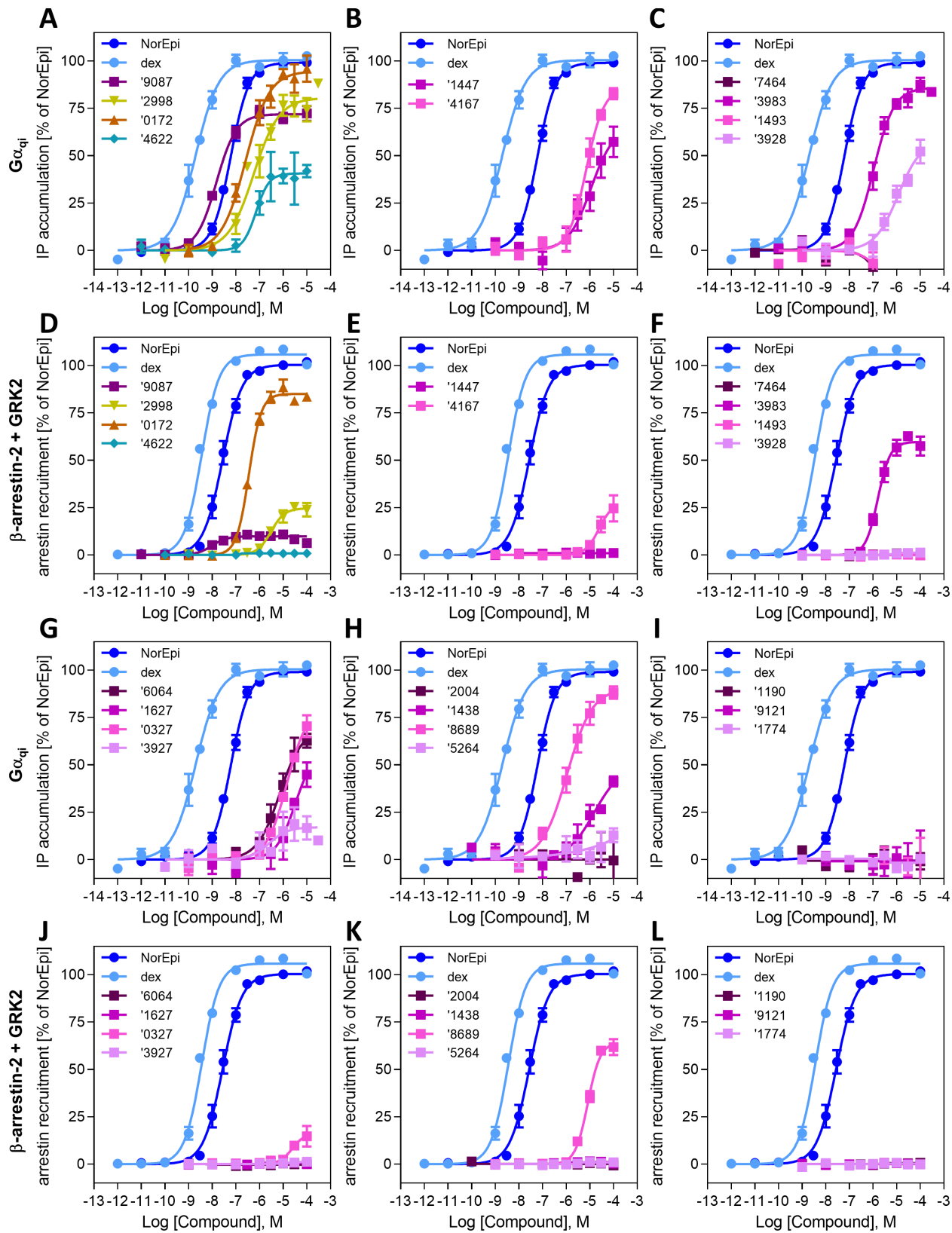
(A) Sequence alignment of defined orthosteric binding site residues of adrenergic receptors. (B) Structural similarities of  $\alpha_{2A}$ AR and  $\alpha_{2B}$ AR bound to dexmedetomidine (PDB 7EJA and 6K41, respectively) show high residue identity between receptor subtype and highly similar side chain positions when bound to dexmedetomidine. The  $\alpha_{2B}$ AR-dexmedetomidine structure was used for docking screens prior to the determination of the  $\alpha_{2A}$ AR-dexmedetomidine structure.





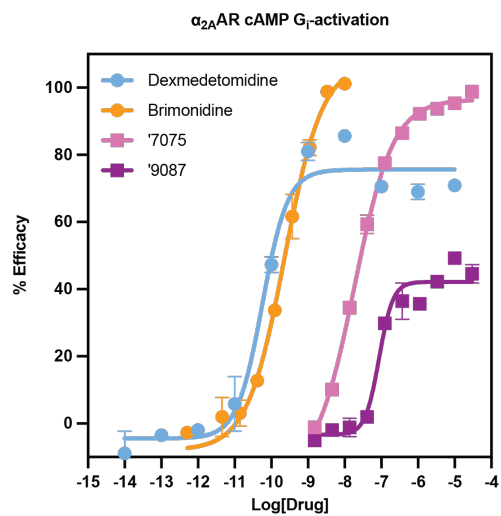
**Figure 2.S2 Functional data for docking hits against  $\alpha_{2A}AR$ .**

(A,B)  $G_i$  signaling for docking hits against human  $\alpha_{2A}AR$  in the IP-One assay. (C-E)  $G_i$  activation for docking hits against murine  $\alpha_{2A}AR$  in the IP-One assay. (F,G)  $\beta$ -arrestin-2 recruitment for docking hits against human  $\alpha_{2A}AR$  in the PathHunter assay. For (A-G), data are mean  $\pm$  s.e.m. of normalized results (n=3-11 measurements).



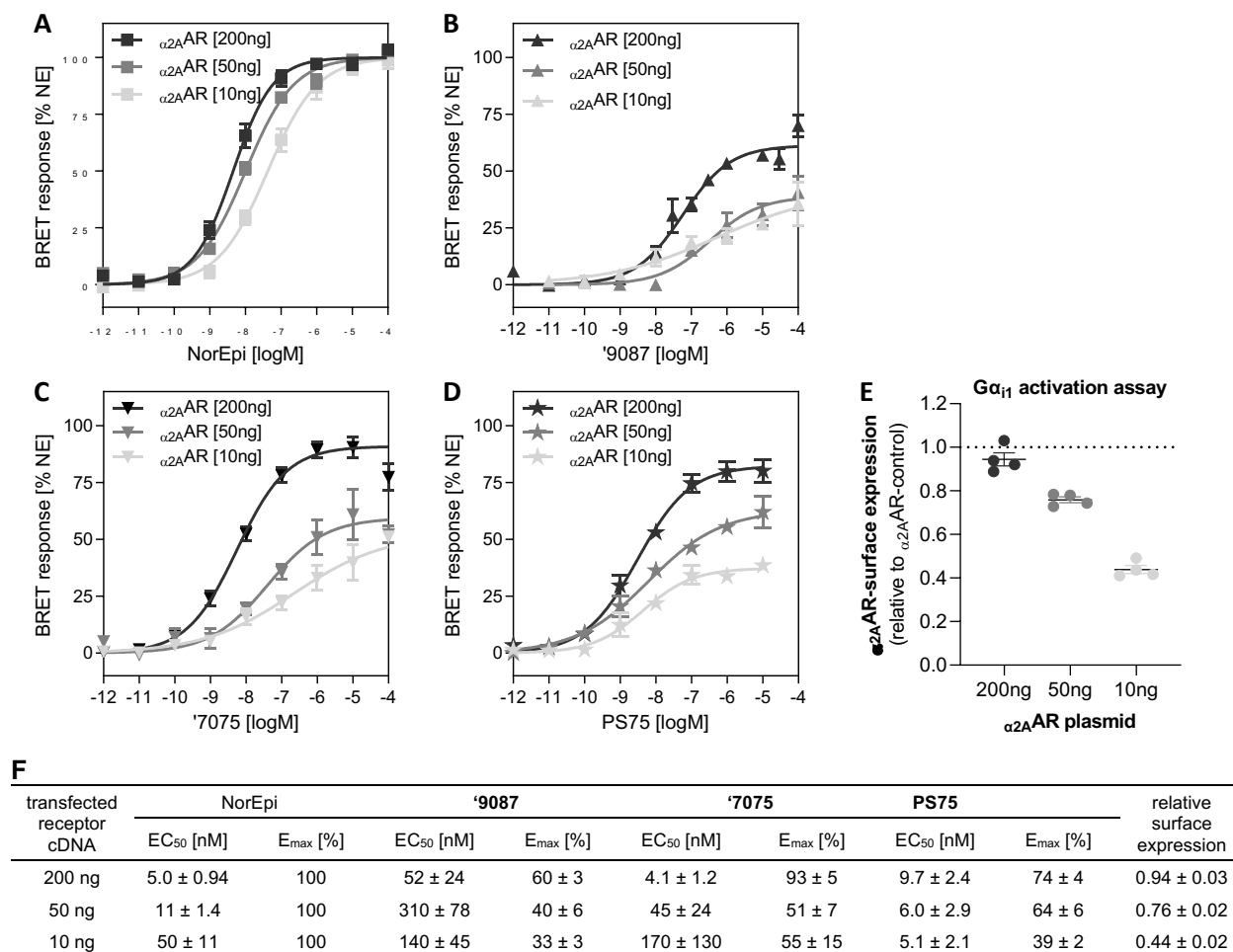
**Figure 2.S3 Functional data for docking hits against  $\alpha_{2B}AR$ .**

(**A-C, G-I**)  $G_i$  signaling for docking hits against human  $\alpha_{2B}AR$  in the IP-One assay. (**D-F, J-L**)  $\beta$ -arrestin-2 recruitment for docking hits against human  $\alpha_{2B}AR$  in the PathHunter assay. For (A-L), data are mean  $\pm$  s.e.m. of normalized results (n=3-9 measurements).



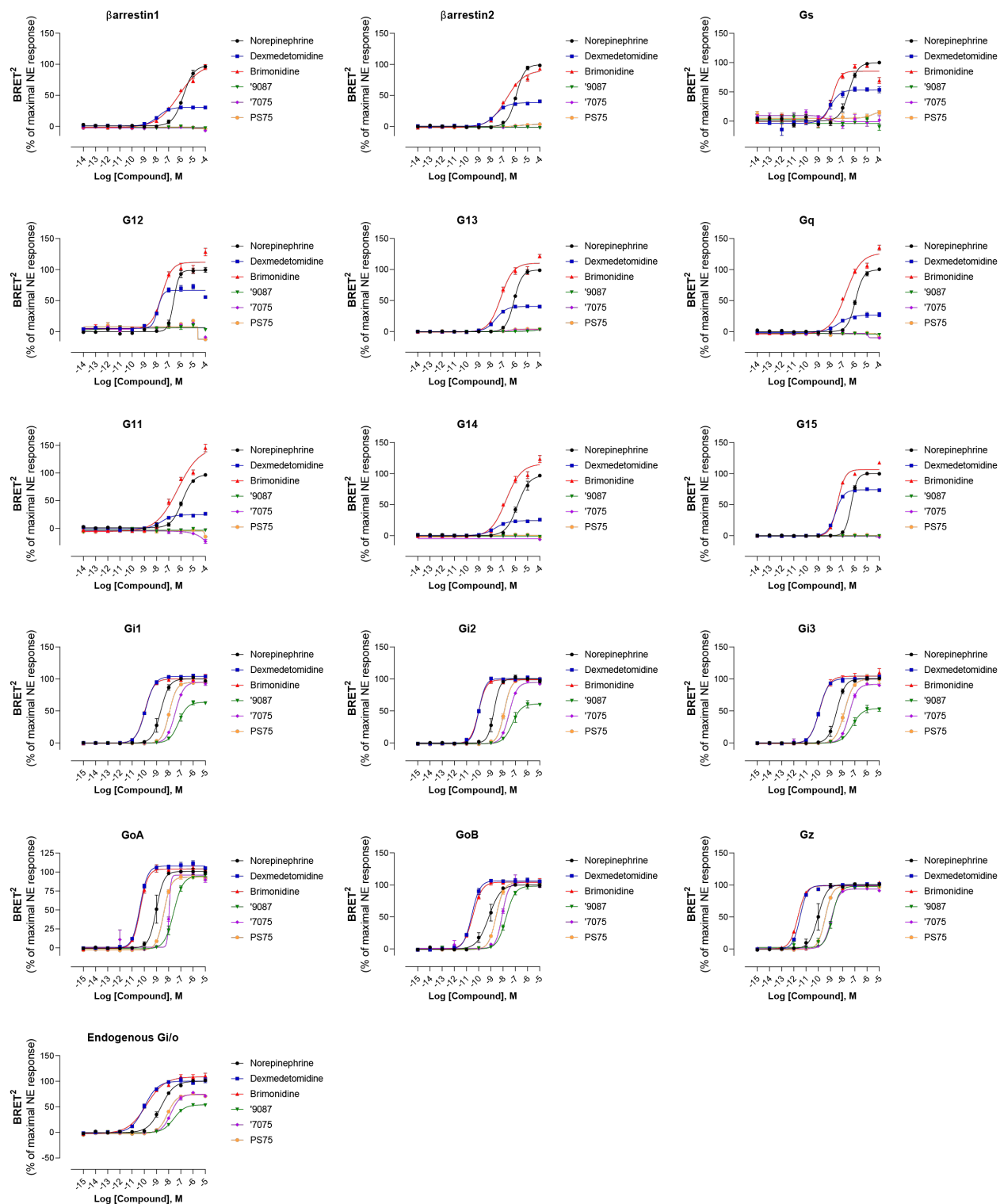
**Figure 2.S4 G<sub>i</sub>-activation induced cAMP inhibition assay against α<sub>2A</sub>AR.**

G<sub>i</sub> signaling for docking hits against human α<sub>2A</sub>AR in the DiscoverX HitHunter cAMP assay. Data are mean ± s.e.m. of normalized results (n=2 measurements).



**Figure 2.S5 Functional properties of norepinephrine, selected docking agonists, and the bespoke synthesized analog PS75 are dependent on receptor density.**

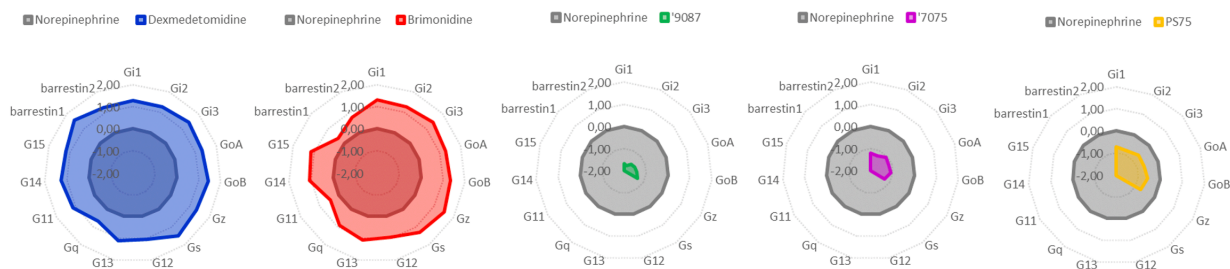
Comparison of EC<sub>50</sub> and E<sub>max</sub> values at the standard receptor density of 200 ng transfected cDNA and results derived at receptor expression at 50 ng and 10 ng of DNA for (A) norepinephrine (NorEpi), (B) '9087, (C) '7075, and (D) PS75 for α<sub>2A</sub>AR. (E) Relative surface expression of α<sub>2A</sub>AR coexpressed with G<sub>αi1</sub>-RlucII, G<sub>β1</sub> and G<sub>γ2</sub>-GFP10 determined by ELISA directed against the N-terminal FLAG-tag. Individual data points are shown relative to cells transfected with 200 ng α<sub>2A</sub>AR-plasmid, only. (F) summary of all activation data for compounds at different α<sub>2A</sub>AR receptor densities. For (A-D, F), all data shown are for G<sub>i</sub> activation monitored in a BRET-biosensor based assay with 5 to 18 experiments in duplicates; mean EC<sub>50</sub> values are displayed as in [nM ± s.e.m.]. Normalization was done relative to the maximum effect of norepinephrine (NE) and is displayed in [% ± s.e.m.]. For (E), data are mean ± s.e.m. of normalized results (n=4 of quadruplicates).



**Figure 2.S6 EMTA coupling panel for select docking compounds against  $\alpha_{2A}AR$ .**

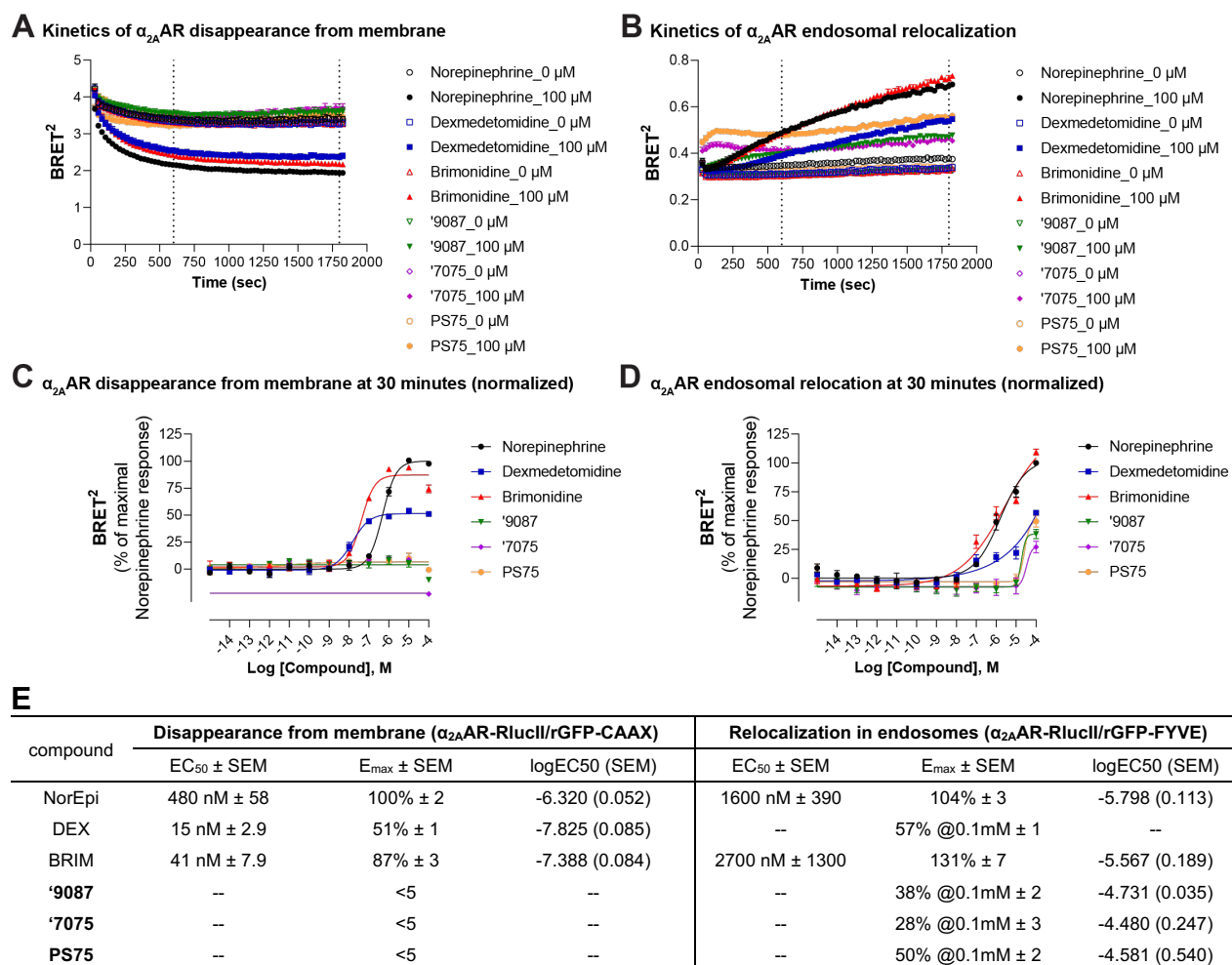
G-protein and  $\beta$ -Arrestin signaling profiles for docking compounds in BRET biosensor-based assays in HEK293 cells expressing the human  $\alpha_{2A}AR$ . Data are mean  $\pm$  s.e.m. of normalized results ( $n=3-5$  measurements). Endogenous  $G_{i/o}$  indicates activation of the protein family in the absence of heterologously expressed G proteins.

$\Delta\log(\text{Emax}/\text{EC50})$ ; Relative activity; (4-parameters fitted curves)



**Figure 2.S7 Relative activities for select docking compounds against  $\alpha_{2A}AR$  EMTA coupling panel.**

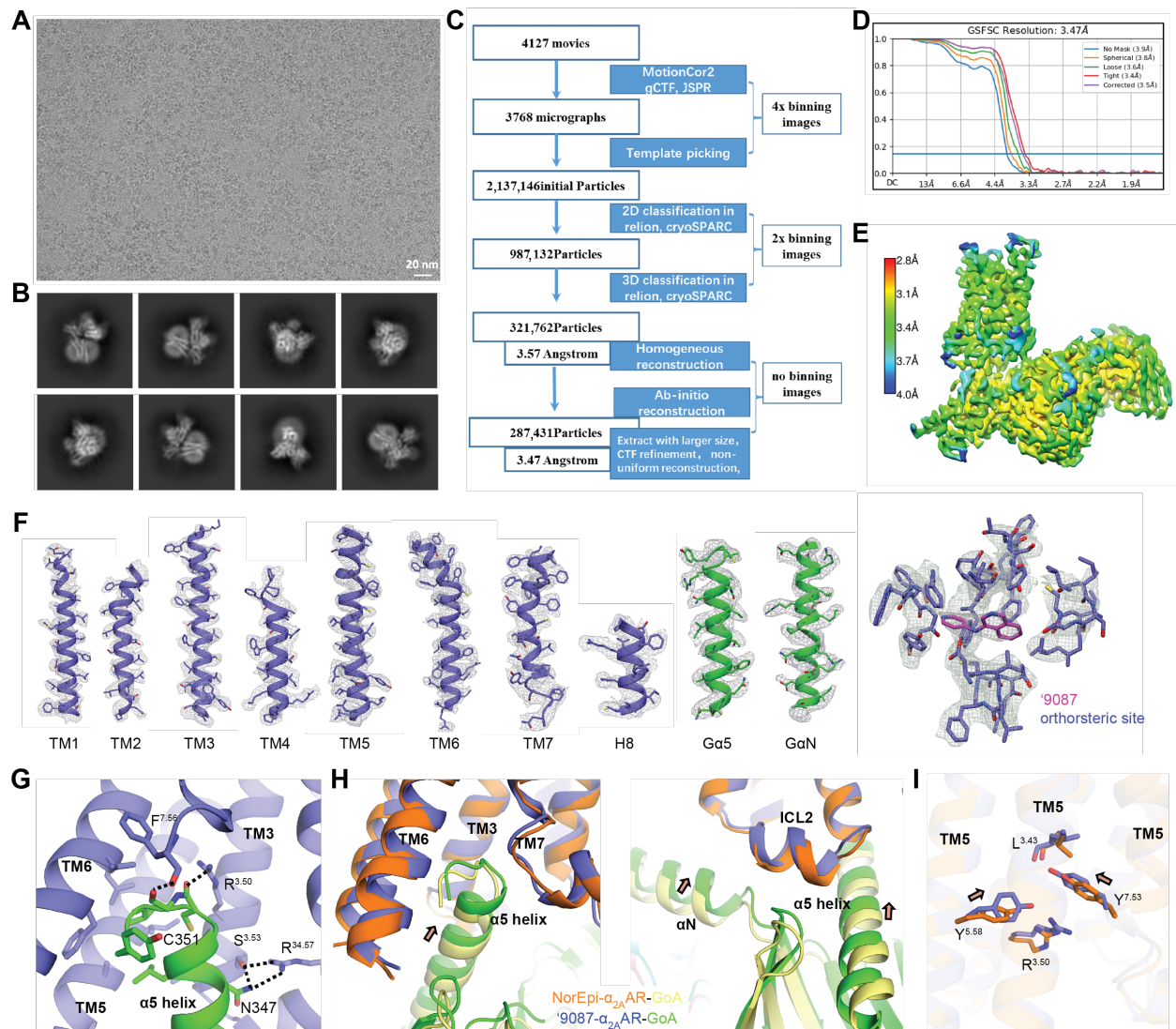
G-protein and  $\beta$ -Arrestin signaling profiles for docking compounds in BRET biosensor-based assays in HEK293 cells expressing the human  $\alpha_{2A}AR$ . Relative activities are from Supplementary Table 2.4.



**Figure 2.S8 Internalization behavior of  $\alpha_{2A}$ AR following compound treatment.**

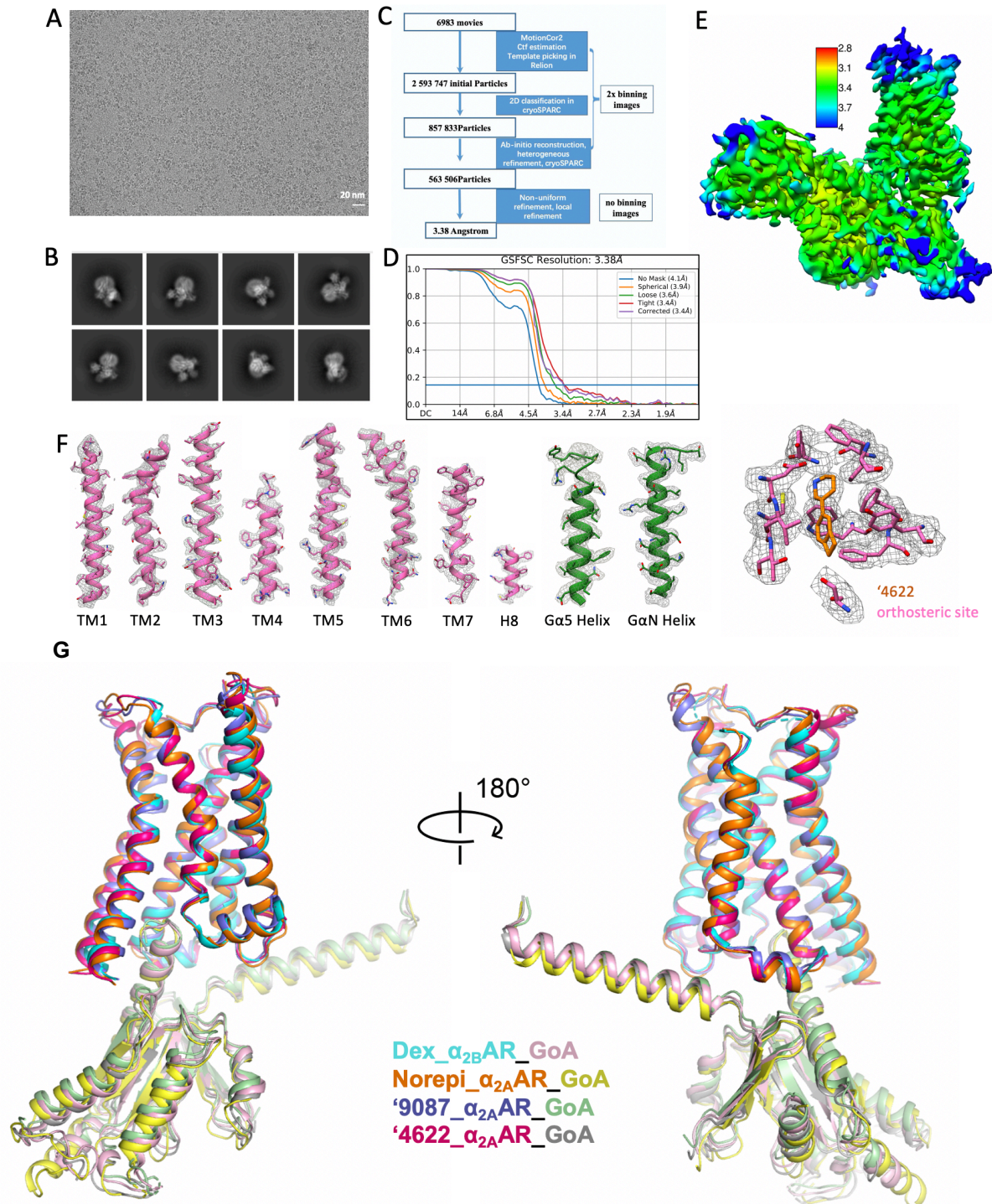
(A) Kinetics of  $\alpha_{2A}$ AR disappearance from the plasma membrane following 100 $\mu$ M compound treatment using an human  $\alpha_{2A}$ AR-RlucII/rGFP-CAAX biosensor. (B) Kinetics of  $\alpha_{2A}$ AR relocation in endosomes following 100 $\mu$ M compound treatment using a human  $\alpha_{2A}$ AR-RlucII/rGFP-FYVE biosensor. (C) Concentration-response curves of  $\alpha_{2A}$ AR disappearance from the plasma membrane using an human  $\alpha_{2A}$ AR-RlucII/rGFP-CAAX biosensor. (D) Concentration-response curves of  $\alpha_{2A}$ AR relocation in endosomes using an human  $\alpha_{2A}$ AR-RlucII/rGFP-CAAX biosensor. (E) Summary of biosensor data for compounds. For (C-D) Normalization was done relative to the maximum effect of norepinephrine. Data are mean  $\pm$  s.e.m. of normalized results (n=3 measurements).





**Figure 2.S9 Structural determination of '9087- $\alpha_{2A}$ AR-G $_{oA}$ .**

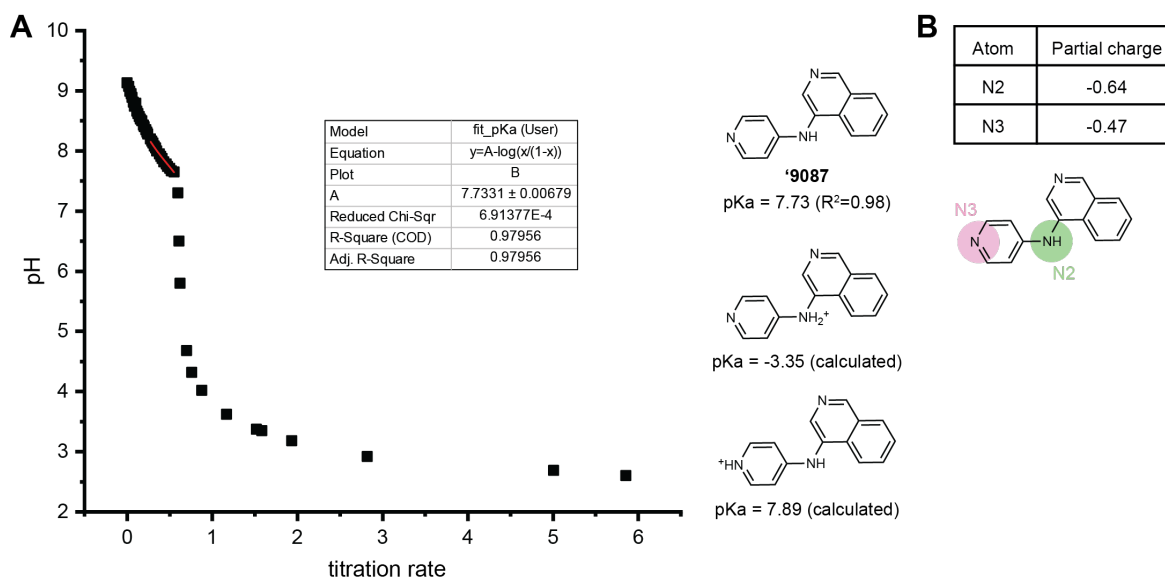
(A) Representative cryo-EM image of the '9087- $\alpha_{2A}$ AR-G $_{oA}$  complex. (B) Representative '9087- $\alpha_{2A}$ AR-G $_{oA}$  complex 2D class averages with different views. (C) Work flow for cryo-EM image processing. (D) Gold standard Fourier shell correlation (FSC) curve indicates overall nominal resolution at 3.47 Å using the FSC=0.143 criterion. (E) Local resolution mapped onto density of '9087- $\alpha_{2A}$ AR-G $_{oA}$  complex. (F) Densities surrounding coordinates of TM 1-7, H8, G $\alpha_5$ , G $\alpha_N$  helices, '9087, and orthosteric site residues. (G) Interactions between the  $\alpha_5$  helix of G $_{o\alpha}$  (green) and  $\alpha_{2A}$ AR (blue). (H) Tight coupling between  $\alpha_{2A}$ AR and G $_{o\alpha}$  with a 1.5 Å shift of the  $\alpha_5$  and  $\alpha_N$  helices of G $_{o\alpha}$  towards  $\alpha_{2A}$ AR, as compared to the norepinephrine- $\alpha_{2A}$ AR-G $_{o\alpha}$  complex (PDB 7EJ0). (I) Conserved Y $^{5.58}$  and Y $^{7.53}$  move ~1.5 Å closer in the '9087- $\alpha_{2A}$ AR-G $_{oA}$  complex (blue) compared to the norepinephrine- $\alpha_{2A}$ AR-G $_{oA}$  complex (orange). For F, the zone size is 2 and the map threshold is between 0.75 and 0.85.



**Figure 2.S10 Structural determination of '4622- $\alpha_{2A}$ AR-G $_{oA}$ .**

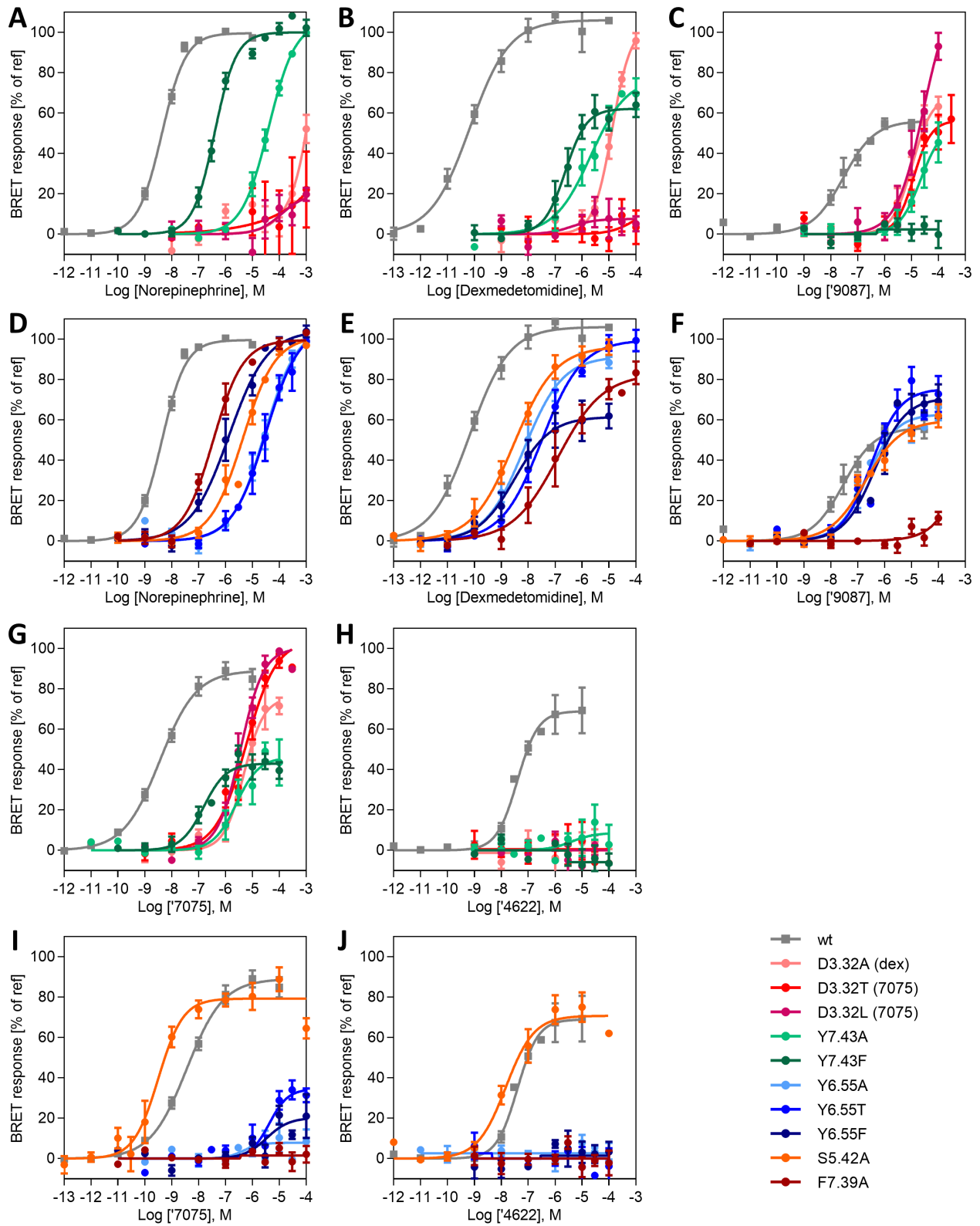
(A) Representative cryo-EM image of the '4622- $\alpha_{2A}$ AR-G $_{oA}$  complex. (B) Representative '4622- $\alpha_{2A}$ AR-G $_{oA}$  complex 2D class averages with different views. (C) Work flow for cryo-EM image processing, (D) Gold standard Fourier shell correlation (FSC) curve indicates overall nominal resolution at 3.47 Å using the FSC=0.143 criterion. (E) Local resolution mapped onto density of '4622- $\alpha_{2A}$ AR-G $_{oA}$  complex. (F) Densities surrounding coordinates of TM 1-7, H8, G $\alpha$ 5, G $\alpha$ N

helices, and orthosteric site. (G) Comparison of highly similar  $G_{\alpha A}$  interfaces for  $\alpha_{2A}AR$  and  $\alpha_{2B}AR$  bound to dexmedetomidine (PDB 7EJA), norepinephrine (PDB 7EJ0), **'9087**, or **'4622**. For F, the zone size is 2 and the map threshold is between 0.75 and 0.85.



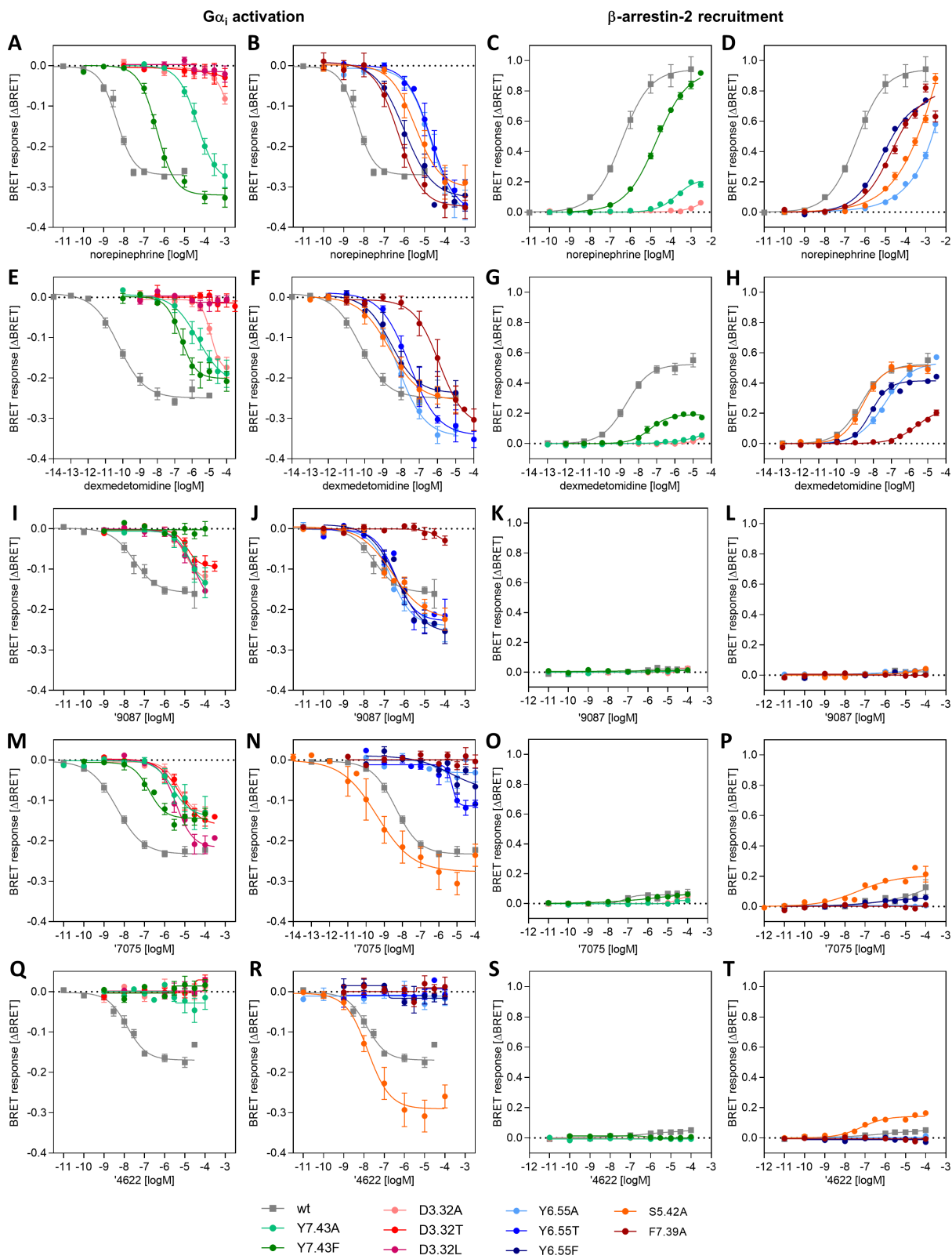
**Figure 2.S11 '9087 formal and partial charges.**

(A) Formal charge of '9087 determined at the pyrimidine nitrogen with a pKa of 7.73. (B) Partial charge of pyrimidine (N3, pink) and exocyclic (N2, green) nitrogens as calculated by semi-empirical quantum mechanics for docking model.



**Figure 2.S12 Substitutions of binding site residues reveals distinct interaction patterns for known agonists and potent docking agonists.**

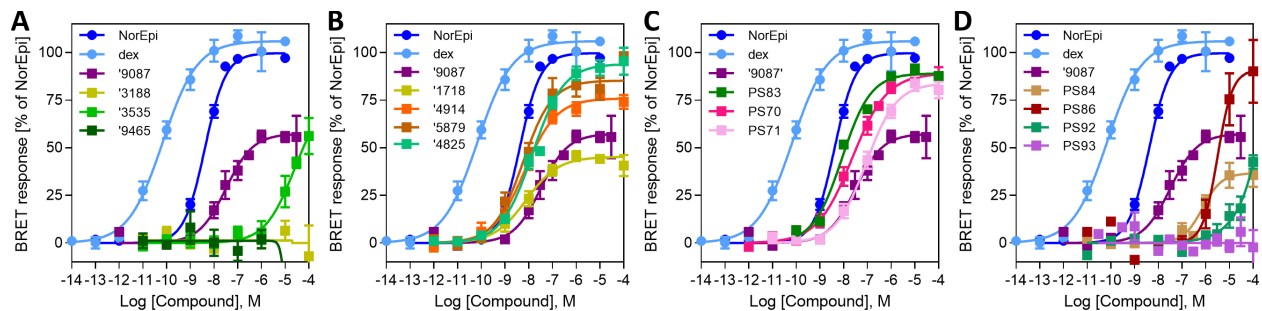
Substitutions of residues compared to wildtype (wt) for (A,D) norepinephrine, (B,E) dexmedetomidine, (C,F) '9087, (G,I) '7075, and (H,J) '4622 for  $\alpha_{2A}$ AR. Gi activation was monitored in a BRET-biosensor based assay with 3-7 experiments for each mutant. Normalization was done relative to the maximum effect of NorEpi with the exception of mutants D128<sup>3.32</sup>A (DEX), D128<sup>3.32</sup>T and D128<sup>3.32</sup>L ('7075) as indicated in figure legend.



**Figure 2.S13 Substitution of binding site residues reveals distinct interaction patterns for known agonists and potent docking agonists.**

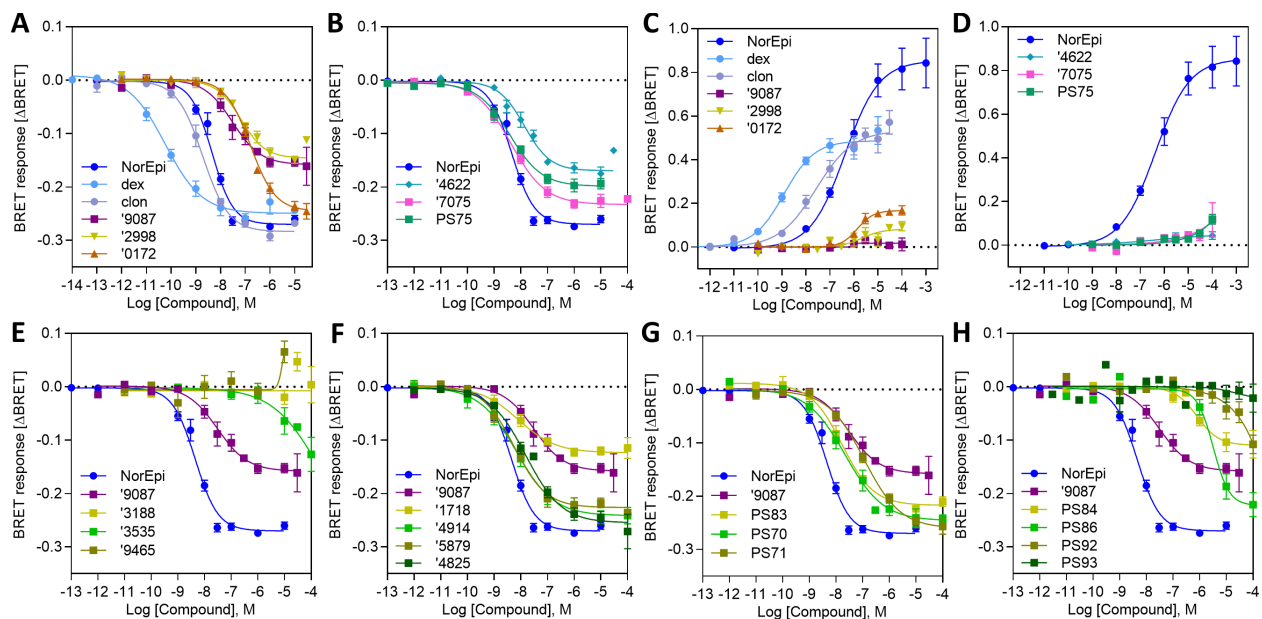
G-protein (A,B,E,F,I,J,M,N,Q,R) and  $\beta$ -arrestin signaling (C,D,G,H,K,L,O,P,S,T) for norepinephrine (A-D) and dexmedetomidine (E-H) in comparison to the docking hits '9087 (I-L), '7075 (M-P), and '4622 (Q-T) in BRET-biosensor based assays in HEK293T cells expressing the human  $\alpha_{2A}$ AR wild-type or mutants. Data are mean  $\pm$  s.e.m. displayed as delta BRET values (n=3 to 17 measurements).



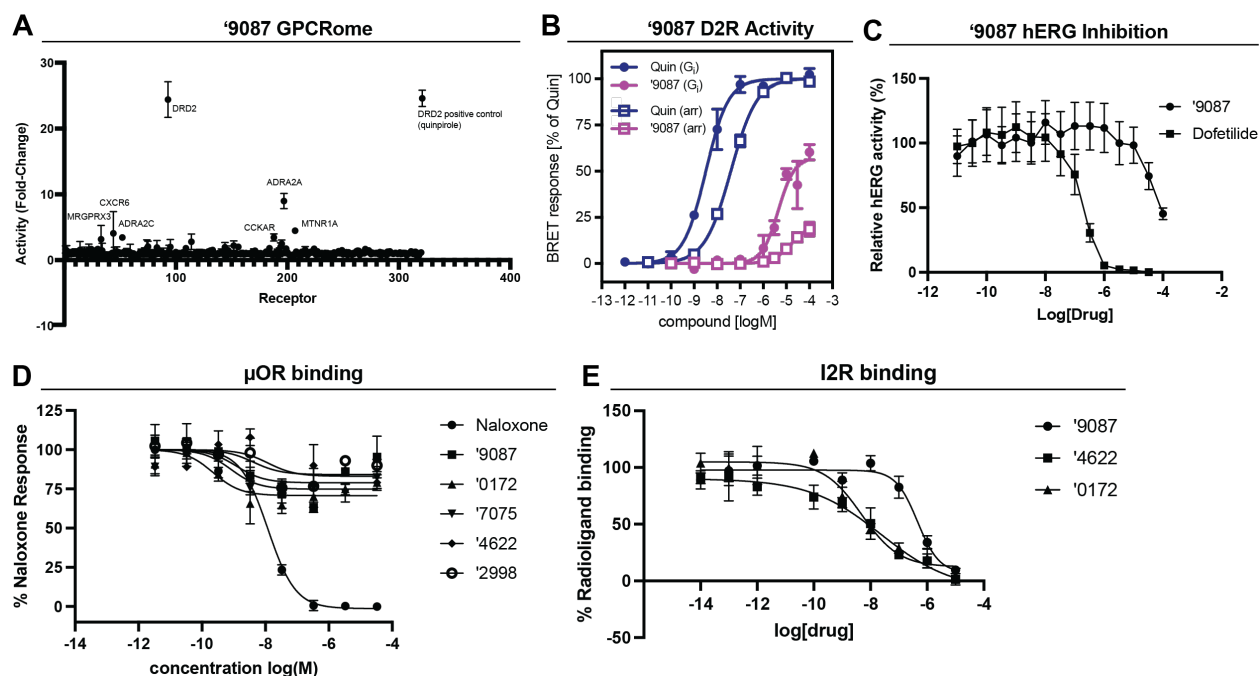


**Figure 2.S14 Analogs of '9087 reveal key SAR.**

Gi activation for human  $\alpha_{2A}AR$  of analogs with changes to the pyridine ring, exocyclic nitrogen, and isoquinoline of '9087 monitored in a BRET-biosensor based assay. Data are mean  $\pm$  s.e.m. of normalized results (n=3-9 measurements).

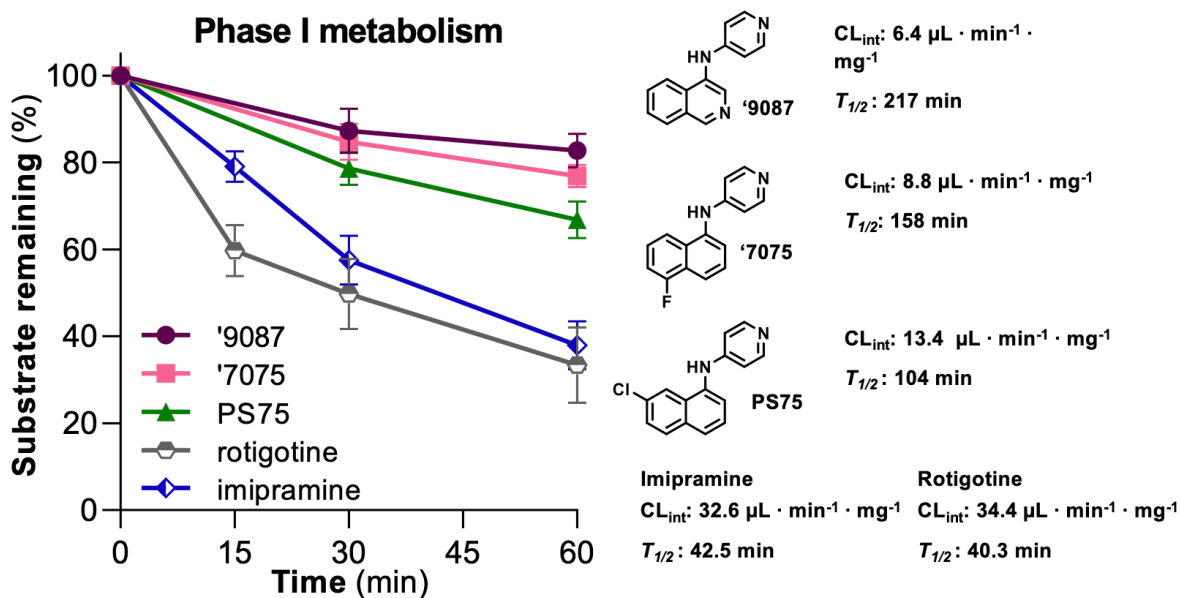


**Figure 2.S15 Functional data for selected docking hits and references against  $\alpha_{2A}AR$ .** (A, B, E-H) Gi activation and (C,D)  $\beta$ -arrestin-2 recruitment determined in BRET-biosensor based assays in HEK293T cells expressing the human  $\alpha_{2A}AR$  wild-type. Data are mean  $\pm$  s.e.m. (n=3 to 17 measurements) displayed as delta BRET values corresponding to functional data of selected docking hits and references referred to in Figure 1C, 3B (A-D), and Fig. 2.S14 (E-H) in the order of their appearance.

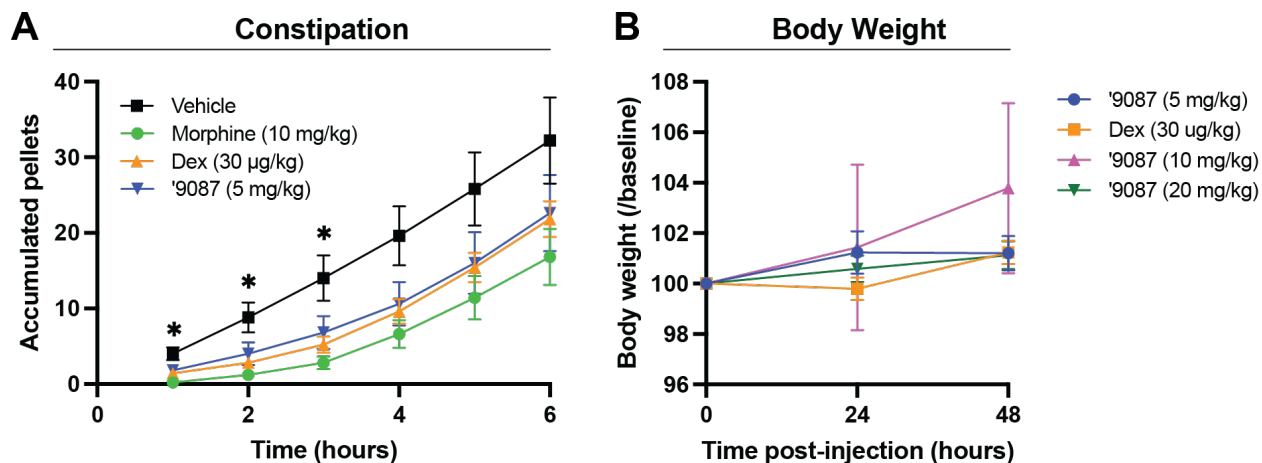


**Figure 2.S16 Off-target activity for  $\alpha_{2A}AR$  agonists.**

(A) GPCRome of '9087. Labeled targets indicate an increase of 3-fold or higher signaling compared to basal activity. Positive control is shown with D2R<sub>long</sub> and quinpirole. (B) D2R<sub>long</sub> G<sub>i</sub> and  $\beta$ -arrestin-2 (arr) recruitment for '9087 with EC<sub>50</sub>= 4.5  $\mu$ M and E<sub>max</sub>=57% for G<sub>i</sub> signaling and EC<sub>50</sub>= 16  $\mu$ M and E<sub>max</sub>=21% for arrestin recruitment, respectively. Positive control quinpirole (Quin) also shown. (C) hERG inhibition of '9087 and positive control dofetilide. (D)  $\mu$ OR binding of  $\alpha_{2A}AR$  docking agonists and analogs. (E) I2R binding of  $\alpha_{2A}AR$  agonists. For (A-E), data are shown as mean  $\pm$  s.e.m. (n=3 to 4 measurements).



**Figure 2.S17 Phase I metabolism of '9087, '7075, and PS75 in male rat liver microsomes.** Rotigotine and imipramine serve as positive controls for extensive phase I metabolism. Data are percent of non-metabolized compound remaining shown as mean  $\pm$  s.e.m. (n=4 individual experiments for imipramine and rotigotine, n=5 for '9087, '7075, PS75).

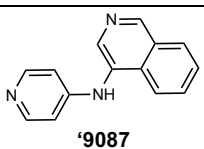
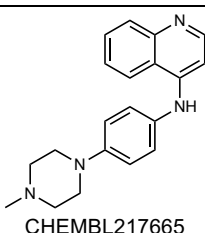
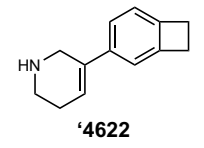
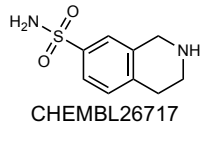
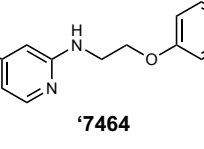
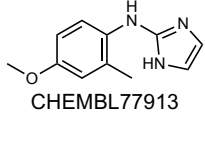
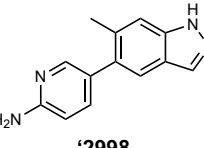
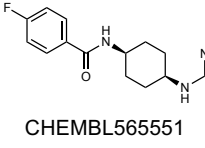
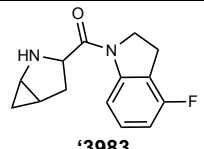
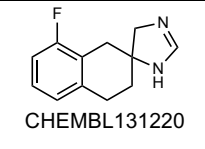


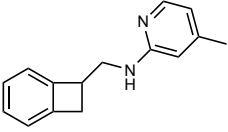
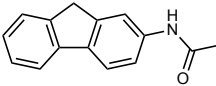
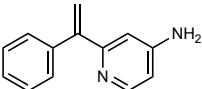
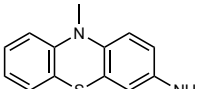
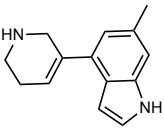
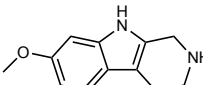
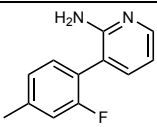
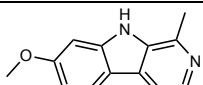
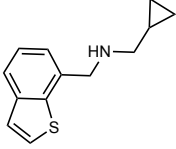
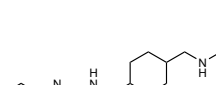
**Figure 2.S18 In vivo side effects of constipation and body weight.**

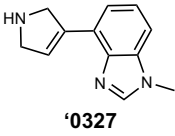
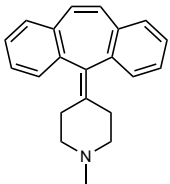
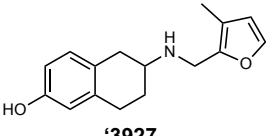
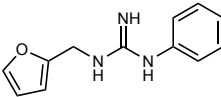
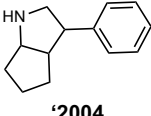
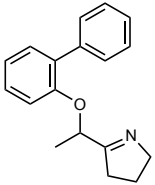
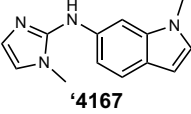
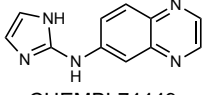
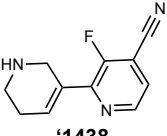
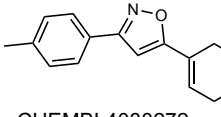
(A) Constipation monitored up to 6 hours following vehicle or compound i.p. injection (two-way ANOVA; time  $\times$  treatment interaction:  $F(15,80) = 1.501$ ,  $P = 0.1250$ ; time:  $F(1.071,17.14) = 111.7$ ,  $P < 0.0001$ ; treatment:  $F(3,16) = 3.784$ ,  $P = 0.0316$ ; all treatment groups ( $n = 6$ )); each compound time point compared to vehicle, asterisks define difference between morphine and vehicle at 1 h ( $P = 0.0209$ ), 2 h ( $P = 0.0372$ ) and 3 h ( $P = 0.0417$ ) for simplicity, all other points compared to vehicle are not significant; \*  $p < 0.05$ ). Data are mean  $\pm$  s.e.m. (B) Body weight measured over 48 hours following vehicle or compound i.p. injection (two-way ANOVA; time  $\times$  treatment interaction:  $F(6,32) = 0.5174$ ,  $P = 0.7907$ ; time:  $F(1.161,18.57) = 3.177$ ,  $P = 0.0863$ ; treatment:  $F(3,16) = 0.2854$ ,  $P = 0.8358$ ; all treatment groups ( $n = 3$ )); time points compared within same treatment groups, all comparisons are not significant). Data are mean  $\pm$  s.e.m.

## 2.6 Tables

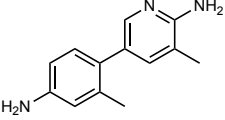
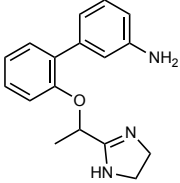
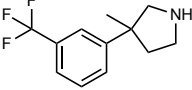
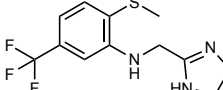
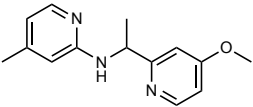
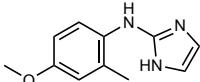
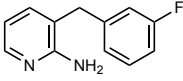
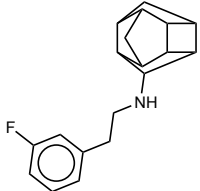
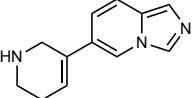
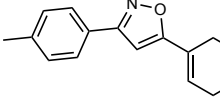
**Table 2.S1 Binding affinities and functional properties at  $\alpha_{2A}$ AR or  $\alpha_{2B}$ AR for hits identified in initial docking screen.**

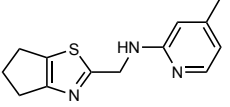
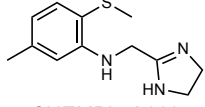
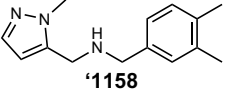
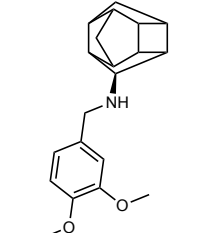
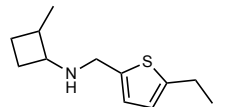
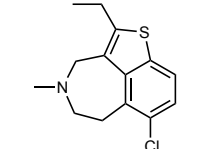
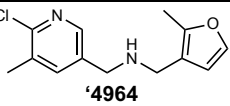
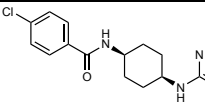
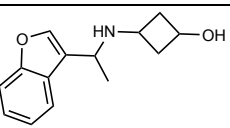
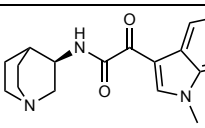
Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sub>f,g,h</sub>	$\alpha_{2A}$ AR	$\alpha_{2B}$ AR	TC <sup>i</sup>	Nearest ChEMBL $\alpha_{2A}$ AR ligand <sup>j</sup>
 <b>'9087</b>	4560 (12687) <sup>a</sup>	$K_i^e$ n	72 nM ± 21 6	1.7 nM ± 0.65 7	0.30	 ChEMBL217665
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	27 nM ± 5.7 71% ± 3 7	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	79 nM ± 18 78% ± 3 10	1.6 nM ± 0.21 72% ± 3 4		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	11 nM ± 1.8 10% ± 1 3		
 <b>'4622</b>	406 (1057) <sup>a</sup>	$K_i^e$ n	110 nM ± 18 4	60 nM ± 25 4	0.32	 ChEMBL26717
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	25 nM ± 3.8 69% ± 0 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	16 nM ± 2.8 74% ± 3 5	84 nM ± 25 41% ± 4 4		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	710 nM ± 180 9% ± 3 3	-- nM <5% 3		
 <b>'7464</b>	1703 (4522) <sup>a</sup>	$K_i^e$ n	340 nM ± 66 4	170 nM ± 80 3	0.24	 ChEMBL77913
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <b>'2998</b>	2451 (6666) <sup>b</sup>	$K_i^e$ n	250 nM ± 47 7	160 nM ± 44 6	0.19	 ChEMBL565551
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	6.3 nM ± 0.78 87% ± 1 7	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	53 nM ± 12 78% ± 3 8	65 nM ± 22 81% ± 5 4		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	3000 nM ± 300 24% ± 3 4		
 <b>'3983</b>	3909 (11204) <sup>b</sup>	$K_i^e$ n	650 nM ± 150 5	330 nM ± 54 5	0.29	 ChEMBL131220
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	69 nM ± 2.7 76% ± 2 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	100 nM ± 11 41% ± 5 4	120 nM ± 28 86% ± 2 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	1600 nM ± 160 58% ± 5 4		

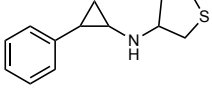
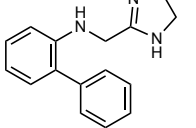
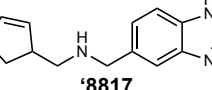

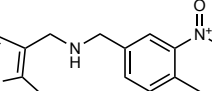
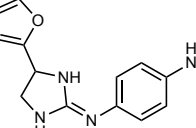
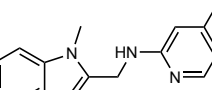
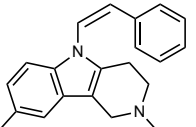
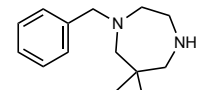
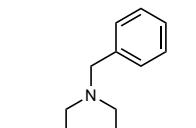
Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sup>f,g,h</sup>	$\alpha_{2A}AR$	$\alpha_{2B}AR$	TC <sup>i</sup>	Nearest ChEMBL $\alpha_{2A}AR$ ligand <sup>j</sup>
 <p>'1493</p>	91 (163) <sup>b</sup>	$K_i^e$ n	1200 nM ± 260 3	100 nM ± 44 3	0.29	 <p>ChEMBL311469</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p>'0172</p>	341 (879) <sup>a</sup>	$K_i^e$ n	1200 nM ± 290 3	260 nM ± 110 4	0.23	 <p>ChEMBL1323615</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	40 nM ± 17 94% ± 2 4	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	32 nM ± 2.4 91% ± 1 7	27 nM ± 4 95% ± 5 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	660 nM ± 84 42% ± 9 3	350 nM ± 23 85% ± 2 3		
 <p>'3928</p>	590 (1541) <sup>a</sup>	$K_i^e$ n	1500 nM ± 500 2	830 nM ± 35 2	0.24	 <p>ChEMBL6310</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	-- nM 38% ± 6 @10μM 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	2000 nM ± 1300 65% ± 15 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p>'6064</p>	93 (190) <sup>a</sup>	$K_i^e$ n	3500 nM ± 420 2	1500 nM ± 1300 2	0.24	 <p>ChEMBL269538</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	700 nM ± 200 66% ± 4 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p>'1627</p>	452 (1190) <sup>a</sup>	$K_i^e$ n	4300 nM ± 1400 2	1500 nM ± 1400 2	0.28	 <p>ChEMBL360117</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	4300 nM ± 2000 66% ± 18 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		

Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sup>f,g,h</sup>	$\alpha_2AAR$	$\alpha_2BAR$	TC <sup>i</sup>	Nearest ChEMBL $\alpha_2AAR$ ligand <sup>j</sup>
 <p><b>'0327</b></p>	4419 (13032) <sup>b</sup>	$K_i^e$ n	4300 nM ± 3000 2	6600 nM ± 2900 2	0.21	 <p>ChEMBL516</p>
		$G\alpha_{q1}^f$ EC <sub>50</sub> E <sub>max</sub> n	1600 nM ± 190 50% ± 0 3	--		
		$G\alpha_{q9}$ EC <sub>50</sub> E <sub>max</sub> n	--	1500 nM ± 290 81% ± 9 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	23000 nM ± 4700 17% ± 7 3		
 <p><b>'3927</b></p>	1878 (4691) <sup>c</sup>	$K_i^e$ n	4600 nM ± 1800 3	370 nM ± 160 3	0.25	 <p>ChEMBL3289531</p>
		$G\alpha_{q1}^f$ EC <sub>50</sub> E <sub>max</sub> n	3000 nM ± 640 45% ± 2 3	--		
		$G\alpha_{q9}$ EC <sub>50</sub> E <sub>max</sub> n	--	990 nM ± 710 18% ± 8 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p><b>'2004</b></p>	2474 (6733) <sup>b</sup>	$K_i^e$ n	5600 nM ± 2200 3	690 nM ± 150 3	0.23	 <p>ChEMBL222593</p>
		$G\alpha_{q1}^f$ EC <sub>50</sub> E <sub>max</sub> n	-- nM <5% 3	--		
		$G\alpha_{q9}$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 4		
 <p><b>'4167</b></p>	12 (17) <sup>a</sup>	$K_i^e$ n	6400 nM ± 2500 2	3700 nM ± 2100 2	0.25	 <p>ChEMBL74449</p>
		$G\alpha_{q1}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{q9}$ EC <sub>50</sub> E <sub>max</sub> n	--	800 nM ± 230 88% ± 5 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	24000 nM ± 3200 28% ± 8 3		
 <p><b>'1438</b></p>	934 (2486) <sup>a</sup>	$K_i^e$ n	7300 nM ± 2900 2	5000 nM ± 1300 2	0.19	 <p>ChEMBL4088272</p>
		$G\alpha_{q1}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{q9}$ EC <sub>50</sub> E <sub>max</sub> n	--	4500 nM ± 3800 60% ± 17 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 4		



Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sup>f,g,h</sup>	$\alpha_2AAR$	$\alpha_2BAR$	TC <sup>i</sup>	Nearest ChEMBL $\alpha_2AAR$ ligand <sup>j</sup>
 <p><b>'8689</b></p>	79 (139) <sup>b</sup>	$K_i^e$ n	8900 nM $\pm$ 3100 3	1100 nM $\pm$ 180 3	0.23	 <p>ChEMBL222371</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	98 nM $\pm$ 18 96% $\pm$ 4 3	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	270 nM $\pm$ 52 87% $\pm$ 5 4	130 nM $\pm$ 42 91% $\pm$ 2 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	28000 nM $\pm$ 8700 31% $\pm$ 8 3	8200 nM $\pm$ 360 64% $\pm$ 3 3		
 <p><b>'5264</b></p>	693 (1839) <sup>a</sup>	$K_i^e$ n	9400 nM $\pm$ 140 2	2900 nM $\pm$ 780 2	0.24	 <p>ChEMBL72724</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM 13% $\pm$ 4 @10 $\mu$ M 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p><b>'1190</b></p>	81 (131) <sup>a</sup>	$K_i^e$ n	>10000 nM 2	1600 nM $\pm$ 70 2	0.33	 <p>ChEMBL77913</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p><b>'1447</b></p>	1494 (3922) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	5000 nM $\pm$ 4500 2	0.29	 <p>ChEMBL1643900</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	1300 nM $\pm$ 440 65% $\pm$ 8 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <p><b>'9121</b></p>	2458 (6674) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	2200 nM $\pm$ 1600 2	0.25	 <p>ChEMBL4088272</p>
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		

Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sup>f,g,h</sup>	$\alpha_2AAR$	$\alpha_2BAR$	TC <sup>i</sup>	Nearest ChEMBL $\alpha_2AAR$ ligand <sup>j</sup>
 <b>'3788</b>	413 (1083) <sup>a</sup>	$K_i^e$ n	>10000 nM 3	3100 nM $\pm$ 1300 2	0.22	 ChEMBL72441
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'1158</b>	4320 (12733) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	4000 nM $\pm$ 2300 2	0.20	 ChEMBL2432046
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'2964</b>	984 (2647) <sup>a</sup>	$K_i^e$ n	>10000 nM 4	6500 nM $\pm$ 3900 2	0.25	 ChEMBL123448
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'4964</b>	3004 (8103) <sup>c</sup>	$K_i^e$ n	>10000 nM 2	7500 nM $\pm$ 1500 2	0.20	 ChEMBL577912
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'1099</b>	379 (883) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	7700 nM $\pm$ 6100 2	0.22	 ChEMBL294649
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		

Compound	Cluster rank (Global rank) <sup>d</sup>	Binding <sup>e</sup> and functional properties <sup>f,g,h</sup>	$\alpha_{2A}AR$	$\alpha_{2B}AR$	TC <sup>i</sup>	Nearest ChEMBL $\alpha_{2A}AR$ ligand <sup>j</sup>
 <b>'1774</b>	1631 (4337) <sup>a</sup>	$K_i^e$ n	>10000 nM 2	8200 nM ± 4000 2	0.21	 ChEMBL337862
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	-- nM <5% 3		
 <b>'8817</b>	4909 (14777) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	8300 nM ± 5300 2	0.33	 ChEMBL74544
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'7963</b>	4660 (13691) <sup>c</sup>	$K_i^e$ n	>10000 nM 2	8700 nM ± 1300 2	0.21	 ChEMBL3289549
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'8695</b>	14744 (53883) <sup>c</sup>	$K_i^e$ n	>10000 nM 2	9100 nM ± 2700 2	0.23	 ChEMBL605405
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		
 <b>'5343</b>	1201 (3112) <sup>b</sup>	$K_i^e$ n	>10000 nM 2	9400 nM ± 2300 2	0.33	 ChEMBL141209
		$G\alpha_{qi}^f$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		$G\alpha_{qi}^g$ EC <sub>50</sub> E <sub>max</sub> n	--	--		
		Arr2 <sup>h</sup> EC <sub>50</sub> E <sub>max</sub> n	--	--		

<sup>abc</sup> Correspond to three docking screens (see Methods)

<sup>d</sup> Cluster rank and global rank from docking screens; global rank in brackets (Methods)

<sup>e</sup> Binding affinity to murine  $\alpha_{2A}AR$  or human  $\alpha_{2B}AR$  as mean  $K_i$  value ± s.e.m. from 3-7 experiments or as mean  $K_i$  value ± SD for 2 experiments

<sup>f</sup> G-protein signaling determined in an IP-One assay with HEK293T cells expressing murine  $\alpha_{2A}AR$  and the hybrid G-protein  $G\alpha_{qi}$  from 3-7 experiments

<sup>g</sup> G-protein signaling (IP-One assay) with the human  $\alpha_{2A}AR$  or human  $\alpha_{2B}AR$  and  $G\alpha_{qi}$  from 4-10 experiments

<sup>h</sup>  $\beta$ -Arrestin-2 recruitment with human  $\alpha_{2A}AR$  or  $\alpha_{2B}AR$  derived from 3 or 3-4 experiments, respectively

<sup>i</sup> ECFP4 Tanimoto coefficient (Tc) to most similar known  $\alpha_{2A}AR$  ligand in ChEMBL29

<sup>j</sup> Most similar  $\alpha_{2A}AR$  ChEMBL29 ligand with activity < 10 $\mu$ M to the docking hit

--, not determined

**Table 2.S2 Binding affinity and functional properties of selected compounds in comparison to the reference compounds norepinephrine, dexmedetomidine, and clonidine to the human  $\alpha_{2A}$ AR.**

compound	$K_i$ for $\alpha_{2A}$ AR <sup>a</sup>		$G\alpha_i$ <sup>b</sup>			$G\beta_6$ <sup>c</sup>			$\beta$ -arrestin-2 <sup>d</sup>		
	[nM $\pm$ s.e.m.]	n	EC <sub>50</sub> [nM $\pm$ s.e.m.]	E <sub>max</sub> [% $\pm$ s.e.m.]	n	EC <sub>50</sub> [nM $\pm$ s.e.m.]	E <sub>max</sub> [% $\pm$ s.e.m.]	n	EC <sub>50</sub> [nM $\pm$ s.e.m.]	E <sub>max</sub> [% $\pm$ s.e.m.]	n
NorEpi	4100 $\pm$ 1500	4	5.0 $\pm$ 0.94	100	17	--	--	--	450 $\pm$ 30	100	8
DEX	10 $\pm$ 1.5	8	0.077 $\pm$ 0.015	105 $\pm$ 4	7	0.057 $\pm$ 0.015	76 $\pm$ 3	2	1.3 $\pm$ 0.43	60 $\pm$ 3	5
CLON	84 $\pm$ 21 <sup>e</sup>	2	2.4 $\pm$ 0.96	115 $\pm$ 4	5	--	--	--	17 $\pm$ 2.5	41 $\pm$ 3	5
BRIM	--	--	--	--	--	0.232 $\pm$ 0.043	107 $\pm$ 7	2	--	--	--
'9087	51 $\pm$ 8.8	6	52 $\pm$ 24	60 $\pm$ 3	9	87 $\pm$ 13	42 $\pm$ 2	2	--	<5	6
'2998	180 $\pm$ 34	5	73 $\pm$ 14	61 $\pm$ 3	7	--	--	--	98 $\pm$ 32	8 $\pm$ 3	4
'0172	260 $\pm$ 33	4	210 $\pm$ 38	95 $\pm$ 7	4	--	--	--	1700 $\pm$ 290	22 $\pm$ 2	3
'4622	12 $\pm$ 3.6	4	9.7 $\pm$ 2.4	74 $\pm$ 4	6	--	--	--	--	<5	3
'7075	37 $\pm$ 5.1	5	4.1 $\pm$ 1.2	93 $\pm$ 5.2	9	18 $\pm$ 1.9	96 $\pm$ 2	2	--	10 $\pm$ 7 @ 0.1mM	7
PS75	8.2 $\pm$ 0.48	4	4.8 $\pm$ 1.3	82 $\pm$ 4	7	--	--	--	--	15 $\pm$ 3 @ 0.1mM	6

<sup>a</sup> Binding affinity to human  $\alpha_{2A}$ AR as mean  $K_i$  value  $\pm$  s.e.m derived from 4-8 experiments each done in triplicate.

<sup>b</sup> G-protein signaling monitored in a BRET-biosensor based assay in HEK293T cells expressing the human  $\alpha_{2A}$ AR performed in duplicates.

<sup>c</sup> G-protein signaling monitored in DiscoverX HitHunter cAMP assay expressing the human  $\alpha_{2A}$ AR.

<sup>d</sup>  $\beta$ -Arrestin-2 recruitment monitored in a BRET-biosensor based assay in HEK293T cells expressing the human  $\alpha_{2A}$ AR performed in duplicates.

<sup>e</sup>  $K_i$  value  $\pm$  SD derived from 2 experiments.

NorEpi, norepinephrine

DEX, dexmedetomidine

CLON, clonidine

BRIM, brimonidine

--, not determined

**Table 2.S3 Functional properties in EMTA coupling panel of '9087 and its analogs in comparison to norepinephrine, dexmedetomidine, and brimonidine against the human  $\alpha_2$ AR.**

Transducer		Compound functional data					
Endogenous Gi/o*		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	101	111	55	75	76
	SEM	1.99	2.28	2.71	1.63	1.66	1.22
	mean	-8.63	-10.01	-9.83	-7.58	-7.82	-8.11
	SEM	0.05	0.06	0.07	0.07	0.05	0.03
N		3	5	5	5	5	5
<b>Gi1</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	104	100	64	95	95
	SEM	2.42	0.92	1.09	1.87	0.97	0.91
LogEC50	mean	-8.66	-9.92	-9.96	-7.20	-7.47	-7.97
	SEM	0.06	0.02	0.02	0.05	0.02	0.02
N		3	5	5	5	5	5
<b>Gi2</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	101	99	62	96	101
	SEM	2.25	1.23	1.12	2.63	0.92	1.17
LogEC50	mean	-8.76	-10.00	-10.02	-7.26	-7.51	-7.92
	SEM	0.06	0.02	0.02	0.08	0.02	0.02
N		3	5	5	5	5	5
<b>Gi3</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	101	105	54	91	101
	SEM	1.91	1.51	2.40	2.71	1.58	2.75
LogEC50	mean	-8.49	-9.92	-9.88	-7.22	-7.46	-7.85
	SEM	0.05	0.03	0.05	0.10	0.04	0.05
N		3	5	5	5	5	5
<b>GoA</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	108	105	96	95	96
	SEM	3.86	1.46	1.58	1.94	3.55	1.20
LogEC50	mean	-9.01	-10.36	-10.33	-7.59	-7.97	-8.41
	SEM	0.07	0.03	0.04	0.04	33.55	0.03
N		3	5	5	5	5	5
<b>GoB</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	107	105	98	103	105
	SEM	3.43	1.60	1.98	1.77	2.58	1.66
LogEC50	mean	-9.17	-10.57	-10.48	-7.76	-8.08	-8.57
	SEM	0.09	0.04	0.05	0.04	0.04	0.04
N		3	5	5	5	5	5
<b>Gz</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	99	98	95	93	99
	SEM	4.38	1.46	1.52	2.45	1.44	0.99
LogEC50	mean	-10.02	-11.50	-11.68	-8.86	-8.93	-9.43
	SEM	0.09	0.04	0.04	0.05	0.03	0.03
N		3	5	5	5	5	5
<b>Gs</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	57	85	<5	<5	<5
	SEM	3.82	4.21	3.92	--	--	--
LogEC50	mean	-6.50	-8.04	-7.82	--	--	--
	SEM	0.08	0.17	0.10	--	--	--
N		5	5	5	5	5	5
<b>G12</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	61	107	<5	<5	<5
	SEM	2.48	2.26	4.00	--	--	--
LogEC50	mean	-6.60	-7.91	-7.58	--	--	--
	SEM	0.07	0.09	0.09	--	--	--
N		5	5	5	5	5	5
<b>G13</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	41	111	<5	<5	<5
	SEM	1.39	1.04	3.19	--	--	--
LogEC50	mean	-6.10	-7.60	-7.18	--	--	--
	SEM	0.02	0.06	0.07	--	--	--
N		5	5	5	5	5	5
<b>Gq</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
Emax	mean	100	28	130	<5	<5	<5
	SEM	1.70	1.59	3.77	--	--	--
LogEC50	mean	-5.95	-7.34	-6.74	--	--	--
	SEM	0.03	0.14	0.07	--	--	--
N		5	5	5	5	5	5

Transducer		Compound functional data					
<b>G11</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
E <sub>max</sub>	mean	100	25	153	<5	<5	<5
	SEM	3.22	1.44	10.12	--	--	--
LogEC <sub>50</sub>	mean	-5.90	-7.62	-6.26	--	--	--
	SEM	0.06	0.14	0.16	--	--	--
N		5	5	5	5	5	5
<b>G14</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
E <sub>max</sub>	mean	100	24	117	<5	<5	<5
	SEM	3.48	0.94	4.06	--	--	--
LogEC <sub>50</sub>	mean	-5.83	-7.64	-6.77	--	--	--
	SEM	0.06	0.10	0.08	--	--	--
N		5	5	5	5	5	5
<b>G15</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
E <sub>max</sub>	mean	100	74	107	<5	<5	<5
	SEM	0.02	0.02	0.04	--	--	--
LogEC <sub>50</sub>	mean	-6.30	-7.59	-7.42	--	--	--
	SEM	0.94	0.75	1.65	--	--	--
N		5	5	5	5	5	5
<b>barrestin1</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
E <sub>max</sub>	mean	100	30	103	<5	<5	<5
	SEM	2.84	1.10	5.05	--	--	--
LogEC <sub>50</sub>	mean	-5.80	-7.89	-6.17	--	--	--
	SEM	0.05	0.09	0.12	--	--	--
N		5	5	5	5	5	5
<b>barrestin2</b>		Norepinephrine	Dexmedetomidine	Brimonidine	'9087	'7075	PS75
E <sub>max</sub>	mean	100	38	92	<5	<5	<5
	SEM	1.53	0.94	2.58	--	--	--
LogEC <sub>50</sub>	mean	-5.94	-7.62	-6.74	--	--	--
	SEM	0.02	0.06	0.07	--	--	--
N		5	5	5	5	5	5

G-protein and  $\beta$ -Arrestin signaling for docking compounds in BRET-biosensor based assays in HEK293T cells expressing the human  $\alpha_{2A}$ AR. All compound responses normalized to norepinephrine response.

<sup>a</sup> Endogenous G<sub>i/o</sub> protein family activation in the absence of heterologously expressed G proteins.

--, not determined

**Table 2.S4 EMTA coupling panel relative activities of '9087 and its analogs in comparison to norepinephrine, dexmedetomidine, and brimonidine against the human  $\alpha_{2A}$ AR.**

Transducer	Compound functional data					
	Mean log(Emax/EC50)	SEM log(Emax/EC50)	Mean for $\Delta$ log(Emax/EC50)	SEM for $\Delta$ log(Emax/EC50)	unpaired t-test compared to NE <sup>b</sup>	RE: 10 <sup>^</sup> $\Delta$ log(Emax/EC50)
<b>Endogenous Gi/o<sup>a</sup></b>						
Norepi	10.64	0.11	0.00	0.16		1.00
Dex	12.00	0.07	1.36	0.13	***	23.17
Brimonidine	11.87	0.05	1.23	0.13	**	16.97
'9087	9.32	0.09	-1.32	0.15	**	0.05
'7075	9.69	0.06	-0.95	0.13	**	0.11
PS75	9.98	0.04	-0.66	0.12	*	0.22
<b>Gi1</b>						
Norepi	10.66	0.16	0.00	0.23		1.00
Dex	11.93	0.02	1.27	0.16	**	18.71
Brimonidine	11.96	0.02	1.30	0.16	**	20.17
'9087	8.99	0.12	-1.67	0.20	**	0.02
'7075	9.45	0.03	-1.21	0.17	**	0.06
PS75	9.94	0.03	-0.71	0.17	*	0.19
<b>Gi2</b>						
Norepi	10.75	0.12	0.00	0.18		1.00
Dex	12.01	0.03	1.27	0.13	**	18.41
Brimonidine	12.01	0.01	1.26	0.12	***	18.17
'9087	9.01	0.14	-1.74	0.19	***	0.02
'7075	9.49	0.03	-1.26	0.13	**	0.05
PS75	9.94	0.04	-0.81	0.13	**	0.16
<b>Gi3</b>						
Norepi	10.49	0.13	0.00	0.18		1.00
Dex	11.93	0.04	1.44	0.13	***	27.34
Brimonidine	11.90	0.02	1.41	0.13	***	25.87
'9087	8.92	0.13	-1.57	0.18	**	0.03
'7075	9.42	0.05	-1.07	0.14	**	0.08
PS75	9.85	0.05	-0.64	0.14	*	0.23
<b>GoA</b>						
Norepi	11.08	0.20	0.00	0.28		1.00
Dex	12.39	0.03	1.31	0.20	**	20.39
Brimonidine	12.35	0.05	1.27	0.21	*	18.51
'9087	9.58	0.07	-1.50	0.21	**	0.03
'7075	9.93	0.02	-1.14	0.20	*	0.07
PS75	10.39	0.03	-0.68	0.20	ns	0.21
<b>GoB</b>						
Norepi	11.16	0.13	0.00	0.19		1.00
Dex	12.59	0.05	1.44	0.14	***	27.41
Brimonidine	12.50	0.03	1.34	0.14	**	21.96
'9087	9.75	0.07	-1.41	0.15	**	0.04
'7075	10.08	0.03	-1.07	0.14	**	0.08
PS75	10.59	0.04	-0.57	0.14	*	0.27
<b>Gz</b>						
Norepi	12.16	0.19	0.00	0.27		1.00
Dex	13.49	0.03	1.33	0.19	**	21.30
Brimonidine	13.67	0.03	1.51	0.19	**	32.35
'9087	10.86	0.07	-1.30	0.20	**	0.05
'7075	10.91	0.03	-1.26	0.19	**	0.06
PS75	11.43	0.03	-0.74	0.19	ns	0.18
<b>Gs</b>						
Norepi	8.49	0.12	0.00	0.17		1.00
Dex	10.00	0.36	1.52	0.38	**	32.93
Brimonidine	9.77	0.06	1.29	0.13	***	19.40
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--

Transducer	Compound functional data					
	Mean log(Emax/EC50)	SEM log(Emax/EC50)	Mean for $\Delta\log(\text{Emax/EC50})$	SEM for $\Delta\log(\text{Emax/EC50})$	unpaired t-test compared to NE <sup>b</sup>	RE: $10^{\Delta\log(\text{Emax/EC50})}$
<b>G12</b>						
Norepi	8.65	0.11	0.00	0.16		1.00
Dex	9.72	0.02	1.07	0.11	***	11.83
Brimonidine	9.61	0.03	0.96	0.11	**	9.11
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>G13</b>						
Norepi	8.08	0.04	0.00	0.05		1.00
Dex	9.22	0.03	1.13	0.05	***	13.60
Brimonidine	9.18	0.07	1.10	0.08	***	12.50
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>Gq</b>						
Norepi	7.96	0.04	0.00	0.06		1.00
Dex	8.66	0.18	0.70	0.18	**	4.97
Brimonidine	8.86	0.03	0.90	0.05	***	7.90
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>G11</b>						
Norepi	7.88	0.08	0.00	0.11		1.00
Dex	9.02	0.16	1.14	0.18	***	13.71
Brimonidine	8.33	0.19	0.45	0.20	ns	2.80
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>G14</b>						
Norepi	7.76	0.08	0.00	0.12		1.00
Dex	9.04	0.15	1.28	0.17	***	18.95
Brimonidine	8.83	0.06	1.07	0.10	***	11.70
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>G15</b>						
Norepi	8.29	0.07	0.00	0.10		1.00
Dex	9.46	0.03	1.17	0.07	***	14.91
Brimonidine	9.45	0.01	1.16	0.07	***	14.60
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>barrestin1</b>						
Norepi	7.78	0.07	0.00	0.09		1.00
Dex	9.35	0.08	1.57	0.10	***	37.50
Brimonidine	8.13	0.14	0.36	0.15	ns	2.28
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--
<b>barrestin2</b>						
Norepi	7.94	0.04	0.00	0.05		1.00
Dex	9.19	0.06	1.25	0.08	***	17.70
Brimonidine	8.69	0.05	0.75	0.06	***	5.58
'9087	--	--	--	--	--	--
'7075	--	--	--	--	--	--
PS75	--	--	--	--	--	--

G-protein and  $\beta$ -Arrestin signaling for docking compounds in BRET-biosensor based assays in HEK293T cells expressing the human  $\alpha_{2A}$ AR (logEC50 and Emax values located in Table S3). All compound responses normalized to norepinephrine response.

<sup>a</sup> Endogenous G<sub>10</sub> protein family activation in the absence of heterologously expressed G proteins.

<sup>b</sup> Statistical significance of all comparisons (Mean  $\Delta\log(\text{Emax/EC50})$ ) by unpaired t-test. ns = not significant, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

NE, norepinephrine

RE, relative efficacy

--, not determined



**Table 2.S5 Summary of G<sub>ai</sub> BRET activation data for compounds and α<sub>2A</sub>AR mutations.**

receptor	NorEpi		DEX		'9087		'7075		'4622		
	activity	n	activity	n	activity	n	activity	n	activity	n	
G <sub>ai</sub> <sup>a,b</sup>											
wt	EC <sub>50</sub>	5.0 ± 0.94	17	0.077 ± 0.015	7	52 ± 24	9	4.1 ± 1.2	9	9.7±2.4	6
	E <sub>max</sub>	<b>100</b>		105 ± 4		60 ± 3		93 ± 5		74±4	
D128 <sup>3.32</sup> A	EC <sub>50</sub>	--	4	13000 ± 2600	4	16000 ± 5800	5	5600 ± 1500	4	--	4
	E <sub>max</sub>	<b>50 @1mM</b>		<b>100</b>		72 ± 5		76 ± 6		<5	
D128 <sup>3.32</sup> T	EC <sub>50</sub>	--	3	--	4	14000 ± 3100	7	6600 ± 1600	5	--	4
	E <sub>max</sub>	<b>20 @1mM</b>		<b>10 @0.1mM</b>		59 ± 7		<b>100</b>		<5	
D128 <sup>3.32</sup> L	EC <sub>50</sub>	--	3	--	4	26000 ± 6400	5	3900 ± 700	5	--	4
	E <sub>max</sub>	<b>20 @1mM</b>		<b>10 @0.1mM</b>		113 ± 9		<b>100</b>		<5	
S215 <sup>5.42</sup> A	EC <sub>50</sub>	5600 ± 1700	6	4.4 ± 1.6	6	170 ± 66	4	0.43 ± 0.095	7	29±15	4
	E <sub>max</sub>	<b>100</b>		97 ± 5	61±5	61 ± 5		77 ± 4		73±7	
Y409 <sup>6.55</sup> A	EC <sub>50</sub>	41000 ± 17000	5	16 ± 6.4	5	300 ± 79	6	690 ± 460	6	--	4
	E <sub>max</sub>	<b>100</b>		95 ± 2		64 ± 7		11 ± 6		<5	
Y409 <sup>6.55</sup> T	EC <sub>50</sub>	52000 ± 17000	6	50 ± 28	5	510 ± 360	4	5300 ± 2400	4	--	3
	E <sub>max</sub>	<b>100</b>		100 ± 2		80 ± 9		36 ± 3		<5	
Y409 <sup>6.55</sup> F	EC <sub>50</sub>	2300 ± 1100	6	4.3 ± 2.0	3	500 ± 290	4	3200 ± 1400	5	--	4
	E <sub>max</sub>	<b>100</b>		62 ± 8		73 ± 7		17 ± 3		<5	
F427 <sup>7.39</sup> A	EC <sub>50</sub>	420 ± 160	4	330 ± 180	3	--	5	--	3	--	4
	E <sub>max</sub>	<b>100</b>		87 ± 7		11 @0.1mM		<5		<5	
Y431 <sup>7.43</sup> A	EC <sub>50</sub>	35000 ± 6400	4	850 ± 330	4	26000 ± 4700	5	3900 ± 1500	3	9300±4700	4
	E <sub>max</sub>	<b>100</b>		70 ± 1		54 ± 10		42 ± 9		13±8	
Y431 <sup>7.43</sup> F	EC <sub>50</sub>	430 ± 61	5	350 ± 89	6	--	5	120 ± 43	6	--	4
	E <sub>max</sub>	<b>100</b>		63 ± 4		<5		42 ± 6		<5	

<sup>a</sup> G-protein activation monitored in a BRET-biosensor based assay in HEK293T cells expressing the human α<sub>2A</sub>AR performed in duplicates. Mean EC<sub>50</sub> values are displayed as in nM±s.e.m.

<sup>b</sup> Normalization was done relative to the maximum effect of NorEpi with the exception of mutants D128<sup>3.32</sup>A (DEX), D128<sup>3.32</sup>T and D128<sup>3.32</sup>L ('7075) highlighted in bold and displayed in %±s.e.m.

NorEpi, norepinephrine

DEX, dexmedetomidine

--, not determined

**Table 2.S6 Summary of G<sub>ai</sub> BRET activation and arrestin BRET recruitment data for selected compounds at the α<sub>2A</sub>AR and selected mutants displayed as delta BRET values.**

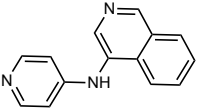
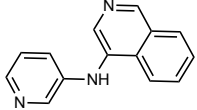
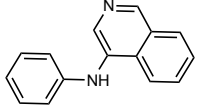
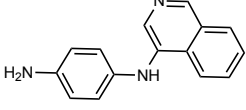
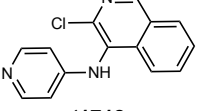
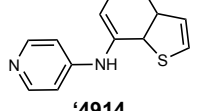
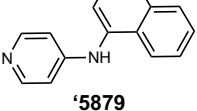
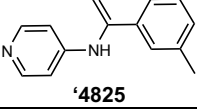
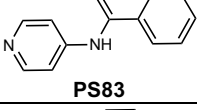
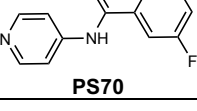
receptor		NorEpi		DEX		'9087		'7075		'4622	
		activity	n	activity	n	activity	n	activity	n	activity	n
Gα <sup>a</sup>											
wt	EC <sub>50</sub>	5.0 ± 0.87	17	0.082 ± 0.015	7	56 ± 27	9	6.9 ± 3.0	9	17 ± 6.4	6
	E <sub>max</sub>	0.270 ± 0.006		0.249 ± 0.012		0.164 ± 0.011		0.237 ± 0.010		0.173 ± 0.009	
D128 <sup>3,32A</sup>	EC <sub>50</sub>	--	4	12000 ± 1800	4	12000 ± 5700	5	5300 ± 1300	4	--	4
	E <sub>max</sub>	0.082 ± 0.013 @ 1mM		0.175 ± 0.024		0.140 ± 0.009		0.134 ± 0.011		<0.05	
S215 <sup>5,42A</sup>	EC <sub>50</sub>	5200 ± 1600	6	4.4 ± 1.6	6	240 ± 120	4	0.37 ± 0.18	7	14 ± 7.0	4
	E <sub>max</sub>	0.311 ± 0.039		0.259 ± 0.023		0.209 ± 0.021		0.257 ± 0.026		0.299 ± 0.034	
Y409 <sup>6,55A</sup>	EC <sub>50</sub>	15000 ± 3800	5	16 ± 6.6	5	290 ± 70	6	--	6	--	4
	E <sub>max</sub>	0.414 ± 0.036		0.362 ± 0.019		0.241 ± 0.030		<0.05		<0.05	
Y409 <sup>6,55F</sup>	EC <sub>50</sub>	2000 ± 940	6	4.4 ± 2.0	3	600 ± 360	4	5200 ± 2600	5	--	4
	E <sub>max</sub>	0.326 ± 0.038		0.240 ± 0.038		0.267 ± 0.030		0.089 ± 0.048		<0.05	
F427 <sup>7,39A</sup>	EC <sub>50</sub>	410 ± 96	4	2800 ± 1400	4	--	5	--	3	--	4
	E <sub>max</sub>	0.350 ± 0.023		0.332 ± 0.021		<0.05		<0.05		<0.05	
Y431 <sup>7,43A</sup>	EC <sub>50</sub>	47000 ± 9100	4	1300 ± 350	4	16000 ± 2600	5	2400 ± 1000	3	--	4
	E <sub>max</sub>	0.290 ± 0.031		0.199 ± 0.024		0.153 ± 0.044		0.148 ± 0.043		<0.05	
Y431 <sup>7,43F</sup>	EC <sub>50</sub>	540 ± 120	5	360 ± 97	6	--	5	120 ± 44	6	--	4
	E <sub>max</sub>	0.322 ± 0.018		0.207 ± 0.019		<0.05		0.140 ± 0.014		<0.05	
β-arrestin-2 recruitment <sup>b</sup>											
wt	EC <sub>50</sub>	400 ± 30	12	3.0 ± 1.2	8	--	8	--	9	--	7
	E <sub>max</sub>	0.946 ± 0.074		0.522 ± 0.034		<0.05		0.065 ± 0.032 @ 100μM		0.051 ± 0.010 @ 100μM	
D128 <sup>3,32A</sup>	EC <sub>50</sub>	--	5	--	3	--	4	--	4	--	4
	E <sub>max</sub>	<0.05		<0.05		<0.05		<0.05		<0.05	
S215 <sup>5,42A</sup>	EC <sub>50</sub>	--	6	3.5 ± 1.0	5	--	3	37 ± 8.6	7	110 ± 39	7
	E <sub>max</sub>	0.615 ± 0.023 @ 1mM		0.523 ± 0.025		<0.05		0.195 ± 0.019		0.144 ± 0.009	
Y409 <sup>6,55A</sup>	EC <sub>50</sub>	--	5	44 ± 11	5	--	5	--	5	--	4
	E <sub>max</sub>	0.383 ± 0.020 @ 1mM		0.527 ± 0.013		<0.05		<0.05		<0.05	
Y409 <sup>6,55F</sup>	EC <sub>50</sub>	8300 ± 1500	4	8.2 ± 1.0	4	--	3	120000 ± 71000	4	--	3
	E <sub>max</sub>	0.756 ± 0.007		0.420 ± 0.012		<0.05		0.104 ± 0.023		<0.05	
F427 <sup>7,39A</sup>	EC <sub>50</sub>	25000 ± 2700	6	3200 ± 810	5	--	5	--	5	--	5
	E <sub>max</sub>	0.848 ± 0.045		0.252 ± 0.027		<0.05		<0.05		<0.05	
Y431 <sup>7,43A</sup>	EC <sub>50</sub>	250000 ± 37000	6	--	4	--	4	--	4	--	4
	E <sub>max</sub>	0.941 ± 0.025		<0.05		<0.05		<0.05		<0.05	
Y431 <sup>7,43F</sup>	EC <sub>50</sub>	21000 ± 3800	5	49 ± 12	4	--	3	52000 ± 28000	5	--	3
	E <sub>max</sub>	0.234 ± 0.011		0.195 ± 0.009		<0.05		0.098 ± 0.015		<0.05	

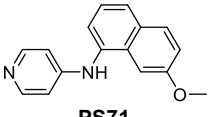
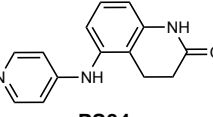
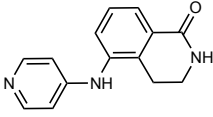
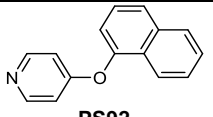
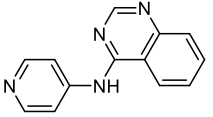
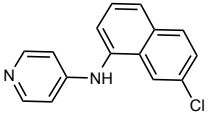
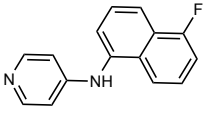
<sup>a</sup> G-protein activation monitored in a BRET-biosensor based assay in HEK293T cells expressing the human α<sub>2A</sub>AR or a mutant performed in duplicates. Mean EC<sub>50</sub> values are displayed as nM±s.e.m.; E<sub>max</sub> values are listed as absolute mean values±s.e.m. of delta BRET with numbers of experiments (n) listed separately.

<sup>b</sup> β-Arrestin-2 recruitment was determined in a BRET-biosensor based assay in HEK293T cells expressing the human α<sub>2A</sub>AR or a mutant performed in duplicates. Mean EC<sub>50</sub> values are displayed as nM±s.e.m.; E<sub>max</sub> values are listed as mean values±s.e.m. of delta BRET; numbers of experiments (n) are specified individually.

NorEpi, norepinephrine  
DEX, dexmedetomidine  
--, not determined

**Table 2.S7 Binding affinity and functional properties of selected '9087 analogs to the human  $\alpha_2AAR$ .**

Compound	$K_i$ (nM) <sup>a</sup> n	$G_i$ EC <sub>50</sub> ± SE <sup>b</sup> E <sub>max</sub> ± SEM <sup>b</sup> Mean logEC <sub>50</sub> (SEM) <sup>c</sup> Relative activity <sup>d</sup> n	$\beta$ -Arr-2 <sup>e</sup> EC <sub>50</sub> E <sub>max</sub> n	Statistical differences for $G_i$ activity <sup>f</sup>
 <b>'9087</b>	51 ± 8.8 6	52 nM ± 24 60% ± 3 -7.585 (0.169) 9.354 (0.171) 9	--nM <5% 6	'5879 * PS84 **** PS86 **** PS75 ** '7075 ****
 <b>'3188</b>	>10000 3	--nM <5% ND ND 4	--	--
 <b>'3535</b>	>10000 3	--nM 57% ± 9 @0.1mM ND ND 4	--	--
 <b>'9465</b>	4600 ± 2100 3	--nM <5% ND ND 3	--	--
 <b>'1718</b>	22 ± 6.7 3	14 nM ± 4.8 47% ± 2 -8.180 (0.296) 9.845 (0.294) 7	--nM <5% 3	PS84 **** PS86 ****
 <b>'4914</b>	7.2 ± 1.0 4	9.1 nM ± 3.2 77% ± 2 -8.175 (0.183) 10.058 (0.184) 5	--nM <5% 3	PS71 * PS84 **** PS86 ****
 <b>'5879</b>	8.8 ± 3.2 3	6.1 nM ± 1.2 84% ± 6 -8.251 (0.094) 10.170 (0.117) 5	--nM 6% ± 1 @0.1mM 4	'9087 * PS71 ** PS84 **** PS86 ****
 <b>'4825</b>	19 ± 1.5 4	18 nM ± 5 94% ± 4 -7.850 (0.098) 9.823 (0.091) 9	--nM 18% ± 9 @0.1mM 4	PS71 * PS84 **** PS86 ****
 <b>PS83</b>	72 ± 15 3	13 nM ± 2.7 94% ± 9 -7.899 (0.093) 9.868 (0.073) 3	--nM ±	PS84 **** PS86 ****
 <b>PS70</b>	9.2 ± 0.99 3	36 nM ± 14 91% ± 3 -7.615 (0.175) 9.570 (0.171) 6	--nM <5% 4	PS84 ** PS86 **** '7075 **

Compound	$K_i$ (nM) <sup>a</sup> n	$G_i$ EC <sub>50</sub> ± SE <sup>b</sup> E <sub>max</sub> ± SEM <sup>b</sup> Mean logEC <sub>50</sub> (SEM) <sup>c</sup> Relative activity <sup>d</sup> n	β-Arr-2 <sup>e</sup> EC <sub>50</sub> E <sub>max</sub> n	Statistical differences for $G_i$ activity <sup>f</sup>
 <b>PS71</b>	240 ± 120 3	100 nM ± 19 85% ± 3 -7.020 (0.100) 8.948 (0.113) 4	--nM <5% 4	'4914 * '5879 ** '4825 ** PS84 * PS86 ** PS75 *** '7075 ****
 <b>PS84</b>	500 ± 87 4	840 nM ± 350 37% ± 4 -6.189 (0.250) 7.752 (0.255) 3	--	'9087 **** '1718 **** '4914 **** '5879 **** '4825 **** PS83 **** PS70 **** PS71 * PS75 **** '7075 ****
 <b>PS86</b>	1200 ± 560 3	3400 nM ± 940 91% ± 15 -5.494 (0.109) 7.442 (0.155) 3	--	'9087 **** '1718 **** '4914 **** '5879 **** '4825 **** PS83 **** PS70 **** PS71 ** PS75 **** '7075 ****
 <b>PS92</b>	1800 ± 730 4	--nM 43% ± 4 @ 0.1mM ND ND 5	--	--
 <b>PS93</b>	6600 ± 1800 4	--nM <5% ND ND 6	--	--
 <b>PS75</b>	8.2 ± 0.48 4	4.8 nM ± 1.3 82% ± 4 -8.441 (0.148) 10.352 (0.149) 7	-- nM 15% ± 3 @ 0.1mM 6	'9087 ** PS71 *** PS84 **** PS86 ****
 <b>'7075</b>	37 ± 5.1 5	4.1 nM ± 1.2 93% ± 5.2 -8.509 (0.117) 10.471 (0.095) 9	-- nM 10% ± 7 @ 0.1mM 7	'9087 **** PS70 ** PS71 **** PS84 **** PS86 ****

<sup>a</sup> Binding affinity to human  $\alpha_{2A}$ AR as mean  $K_i$  value in [nM] ± SEM

<sup>b</sup> G-protein activation monitored in a BRET-biosensor based assay using the human  $\alpha_{2A}$ AR

<sup>c</sup> logEC<sub>50</sub> (means calculated from each replicate) of G-protein signaling monitored in a BRET-biosensor based assay using the human  $\alpha_{2A}$ AR

<sup>d</sup> Relative activity, log(E<sub>max</sub>/EC<sub>50</sub>), of G-protein signaling monitored in a BRET-biosensor based assay using the human  $\alpha_{2A}$ AR

<sup>e</sup> β-Arrestin-2 recruitment measured with a BRET-biosensor based assay and the human  $\alpha_{2A}$ AR

<sup>f</sup> Differences of relative compound activities (log(E<sub>max</sub>/EC<sub>50</sub>)) were analyzed by One-way ANOVA applying Tukey's multiple comparisons test in PRISM 9.3.1. \* p<0.05, \*\* p<0.01, \*\*\*, p<0.001; \*\*\*\*, p<0.0001. Only compounds with statistically significant differences are listed.

--, not determined

**Table 2.S8 Binding affinity of '9087 to other adrenergic receptors.**

$\alpha_{2c}AR$	$\alpha_{1A}AR$	$\alpha_{1B}AR$	$\beta_1AR$	$\beta_2AR$
$0.027 \pm 0.004 \mu M$ (3)	$1.3 \pm 0.37 \mu M$ (4)	$7.4 \pm 0.64 \mu M$ (3)	$>10 \mu M$ (3)	$>10 \mu M$ (3)

Binding affinity to related adrenergic receptors as mean  $K_i$  value in [ $\mu M \pm$  s.e.m.] each done in triplicates; number of individual experiments are displayed in parentheses.

**Table 2.S9 Pharmacokinetic properties of docking agonists.**

Compound	Routes (dose)	PK parameters <sup>a,b,c</sup>	Plasma	Brain	CSF	Oral bioavailability <sup>d</sup>	Permeability prediction (kcal/mol) <sup>e</sup>
<b>'9087</b>	I.P. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	1440 110000 0.50	5960 439000 9.85	447 154000 26.83	152% <sup>f</sup>	-2.207
	I.V. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	2020 105000 0.87	6590 420000 0.62	ND		
	P.O. (30 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	2720 506000 1.25	14000 2540000 1.06	ND		
<b>'7075</b> ( <b>'9087</b> analog)	I.P. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	694 57100 0.64	8570 814000 0.73	412 12000 0.22	ND	-3.238
<b>PS75</b> ( <b>'9087</b> analog)	I.P. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	679 55600 0.73	9200 888000 1.07	99.3 7790 2.55	102% <sup>f</sup>	ND
	I.V. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	2586 207,840 1.64	ND	ND		
	P.O. (30 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	2390 637,140 3.17	ND	ND		
<b>'4622</b>	I.P. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	620 34000 1.01	16200 1160000 0.88	44.2 18400 14.5	ND	-3.012
<b>'2998</b>	I.P. (10 mg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	2710 184000 0.4	742 51700 0.46	n/a	ND	-1.121
<b>Dexmedetomidine</b>	I.P. (30 µg/kg)	Cmax (ng/ml) AUC (ng*min/mL) T ½ (hr)	ND	8 285 ND	ND	ND	ND

<sup>a</sup>Cmax, maximum concentration reached in mice plasma, brain, or CSF sample

<sup>b</sup>AUC, area under the concentration-time curve for exposure in mice plasma, brain, or CSF

<sup>c</sup>T ½, half-life of the compound in mice plasma, brain, or CSF

<sup>d</sup>oral bioavailability, see methods for calculation using I.V. and P.O. dosing

<sup>e</sup>Delta-delta-G of insertion of compounds from water to membrane

<sup>f</sup>An oral bioavailability higher than 100% can occur with certain compound and pharmacokinetic conditions, including "non-linear PK" (metabolic saturation) at non-equal I.V. and P.O. doses and entero-hepatic recirculation.

**Table 2.S10 Cryo-EM data collection, refinement, and validation statistics.**

<b>Data collection and processing</b>	'9087- $\alpha_{2A}$ AR-GoA-scfv16 PDB:7W6P EMDB:EMD-32331	'4622- $\alpha_{2A}$ AR-GoA-scfv16 PDB:7W7E EMDB:EMD-32342
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e/ $\text{\AA}$ )	60	60
Defocus range ( $\mu\text{m}$ )	-1.0~-1.6	-1.0~-1.8
Pixel size ( $\text{\AA}$ )	0.85	0.85
Symmetry imposed	C1	C1
Initial particle images (no.)	2137146	2593747
Final particle images (no.)	287431	563506
Map resolution ( $\text{\AA}$ )	3.47	3.38
FSC threshold	0.143	0.143
<b>Refinement</b>		
Initial model used (PDB code)	7EJ0	7EJ0
Map sharpening B-factor ( $\text{\AA}$ )	-120	-80
<b>Model composition</b>		
Non-hydrogen atoms	8429	8524
Protein residues	1083	1090
<b>B factor (<math>\text{\AA}</math>)</b>		
Protein	47.16	92.66
Ligand	50.31	105.49
<b>R.m.s. deviations</b>		
Bond lengths ( $\text{\AA}$ )	0.004	0.004
Bond angles ( $^\circ$ )	0.642	0.653
<b>Validation</b>		
MolProbity score	2.11	1.95
Clashscore	14.17	8.09
<b>Ramachandran plot</b>		
Favored (%)	93.03	91.2
Allowed (%)	6.97	8.8
Disallowed (%)	0	0

## 2.7 Materials and Methods

### Molecular docking

The  $\alpha_{2B}$ AR receptor with dexmedetomidine and  $G_{\alpha A}$  (PDB 6K41) (9) was used for docking calculations prior to the determination of the  $\alpha_{2B}$ AR dexmedetomidine-bound structure (PDB 7EJA; comparison shown in **Fig. S1**). Three screens of the ZINC15 database were run, two for fragment molecules (less than 250 amu,  $cLogP \leq 3.5$ ) and one for lead-like (250-350 amu,  $cLogP \leq 3.5$ ). Docking was performed with DOCK3.7(26). For the first screen, 45 matching spheres (26) were used, 15 from the docked pose of dexmedetomidine and 30 from SPHGEN-generated spheres(58). The receptor structure was protonated using REDUCE (59) and AMBER united atom charges were assigned (60). In control calculations (61) with 15 known agonists from the IUPHAR/BPS database (27) and from the literature (29–31), balanced against 1800 property matched decoys (62), docking parameters were optimized based on adjusted logAUC (61) and based on recapitulation of ligand interactions with residues  $\alpha_{2B}$ AR D92<sup>3,32</sup>, F412<sup>7,39</sup>, F387<sup>6,51</sup>, Y391<sup>6,55</sup>, and F388<sup>6,52</sup> (residues conserved in  $\alpha_{2A}$ AR: D128<sup>3,32</sup>, F427<sup>7,39</sup>, F405<sup>6,51</sup>, Y409<sup>6,55</sup>, and F406<sup>6,52</sup>). An “extrema” set was used to evaluate cationic charge preference, as described (18, 62). The protein low dielectric and desolvation regions, defined by pseudo-atoms calculated with SPHGEN (58), were extended as previously described (63), based on the control calculations, by a radius of 1 Å and 0.3 Å, respectively (10, 64). Energy potential grids were calculated using CHEMGRID for AMBER-based van der Waals potential, QNIFFT (65) for Poisson-Boltzmann-based electrostatic potentials, and SOLVMAP (66) for context-dependent ligand desolvation. In the second and third docking screens, differences included modified matching spheres (added rigid fragments of xylazine docked-pose, only used 40 matching spheres) and extension of the desolvation pseudo-atoms by a radius of 0.2 Å.



For the first screen, 20 million molecules from the ZINC15 (<http://zinc15.docking.org/>) fragments subset were docked in 3,008 core hours or about 6 wall-clock hours on a 500-core cluster. Almost 5 trillion complexes were sampled, on average each molecule sampled 2,405 orientations and 202 conformations. Only about 8 of 20 million could be sterically accommodated in the orthosteric site, reflecting its small size. For the second screen, the same 20 million fragments were docked in 3830 core hours or 7.7 hours on 500-core cluster, sampling over 6 trillion complexes; on average each molecule sampled 3,122 orientations and 203 conformations. About 9 million molecules were accommodated in the site. For the third screen, 281 million molecules from ZINC15 lead-like subset were screened in 71,625 core hours or about 1 week on 500 cores. Over 222 trillion complexes were sampled with an average of 4,553 orientations and 469 conformations per molecule, though ultimately only 13.5 million could sterically fit in the site.

For the first and second screens, the top 161,055 scored compounds were clustered by ECFP4-based Tanimoto coefficient (Tc) of 0.5 to identify unique chemotypes, resulting in 37,150 and 33,378 clusters. For the third screen, the top 300,000 scored compounds were clustered in the same manner resulting in 57,168 clusters. Molecules were filtered for novelty, removing those with Tc > 0.35 to 15  $\alpha_{2A}$ AR agonists used in control calculations. The top 5,000 ranked molecules remaining were visually filtered for interactions at  $\alpha_{2B}$ AR residues D92<sup>3.32</sup>, F412<sup>7.39</sup>, F387<sup>6.51</sup>, Y391<sup>6.55</sup>, and F388<sup>6.52</sup> for the first and second screens; for the third screen, the top 20,000 molecules were examined by the same criteria. Lastly, prioritized molecules were also filtered for internal torsional strain; this was done visually for the first screen, while the second and third screens used a method drawing on CSD torsion populations cutting off at a total energy of 2 Torsion Energy

Units (32). An additional novelty filter was performed removing molecules with TC > 0.35 to ChEMBL29 (28)  $\alpha_{2A}$ AR compounds. Sixty-four molecules were selected for purchasing: 33, 26, and 5 from the first, second and third screens, respectively. Ten were sourced from WuXi and another 54 from Enamine, of which 8 and 40 were successfully synthesized, respectively. Most of these compounds have not previously been synthesized before, to the best of our knowledge, except for some of the smaller fragments, which have been previously used as building blocks ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### **Synthesis of tangible molecules**

Forty-eight molecules prioritized for purchasing were synthesized by Enamine and Wuxi for a total fulfillment rate of 75%. Compounds were sourced from the Enamine REAL database (<https://enamine.net/compound-collections/real-compounds>) or the WuXi GalaXi Virtual library. The purities of active molecules synthesized by Enamine and WuXi were at least 90% and typically above 95%. For bespoke compound synthesized in house purities were at least 96% and typically above 99%. The purity of compounds tested *in vivo* were >95% and typically above 98% ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### **Ligand optimization**

Analogues for four docking hits ('9087', '2998', '0172', '4622) were queried in Arthor and SmallWorld 1.4 and 12 Billion tangible libraries (<https://sw.docking.org/>, <https://arthor.docking.org>), the latter primarily containing Enamine REAL Space compounds (<https://enamine.net/compound-collections/real-compounds/real-space-navigator>). Results from SmallWorld, Bemis-Murcko framework, and substructure queries were pooled, docked into the  $\alpha_{2B}$ AR site prior to '9087- $\alpha_{2A}$ AR structure being determined. Compounds with favorable interactions in the orthosteric site

were prioritized, leading to 13 analogs for **'9087** ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Also, for the 4 docking hits, analogs were designed by modifying the 2D chemical structure to test specific hypotheses, adding another 6 analogs for **'9087** ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). The second round of analogs for **'9087** were designed and prioritized for bespoke synthesis. Some were docked to a preliminary cryo-EM model of the **'9087- $\alpha_{2A}$ AR** structure, while several were designed and synthesized regardless of docked pose to test specific hypotheses; in total 8 of these were synthesized and tested ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Calculation of the contact areas was performed by means of UCSF Chimera (<https://rbvi.ucsf.edu/chimera>).

### **Bespoke synthesis**

**'7075** was designed for hypothesis testing and was bespoke synthesized by Enamine ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### **Molecular modeling of '7075 and PS75**

Maestro (v. 2019-4, Schrödinger, LLC) was used to manually change the chemical structure of **'9087** to **'7075** or **PS75** in a preliminary model of the **'9087- $\alpha_{2A}$ AR** complex cryo-EM structure. The isoquinoline nitrogen was changed to a carbon and the fluorine or chlorine substituent was added to the naphthalene ring for **'7075** and **PS75**, respectively. The resulting complex of **'7075** or **PS75** and  $\alpha_{2A}$ AR coupled to the G protein but without scFv16 was energy minimized following the Protein Preparation Wizard protocol using the OPLS3e force field. The maximum heavy atom deviation from the initial model was 0.3Å.

### **Passive-membrane permeability prediction**

Forty Ligand structures were converted from SMILES strings to three-dimensional structures using LigPrep (v. 53013, Schrodinger, New York). For the passive-membrane permeability prediction (67, 68), we retained only neutral form for each ligand. Passive-membrane permeability of a ligand is predicted from the free-energy of insertion ( $\Delta G_I$ ), i.e., from the energy difference between a conformer in low and high dielectric media. Therefore, we generated conformations of each ligand using ConfGen software (v. 5.1, Schrodinger, New York). We minimized each conformer in a low dielectric medium (chloroform) to mimic the membrane dielectric using Protein Local Optimization Program (PLOP) (69). After finding the lowest energy conformer in the low dielectric medium, we calculated the energy of that energy-minimized conformer in water. We subtracted the energy of the ligand in the high-dielectric water from the low-dielectric medium. We further added a deionization penalty term to account for transforming ionized form of the ligand in water to its neutral form in membrane. We computed the deionization penalty energy using the empirical pKa prediction software Epik (v. 5.1013, Schrodinger Inc.). We rank-ordered the ligands based on their free-energy of insertion.

### **Radioligand binding experiments.**

Receptor binding affinities for the  $\alpha_{2A}$ AR receptor and to  $\alpha_{2B}$ AR as well as the related adrenergic subtypes  $\alpha_{1A}$ ,  $\alpha_{1B}$ ,  $\alpha_{2C}$ ,  $\beta_1$  and  $\beta_2$  were determined as described previously (44, 70). In brief, membranes were prepared from HEK293T cells transiently transfected with the cDNA for human  $\alpha_{2A}$ AR, murine  $\alpha_{2A}$ AR (provided by D. Calebiro, Birmingham, UK), human  $\alpha_{2B}$ AR (obtained from the cDNA resource center, [www.cdna.org](http://www.cdna.org)) or with the cDNAs for the human  $\alpha_{1A}$ ,  $\alpha_{1B}$ ,  $\alpha_{2C}$ ,  $\beta_2$  (all from cDNA resource center) and  $\beta_1$  (provided by R. Sunahara, UCSD). Receptor densities ( $B_{\max}$

value) and specific binding affinities ( $K_D$  value) for the radioligand [ $^3\text{H}$ ]RX82,1002 (specific activity 52 Ci/mmol, Novandi, Södertälje, Sweden) were determined as  $1,400 \pm 210$  fmol/mg protein and  $0.54 \pm 0.024$  nM for human  $\alpha_{2A}\text{AR}$ ,  $4,000 \pm 720$  fmol/mg protein and  $1.8 \pm 0.61$  nM for murine  $\alpha_{2A}\text{AR}$ , and  $3,400 \pm 580$  fmol/mg protein and  $2.3 \pm 0.52$  nM for  $\alpha_{2B}\text{AR}$ , respectively. Further values are  $3,200 \pm 1,900$  fmol/mg protein and  $0.58 \pm 0.11$  nM for  $\alpha_{2C}$ ,  $2,000 \pm 950$  fmol/mg protein and  $0.70 \pm 0.13$  nM for  $\alpha_{1A}$ , and  $4,000$  fmol/mg protein and  $0.11$  nM for  $\alpha_{1B}$ , both determined with [ $^3\text{H}$ ]prazosin (51 Ci/mmol, PerkinElmer, Rodgau, Germany), respectively and  $1,400 \pm 360$  fmol/mg protein and  $0.070 \pm 0.006$  nM for  $\beta_1$ , and  $1,300 \pm 230$  fmol/mg protein and  $0.074 \pm 0.012$  nM for  $\beta_2$ , both determined with [ $^3\text{H}$ ]CGP12,188 (52 Ci/mmol, PerkinElmer).

Competition binding with  $\alpha_2\text{AR}$  subtypes were performed by incubating membranes in buffer A (50 mM TRIS at pH 7.4) at final protein concentrations of 3-10  $\mu\text{g}/\text{well}$  with the radioligand (final concentration 0.5-2.0 nM according to the appropriate  $K_D$  and  $B_{\text{max}}$ ) and varying concentrations of the competing ligands for 60 minutes at 37 °C. Binding to  $\alpha_{1A}$  and  $\alpha_{1B}$  was measured with buffer B (50 mM TRIS, 5 mM  $\text{MgCl}_2$ , 1 mM EDTA, 100  $\mu\text{g}/\text{mL}$  bacitracin and 5  $\mu\text{g}/\text{mL}$  soybean trypsin inhibitor at pH 7.4) at 2-6  $\mu\text{g}/\text{well}$  (radioligand at 0.2-0.3 nM) and binding to  $\beta_1$  and  $\beta_2$  was measured with buffer C (25 mM HEPES, 5 mM  $\text{MgCl}_2$ , 1 mM EDTA, and 0.006% bovine serum albumin at pH 7.4) at 4-8  $\mu\text{g}/\text{well}$  (radioligand 0.2 nM). Non-specific binding was determined in the presence of unlabeled ligand at 10  $\mu\text{M}$ . Protein concentration was measured using the method of Lowry (71).

The resulting competition curves were analyzed by nonlinear regression using the algorithms implemented in Prism 8.0 (GraphPad Software, San Diego, CA) to provide  $\text{IC}_{50}$  values, which

were subsequently transformed into a  $K_i$  values applying the equation of Cheng and Prusoff (72). Mean  $K_i$  values ( $\pm$  s.e.m. for  $n \geq 3$ , or  $\pm$  s.d. for  $n = 2$ ) were derived from 2-7 experiments each performed in triplicates.

## **Functional assays**

### *Plasmids*

The human wild type  $\alpha_{2A}AR$ , its respective receptor mutants (73) and the murine  $\alpha_{2A}AR$ , all carrying an N-terminal HA-signal sequence and a FLAG-tag, as well as the human adrenergic receptor subtypes  $\alpha_{1A}$ ,  $\alpha_{1B}$ ,  $\alpha_{2C}$ ,  $\beta_1$  and  $\beta_2$  and the dopamine receptor  $D_{2long}$  were cloned to pCDNA3.1 for G protein activation assays (BRET, IP accumulation). Human  $\alpha_{2A}AR$  and  $\alpha_{2B}AR$  were fused to the ARMS2-PK2 sequence and cloned to pCMV (DiscoverX, Eurofins) for  $\beta$ -arrestin-2 recruitment assays, respectively, using polymerase chain reaction and Gibson Assembly (New England Biolabs) (70). Sequence integrity was verified by DNA sequencing (Eurofins Genomics).

### *Bioluminescence resonance energy transfer*

G protein activation by human  $\alpha_{2A}AR$  and  $D_{2long}$  was monitored with  $G\alpha_{i1}$ -RLucII (74, 75) together with  $G\beta_1$  and  $G\gamma_2$ -GFP<sub>10</sub>. Assessment of arrestin recruitment was performed by enhanced bystander BRET using CAAX-rGFP and  $\beta$ -arrestin-2-RLucII as biosensors (39, 74) in the presence of GRK2 as described (44, 76). In brief, HEK293T cells (gift from the Chair of Physiology, FAU Erlangen-Nürnberg) were transfected with 200 ng receptor plasmid for G protein activation (receptor: $G\alpha$ : $G\beta$ : $G\gamma$  ratio 2:0.5:1:4) or 100 ng receptor plasmid for  $\beta$ -arrestin recruitment (receptor: $\beta$ -arrestin:GRK2:CAAX ratio 1:0.2:1:3) using linear polyethyleneimine

(PEI, Polysciences, 3:1 PEI:DNA ratio). The DNA was complemented to a total amount of 1  $\mu$ g DNA per  $3 \cdot 10^5$  cells with ssDNA (Sigma Aldrich) and 10,000 cells per well were transferred into 96-well half-area plates (Greiner, Frickenhausen, Germany). Additional experiments were performed using the same amount of G protein plasmids as described above but 50 ng or 10 ng  $\alpha_{2A}$ AR plasmid instead. 48 h after transfection, the cell medium was exchanged with PBS (phosphate buffered saline) and cells were stimulated with ligands at 37°C for 10 min. Coelenterazine 400a (abcr GmbH, Karlsruhe, Germany) at a final concentration of 2.5  $\mu$ M was added 5 min before measurement. BRET was monitored on a Clariostar plate reader (BMG, Ortenberg, Germany) with the appropriate filter sets (donor 410/80 nm, acceptor 515/30 nm) and was calculated as the ratio of acceptor emission to donor emission. BRET ratio was normalized to the effect of buffer (0%) and the maximum effect of norepinephrine (100%) for adrenergic receptors and quinpirole (100%) for  $D_{2long}$ . For each compound 3 to 17 individual experiments were performed each done in duplicates.

Surface expression of the  $\alpha_{2A}$ AR in the G protein activation assays was monitored applying an enzyme-linked immunosorbent assay (ELISA) directed against the N-terminal FLAG tag. HEK293T cells were transfected with the cDNAs encoding  $\alpha_{2A}$ AR,  $G\alpha_{i1}$ -RLucII,  $G\beta_1$ ,  $G\gamma_2$ -GFP10 and ssDNA as described above. As a control, cells transfected with only  $\alpha_{2A}$ AR or mock pcDNA3.1 plasmid and ssDNA were used. Immediately after transfection, 50,000 cells/well were transferred to a 48-well plate (Greiner) pretreated with poly-D-lysine (Sigma Aldrich) and incubated at 37 °C and 5% CO<sub>2</sub> for 48h. The medium was removed, cells were treated with 4% paraformaldehyde for 10 min, washed once (wash buffer, 150 mM NaCl, 25 mM Tris, pH 7.5), and blocked for 60 min (30 g · L<sup>-1</sup> skim milk powder in wash buffer, all steps carried out at room

temperature). After incubation with anti-FLAG M2 mouse IgG (F3165, Sigma Aldrich, 1:4,000 in blocking solution) for 60 min, cells were washed twice, blocked again for 60 min and incubated with anti-mouse rabbit IgG-HPR (A9044, Sigma Aldrich, 1:20,000 in blocking solution) for 60 min. Cells were washed thrice, before 200  $\mu$ L substrate buffer was added (2.8 mM o-phenylenediamine, 35 mM citric acid, 66 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 5.0). Reactions were kept in the dark for 5-15 min and stopped by addition of 1 M H<sub>2</sub>SO<sub>4</sub> (200  $\mu$ L). Resulting mixtures were transferred to a 96-well plate and absorption was determined with the Clariostar microplate reader at 492 nm. Data were normalized using cells transfected with only  $\alpha_{2A}$ AR (100%) and mock pcDNA3.1 (0%), respectively. N = 4 independent experiments were performed with each condition in triplicate.

The sensitivity of selected ligands to the receptor mutants  $\alpha_{2A}$ AR-D128<sup>3.32</sup>A,  $\alpha_{2A}$ AR-D128<sup>3.32</sup>T,  $\alpha_{2A}$ AR-D128<sup>3.32</sup>L,  $\alpha_{2A}$ AR-S215<sup>5.42</sup>A,  $\alpha_{2A}$ AR-Y409<sup>6.55</sup>A,  $\alpha_{2A}$ AR-Y409<sup>6.55</sup>T,  $\alpha_{2A}$ AR-Y409<sup>6.55</sup>F,  $\alpha_{2A}$ AR-F427<sup>7.39</sup>F,  $\alpha_{2A}$ AR-Y431<sup>7.43</sup>A, and  $\alpha_{2A}$ AR-Y431<sup>7.43</sup>F was monitored by G protein activation as described above transfecting the appropriate receptor together with G $\alpha_{i1}$ -RLucII and G $\beta_1$ /G $\gamma_2$ -GFP<sub>10</sub>. Data was analyzed as ligand induced changes in BRET compared to vehicle (deltaBRET) and additionally normalization was done according to the effect of buffer (0%) and norepinephrine (100%) with the exception of  $\alpha_{2A}$ AR-D128<sup>3.32</sup>A (dexmedetomidine = 100%),  $\alpha_{2A}$ AR-D128<sup>3.32</sup>T and  $\alpha_{2A}$ AR-D128<sup>3.32</sup>L ('7075 = 100%). Similarly, the effect of the  $\alpha_{2A}$ AR-D128<sup>3.32</sup>A,  $\alpha_{2A}$ AR-S215<sup>5.42</sup>A,  $\alpha_{2A}$ AR-Y409<sup>6.55</sup>A,  $\alpha_{2A}$ AR-Y409<sup>6.55</sup>F,  $\alpha_{2A}$ AR-F427<sup>7.39</sup>F,  $\alpha_{2A}$ AR-Y431<sup>7.43</sup>A, and  $\alpha_{2A}$ AR-Y431<sup>7.43</sup>F mutations on arrestin recruitment was evaluated as described above transfecting the appropriate receptor together with CAAX-rGFP, GRK2 and  $\beta$ -arrestin-2-RLucII. Data was analyzed as ligand induced changes in BRET compared to vehicle (deltaBRET). Three to seven experiments were done in duplicate.



### *IP accumulation assay*

Determination of G protein mediated signaling by human  $\alpha_{2A}AR$ , murine  $\alpha_{2A}AR$ , and human  $\alpha_{2B}AR$  was performed applying an IP accumulation assay (IP-One HTRF®, Cisbio, Codolet, France) according to the manufacturer's protocol and in analogy to previously described protocols (77, 78). In brief, HEK 293T cells were co-transfected with the cDNA for a receptor and the hybrid G-protein  $G\alpha_{qi}$  ( $G\alpha_q$  protein with the last five amino acids at the C-terminus replaced by the corresponding sequence of  $G\alpha_i$  (gift from The J. David Gladstone Institutes, San Francisco, CA), respectively in a ratio of 1:2. After one day cells were transferred into 384 well micro plates (Greiner) and incubated for further 24 hrs. On the day of the experiment cells were incubated with test compounds for 90 min ( $\alpha_{2A}AR$ ) or 120 min ( $\alpha_{2B}AR$ ) and accumulation of second messenger was stopped by adding detection reagents (IP1-d2 conjugate and Anti-IP1 cryptate TB conjugate). After 60 min TR-FRET was monitored with a Clariostar plate reader. FRET-signals were normalized to buffer (0%) and the maximum effect of norepinephrine (100%). Three to nine (murine  $\alpha_{2A}AR$ ,  $\alpha_{2B}AR$ ) or 4-11 repeats (human  $\alpha_{2A}AR$ ), respectively in duplicate were performed for each test compound all done in duplicate.

### *PathHunter arrestin recruitment assay*

Investigation of  $\alpha_{2A}AR$  and  $\alpha_{2B}AR$  stimulated  $\beta$ -arrestin-2 recruitment was performed applying an assay which is based on fragment complementation of  $\beta$ -galactosidase (PathHunter assay, DiscoverX, Birmingham, U.K.) as described (79). In detail, HEK293T cells stably expressing the enzyme acceptor (EA) tagged  $\beta$ -arrestin-2 were co-transfected with human  $\alpha_{2A}AR$  or  $\alpha_{2B}AR$  each fused to the ProLink-ARMS2-PKS2 fragment for enzyme complementation and GRK2 (cDNA Resource Center) at equal amounts and subsequently transferred into 384 well micro plates

(Greiner) after 1 day. After incubation for further 24 hrs cells were incubated with test compounds for 60 min ( $\alpha_{2A}AR$ ) or 90 min ( $\alpha_{2B}AR$ ), arrestin recruitment was stopped by adding detection reagent and the resulting chemoluminescence was monitored with a Clariostar plate microreader. Data was normalized relative to buffer (0%) and the maximum effect of norepinephrine (100%). Three to nine repeats for  $\alpha_{2A}AR$  (3-6 for  $\alpha_{2B}AR$ ) in duplicate were measured.

### **DiscoverX HitHunter cAMP G-protein activation assay**

Dexmedetomidine, brimonidine, '9087, and '7075 were tested by DiscoverX (catalog item 86-0007P-2270AG; Eurofins; CA, USA) in their HitHunter XS+ assay. Freezer stock cAMP Hunter cell lines were expanded, then seeded in a total volume of 20  $\mu$ L into white walled, 384-well microplates and incubated at 37 °C prior to testing. For agonist determination, cells were incubated with compound samples in the presence of EC<sub>80</sub> forskolin to induce response. Media was aspirated from cells and replaced with 15  $\mu$ L 2:1 HBSS/10 mM Hepes : cAMP XS+ Ab reagent. Intermediate dilution of sample stocks was performed to generate 4X sample in assay buffer containing 4X EC<sub>80</sub> forskolin. 5  $\mu$ L of 4X sample was added to cells and incubated at 37 °C or room temperature for 30 to 60 minutes. Finally assay vehicle concentration was 1%. After sample incubation assay signal was generated through incubation with 20  $\mu$ L cAMP XS+ ED/CL lysis cocktail for one hour followed by incubation with 20  $\mu$ L cAMP XS+ EA reagent for three hours at room temperature. Microplates were read following signal generation with a PerkinElmer Envision Instrument for chemiluminescent signal detection. Compound activity was analyzed using CBIS data analysis suite. For G<sub>i</sub> agonist mode, percentage activity is calculated using the following formula: % Activity = 100% x (1 - (mean RLU of test sample - mean RLU of MAX control) / (mean RLU of

vehicle control - mean RLU of MAX control)). Brimonidine was used as the control agonist. Each measurement was done in duplicate.

### **EMTA coupling panel for $\alpha_2AAR$**

The ebBRET-based effector membrane translocation assay (EMTA) allows detection of each  $G\alpha$  protein subunit activation. Upon receptor activation, G protein-effector proteins fused at their C-terminus to Renilla luciferase (RlucII) translocate from cytoplasm to the plasma membrane to selectively bind activated  $G\alpha$  proteins (p63-RhoGEF-RlucII with  $G_{q/11}$  family, Rap1GAP-RlucII with  $G_{i/o}$  family and PDZ-RhoGEF-RlucII with  $G_{12/13}$  family), thus leading to an increase in ebBRET by becoming in close proximity to the plasma membrane targeted energy acceptor, Renilla green fluorescent protein (rGFP-CAAX). The heterologous co-expression of each  $G\alpha$  subunits allow to identify which specific members of each G protein families (i.e.:  $G_{i1}$ ,  $G_{i2}$ ,  $G_{i3}$ ,  $G_{oA}$ ,  $G_{oB}$ ,  $G_z$ ,  $G_q$ ,  $G_{11}$ ,  $G_{14}$ ,  $G_{15}$ ,  $G_{12}$  and  $G_{13}$ ) is activated by a receptor. The assay is also sensitive enough to detect responses elicited by endogenous  $G_{i/o}$  protein families in the absence of heterologously expressed G protein. The same plasma membrane translocation principle is used to measure  $\beta$ -arrestin-1 or -2 recruitment (39) using  $\beta$ -arrestin-RlucII/rGFP-CAAX biosensors.

### *Cell Culture*

HEK293 clonal cell line (HEK293SL cells), hereafter referred as HEK293 cells, were a gift from S. Laporte (McGill University, Montreal, Quebec, Canada) and previously described (39). Cells were cultured in DMEM medium (Wisent; St-Jean-Baptiste, QC, Canada) supplemented with 10% newborn calf serum iron fortified (NCS; Wisent). Cells were passaged weekly and incubated at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub> and checked for mycoplasma contamination.

### *Transfection*

HEK293 cells (1.2 mL at  $3.5 \times 10^5$  cells per mL) were transfected with a fixed final amount of pre-mixed biosensor-encoding DNA (0.57  $\mu$ g, adjusted with salmon sperm DNA; Invitrogen) and human  $\alpha_{2A}$ AR DNA for  $G_s$ ,  $G_{i/o}$ ,  $G_{q/11}$  and  $\beta$ -arrestins experiments. For  $G_{12/13}$  experiments, cells were transfected with 1  $\mu$ g of total DNA (adjusted with salmon sperm DNA; Invitrogen), including empty pCDNA3.1 vector or human  $\alpha_{2A}$ AR DNA. Transfections were performed using linear polyethylenimine (PEI, 1 mg/mL; Polysciences) diluted in NaCl (150 mM, pH 7.0) as a transfecting agent (3:1 PEI/DNA ratio). Cells were immediately seeded ( $3.5 \times 10^4$  cells/well) into 96-well white microplates (Perkin Elmer), maintained in culture for the next 48 h and BRET experiments carried out. ebBRET (38) was used to monitor the activation of each  $G\alpha$  protein, as well as  $\beta$ -arrestin-1 and -2 recruitment to the plasma membrane.  $G\alpha_s$  protein engagement was measured between the plasma membrane marker rGFP-CAAX and human  $G\alpha_s$ -RlucII in presence of human  $G\beta_1$ ,  $G\gamma_9$  and  $\alpha_{2A}$ AR.  $G\alpha_{12}$  or  $G\alpha_{13}$  protein family activation was assessed using the selective- $G_{12/13}$  effector PDZ-RhoGEF-RlucII and rGFP-CAAX co-expressed with  $G\beta_1$ ,  $G\gamma_1$  and either  $G\alpha_{12}$  or  $G\alpha_{13}$ , in presence of  $\alpha_{2A}$ AR.  $G\alpha_{i/o}$  protein family activation was followed using the selective- $G_{i/o}$  effector Rap1GAP-RlucII and rGFP-CAAX along with the human  $G\alpha_{i1}$ ,  $G\alpha_{i2}$ ,  $G\alpha_{oA}$ ,  $G\alpha_{oB}$  or  $G\alpha_z$  subunits and  $\alpha_{2A}$ AR.  $G\alpha_{q/11}$  protein family activation was determined using the selective- $G_{q/11}$  effector p63-RhoGEF-RlucII and rGFP-CAAX along with the human  $G\alpha_q$ ,  $G\alpha_{11}$ ,  $G\alpha_{14}$  or  $G\alpha_{15}$  subunits and  $\alpha_{2A}$ AR.  $\beta$ -arrestin recruitment to the plasma membrane was determined using DNA mix containing rGFP-CAAX and  $\beta$ -arrestin-1-RlucII or  $\beta$ -arrestin-2-RlucII in presence of  $\alpha_{2A}$ AR.

### *Bioluminescence Resonance Energy Transfer Measurement*

The day of the experiment, cells were washed with phosphate-buffered saline (PBS) and incubated in Tyrode Hepes buffer (137 mM NaCl, 0.9 mM KCl, 1 mM MgCl<sub>2</sub>, 11.9 mM NaHCO<sub>3</sub>, 3.6 mM NaH<sub>2</sub>PO<sub>4</sub>, 25 mM HEPES, 5.5 mM D-Glucose and 1 mM CaCl<sub>2</sub>, pH 7.4) for 1 h at 37 °C. Cells were then treated with increasing concentrations of compounds for 10 min at 37 °C. The luciferase substrate Prolume purple (1 μM, NanoLight Technologies) was added during the last 6 min before the reading. Plates were read on the TriStar<sup>2</sup> LB 942 Multimode Microplate Reader (Berthold Technologies) with the energy donor filter (410 ± 80 nm; RlucII) and energy acceptor filter (515 ± 40 nm; GFP10 and rGFP CAAX). BRET signal (BRET<sup>2</sup>) was determined by calculating the ratio of the light intensity emitted by the acceptor (515 nm) over the light intensity emitted by the donor (410 nm) and data were normalized in percentage of the maximal response elicited by the reference compound Norepinephrine. The data were analyzed in GraphPad Prism 9.1 using “dose–response–stimulation log(agonist) vs response (four parameters)” and data were presented as mean ± s.e.m. of at least 3 different experiments each done in triplicate. E<sub>max</sub> and EC<sub>50</sub> values were determined from dose-response curves to calculate the log(E<sub>max</sub>/EC<sub>50</sub>) value for each pathway and each compound. To determine the relative efficacy of the compounds to activate the different signaling pathways, the difference between the log(E<sub>max</sub>/EC<sub>50</sub>) values was calculated using the following equation:

$$\Delta \text{Log} \left( \frac{E_{max}}{EC_{50}} \right) = \text{Log} \left( \frac{E_{max}}{EC_{50}} \right)_{\text{compound}} - \text{Log} \left( \frac{E_{max}}{EC_{50}} \right)_{\text{Norepinephrine}} \quad (1)$$

The compounds' efficacy toward each pathway, relative to norepinephrine, were calculated as the inverse logarithm of the Δlog(E<sub>max</sub>/EC<sub>50</sub>) using the following equation:

$$\text{Relative Efficacy (RE)} = 10^{\Delta \text{Log} \left( \frac{E_{max}}{EC_{50}} \right)} \quad (2)$$

The SEM was calculated for the  $\log(E_{\max}/EC_{50})$  ratios using the following equation:

$$SEM = \frac{\sigma}{\sqrt{n}} \quad (3)$$

where  $\sigma$  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:

$$SEM_{\left(\Delta\log\left(\frac{E_{\max}}{EC_{50}}\right)\right)} = \sqrt{(SEM_{\text{compound}})^2 + (SEM_{\text{Norepinephrine}})^2} \quad (4)$$

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 norepinephrine for a given pathway, where  $p < 0.05$  was considered to be statistically significant.

### **Internalization assay with rGFP-CAAX and rGFP-FYVE biosensors.**

#### *Plasmids*

Human  $\alpha_{2A}AR$  sequence was fused to RlucII by cloning between the NheI and BamHI sites of pCDNA3.1/Zeo(+)-RlucII vector, using polymerase chain reaction (Q5 Hot Start High-Fidelity DNA Polymerase from NEB), enzymatic digestion (NEB) and ligation (Anza™ T4 DNA Ligase Master Mix; Invitrogen).

#### *Transfection*

The protocol used for transfection is the same as for  $G_{12/13}$  EMTA experiments (i.e.: cells were transfected with 1  $\mu\text{g}$  of total DNA (adjusted with salmon sperm DNA; Invitrogen)). Transfections were performed using linear polyethylenimine (PEI, 1 mg/mL; Polysciences) diluted in NaCl (150 mM, pH 7.0) as a transfecting agent (3:1 PEI/DNA ratio). Cells were immediately seeded ( $3.5 \times 10^4$  cells/well) into 96-well white microplates (Perkin Elmer), maintained in culture for the next

48 h and BRET experiments carried out. Human  $\alpha_{2A}$ AR internalization was evaluated by measuring the disappearance of h $\alpha_{2A}$ AR-RlucII from the plasma membrane labeled with rGFP-CAAX and its relocalization in endosome labeled with rGFP-FYVE (39).

### *Bioluminescence Resonance Energy Transfer Measurement*

The day of the experiment, cells were washed with phosphate-buffered saline (PBS) and incubated in Tyrode Hepes buffer for 1 h at 37 °C. Cells were incubated during 6 min with the luciferase substrate ProLume purple (1  $\mu$ M, NanoLight Technologies) before addition of the indicated compound (0 or 100  $\mu$ M) and kinetics were recorded during 30 min. For concentration-response curves, BRET signal was measured after 30 min incubation. Plates were read on a Spark microplate reader (Tecan; Männedorf, Switzerland) using the BRET2 manufacturer settings. BRET signal (BRET<sup>2</sup>) was determined by calculating the ratio of the light intensity emitted by the acceptor (515 nm) over the light intensity emitted by the donor (410 nm) and for concentration-response curves, data were normalized in percentage of the maximal response elicited by the reference compound norepinephrine. The data were analyzed in GraphPad Prism 9.1 using “log(agonist) vs. response - Variable slope (four parameters)” and data were presented as mean  $\pm$  s.e.m. of 3 experiments performed in triplicate for kinetics or in simplicate for concentration-response curves.

### **Cryo-EM sample preparation and structure determination**

#### *Preparation of the ‘9087- $\alpha_{2A}$ AR-G<sub>oA</sub>-scFv16 and ‘4622- $\alpha_{2A}$ AR-G<sub>oA</sub>-scFv16 complexes*

The human wild type  $\alpha_{2A}$ AR was cloned to pFastBac vector with a N-terminal FLAG tag and a C-terminal histidine Tag. This construct was expressed in Sf9 insect cells using the pfastBac baculovirus system (Expression Systems). Cells were infected at a density of 4 x 10<sup>6</sup> cells per ml

and then incubated for 48 hours at 27 °C. Receptor was extracted and purified following the protocol described previously for  $\alpha_{2B}AR$  (9). Briefly, receptor was purified by Ni-NTA chromatography, Flag affinity chromatography and size exclusion chromatography in the presence of 100  $\mu$ M '9087 or '4622. The monomeric peak fractions of receptor were collected and concentrated to ~20 mg/mL. The freshly purified '9087-bound or '4622-bound  $\alpha_{2A}AR$  was used for complex formation with the G protein.  $G_{oA}$  heterotrimers were expressed and purified as previously described with minor modifications (78). Briefly, Hi5 cells were grown to a density of 3 million per mL and then infected with  $G\alpha_o$  and  $G\beta_1\gamma_2$  baculovirus at a ratio of 10-20 mL/L and 1-2 mL/L, respectively, and then incubated for 48 hours at 27 °C. Cells were solubilized with 1% (w/v) sodium cholate and 0.05% (w/v) DDM. After centrifugation, the supernatant was loaded onto Ni-NTA column and then exchanged to 0.05% DDM. The eluted  $G_{oA}$  heterotrimer was dephosphorylated by lambda phosphatase (homemade) and further purified through ion exchange using a Mono Q 10/100 GL column (GE Healthcare) and the peak fractions were collected and flash frozen in liquid nitrogen until use. The scFv16 (80) protein was expressed in insect Sf9 cells and purified with Ni-NTA column followed by the Superdex 200 Increase 10/300GL column (GE Healthcare) with a buffer composed of 20 mM HEPES, pH 7.5 and 100 mM NaCl. The monomeric peak fractions of receptor were collected and concentrated and stored at -80 °C until use. The complex formation process is same as described. Briefly, the complex of  $\alpha_{2A}AR$  with heterotrimeric  $G_{oA}$  was formed in a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% DDM, 1 mM  $MgCl_2$ , 10  $\mu$ M GDP and 100  $\mu$ M '9087 or '4622. The  $\alpha_{2A}AR$ - $G_{oA}$  complex was then treated with 50 units of apyrase (NEB) on ice overnight, and exchanged on an anti-Flag M1 column into a buffer containing 20 mM HEPES, pH 7.5, 100 mM NaCl, 0.0075% lauryl maltose neopentyl glycol (MNG, NG310 Anatrace), 0.0025% GDN (GDN101, Anatrace), and 0.001% CHS, 100  $\mu$ M



'9087 or '4622 and 2 mM CaCl<sub>2</sub> in a stepwise manner. After elution by adding 5 mM EDTA and 0.2 mg/mL Flag peptide, the complex was concentrated and incubated with 1.5x molar excess scFv16 for 1 hour on ice, then further purified using Superdex 200 Increase 10/300GL column (GE Healthcare) with a running buffer of 20 mM HEPES, pH 7.5, 100 mM NaCl, 0.00075% MNG, 0.00025% GDN and 0.0001% CHS, 100 μM '9087 or '4622. The monomeric peak fraction of α<sub>2A</sub>AR-G<sub>0A</sub> complex was collected and concentrated to ~5 mg/mL for cryo-EM.

#### *Cryo-EM data collection, processing, and model building*

3 μL of purified complex sample was applied onto the grid (CryoMatrix nickel titanium alloy film, R1.2/1.3, Zhenjiang Lehua Electronic Technology Co., Ltd.) (81) glow discharged at Tergeo-EM plasma cleaner and then blotted for 3 sec with blotting force of 0 and quickly plunged into liquid ethane cooled by liquid nitrogen using Vitrobot Mark IV (Thermo Fisher Scientific, USA) at 10 °C and with 95% humidity. Cryo-EM data was collected on a 300 kV Titan Krios Gi3 microscope. The raw movies were recorded by Gatan K3 BioQuantum Camera at the magnification of 105,000 and the corresponding pixel size is 0.85 Å. Inelastically scattered electrons were excluded by a GIF Quantum energy filter (Gatan, USA) using a slit width of 20 eV. The movie stacks were acquired with the defocus range of -1.0 to -1.6 micron with total exposure time 2.5 s fragmented into 50 frames (0.05 s/frame) with the dose rate of 22.0 e/pixel/s. The imaging mode is super resolution with 2-time hardware binning. The semi-automatic data acquisition was performed using SerialEM (82).

For the '9087-α<sub>2A</sub>AR-G<sub>0A</sub>-scFv16 complex, raw movie frames were aligned with MotionCor2 (83) using a 9 × 7 patch and the contrast transfer function (CTF) parameters were estimated using Gctf

and ctf in JSPR (84). Micrographs with consistent CTF values including defocus and astigmatism parameter were kept for the following image processing, which kept 3,768 micrographs from 4,217 raw movies. Templates for particle auto-picking were generated by projecting the 3D volume of norepinephrine-bound  $\alpha_{2A}$ AR-G<sub>0A</sub> complex (40). The 2,137,146 particles picked from template picking was subjected 2D classification in cryoSPARC (85) and 3D-classification in Relion3.1 (86). The sorted 321,762 particles were then subjected to homogeneous reconstruction in cryoSPARC, yielding a 3.57 Å map. Further 3D Ab-initio reconstruction reduced the particles number to 287,431, which was subjected to CTF refinement and non-uniform refinement after extracting with larger particle box size, and finally yield the 3.47 Å map.

For the '4622- $\alpha_{2A}$ AR-G<sub>0A</sub>-scFv16 complex, 6,983 raw movies were collected and subjected for motion correction using MotionCor2 (83). Contrast transfer function parameters were estimated by CTFFIND4, implemented in Relion3.1 (86). 2,593,747 particles were auto-picked using the templates in RELION3.1 and then subjected to 2D classification using cryoSPARC. Selected particles with appropriate 2D average from 2D classification were further subjected to Ab-initio reconstruction. Particles with appropriate initial model were selected from Ab-initio followed by heterogeneous refinement in cryoSPARC. The particles kept to 563,506 particles were subjected to non-uniform refinement and local refinement and yield a 3.38 Å reconstruction determined by gold standard Fourier shell correlation using the 0.143 criterion.

The norepinephrine- $\alpha_{2A}$ AR-G<sub>0A</sub> complex structure (PDB 7EJ0) (40) was used as the initial template for model building. The model was docked into the cryo-EM density map using UCSF Chimera (<https://rbvi.ucsf.edu/chimera>), followed by iterative manual building in Coot (87) and

real space refinement in Phenix. The statistics of the final models were validated by Molprobit. Cryo-EM data collection, refinement, and validation statistics are summarized in **Table 2.S10**. The ligand symmetry accounted RMSDs between the docked pose and cryo-EM pose of ‘**9087**’ and ‘**4622**’ were calculated by the Hungarian algorithm in DOCK6(88).

### ***pKa* determination for ‘9087**

The *pKa* of ‘**9087**’ (2.90 mg, 0.013 mmol) was determined by potentiometric titration using a Metrohm pH Meter 632 equipped with a glass electrode (Metrohm 6.0259.100). The compound was dissolved in 15 mL of 10% methanol aqueous solution, at an ionic strength of  $I = 0.15$  M using KCl. The resulting solution was stirred throughout the experiment using a magnetic stir bar and a magnetic agitator. The compound was titrated with 0.01 M HCl (Titrisol®) using an automatic burette (Metrohm Dosimat Plus 876). The titrant was added to the analyte stepwise (0.024-2.87 mL). The resulting graph for *pKa*-determination is presented in dependence of  $\tau$  and pH( $\tau$ ). The *pKa* value was then determined using a simplified Henderson-Hasselbalch equation. The data from the titration experiment was evaluated with Origin 9.60.

### **Off-target activity**

#### *GPCRome*

10  $\mu$ M ‘**9087**’ was tested for off-target activity at a panel of 320 non-olfactory GPCRs using PRESTO-Tango GPCRome arrestin-recruitment assay as described (45). Receptors with at least three-fold increased relative luminescence over corresponding basal activity are potential positive hits. Screening was performed by the National Institutes of Mental Health Psychoactive Drug

Screen Program (PDSP). Detailed experimental protocols are available on the NIMH PDSP website at <https://pdsp.unc.edu/pdspweb/content/PDSP%20Protocols%20II%202013-03-28.pdf>.

### *D2R Activation*

D2R was selected following the GPCRome panel and ‘9087 was re-tested for full dose-response to determine G protein and arrestin recruitment (see above).

### *I2R Binding*

Top docking compounds (‘9087, ‘2998, ‘4622, ‘0172) were tested for I2R binding, performed by Eurofins Cerep (France; catalog #81) as described (78). For compound ‘2998, no binding was seen in a single point radioligand competition experiment tested at 500nM and the compound is not shown.

### *μOR competition binding*

Equilibrium [3H] Diprenorphine competition and saturation binding were carried out in membranes prepared from Chinese Hamster Ovary (CHO-K1) cells stably expressing human μ-opioid receptor, as previously described (89–91). Briefly, binding was performed at 25 °C for 90 min in the dark. Binding in μOR/CHO-K1 cells was carried out in a buffer consisting of 50 mM HEPES-base pH 7.4 (pH adjusted with KOH), 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, and 0.1% (w/v) Bovine Serum Albumin with membranes containing approximately 40 μg/mL protein. Following incubation with radioligand (1 pM to 10 nM for saturation, 500 pM for competition), drugs (33 μM to 3.3 pM) and/or 20 μM cold competitor naloxone, the reaction was rapidly filtered onto GF/B (PerkinElmer #1450-521) glass fiber filtermats which were equilibrated for 1 hour in binding

buffer supplemented with 0.3% (v/v) polyethyleneimine. The filtermats were washed 5 times in ice-cold 50 mM HEPES-base pH 7.4 using a Perkin Elmer semi-automated cell harvester (Perkin Elmer FilterMate Harvester). The filtermats were dried and Meltilex solid scintillant (Perkin Elmer # 1450-442) was melted onto the mats for 10 min at 60 °C. The scintillant was allowed to re-solidify before disintegrations were quantified with a Wallac MicroBeta Scintillation counter using an integration time of 1 min. Non-specific binding, total binding, the number of receptor binding sites, and the  $K_d$  of the radiotracer were determined from saturation binding experiments. Protein concentrations were determined using the microBCA method with BSA as the standard.  $K_i$  values were calculated by non-linear regression analysis and application of the Cheng-Prusoff correction in GraphPad Prism 9.0.

#### *hERG inhibition assays*

‘9087 was tested for hERG inhibition in the FluxOR assay as described (92). hERG experiments used the National Institutes of Mental Health (NIMH) Psychoactive Drug Screening Program (PDSP). Experimental protocols are available on the NIMH PDSP website at <https://pdsp.unc.edu/pdspweb/content/PDSP%20Protocols%20II%202013-03-28.pdf>.

#### **Metabolic stability studies**

Metabolic stability of the test compounds was assessed using male pooled rat liver microsomes (Sprague Dawley, Sigma Aldrich) as previously described (93, 94). The reactions were carried out in 2 mL polyethylene tubes on a rotator carousel (Stuart™ SB3) in an incubator at 37 °C. The incubation mixture contained ‘9087, ‘7075, PS75, or the positive controls rotigotine or imipramine (final concentration 20 µM), and pooled rat liver microsomes (0.25 mg protein/tube) in Tris-MgCl<sub>2</sub>

buffer (50 mM Tris, 5 mM MgCl<sub>2</sub>, pH 7.4, final volume 500 μL). Transformation reactions were initiated by the addition of 50 μL of cofactor solution (NADPH, Carl Roth, final concentration 1 mM). After time intervals of 0, 30 and 60 min for **'9087, '7075, and PS75** or 0, 15, 30, 60 min for rotigotine and imipramine, respectively, 100 μL aliquots of the reaction mixtures were added to 100 μL ice-cold acetonitrile (containing 10 μM chlorpromazine as internal standard) to terminate enzymatic reactions. Precipitated protein was removed by centrifugation (1 min, 16,000 rcf) and the supernatants were analysed by HPLC/MS on a Thermo Scientific Dionex Ultimate 3000 HPLC system equipped with a Zorbax Eclipse XDB-C8 column (4.6 x 150 mm, 5 μm), a DAD detector (210 nm, 230 nm, 254 nm, 310 nm), and a BRUKER amaZon SL mass spectrometer with ESI source. The following binary eluent system (methanol in water + 0.1% (v/v) formic acid) was employed: 10% for 1 min, 10% to 100% in 20 min, 100% for 5 min, 100% to 10% in 2 min, 10% for 2 min, flow 0.4 mL/min. Per compound four (rotigotine, imipramine) or five (**'9087, '7075, PS75**) independent experiments were performed. Control experiments were conducted in the absence of cofactor solution to determine non-specific binding to matrix. The integral (AUC) of the extracted ion chromatograms (EIC) was used to analyze the concentration of the remaining substrates. Concentrations were plotted in their logarithmic form as a function of the incubation time (in min) to calculate the elimination rate constant (*k*) and to determine the half-life (*T*<sub>1/2</sub>) and intrinsic hepatic clearance (*CL*<sub>int</sub>) for each compound with the following equations (95):

$$t_{1/2} [min] = \frac{\ln(2)}{k} \quad (5)$$

$$CL_{int} \left[ \frac{\mu L}{min * mg(protein)} \right] = \frac{\ln(2)}{t_{1/2}} * \frac{V (of incubation in \mu L)}{m (of protein in incubation in mg)} \quad (6)$$

## ***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 #AN181214). 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. The  $\alpha_{2A}$ AR D79N mutant mice were purchased from Jackson (stock #2777), and 7-8 week-old females were used. Sample sizes were modelled on our previous studies and on studies using a similar approach, which were able to detect significant changes (96, 97). 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*

All ligands were synthesized by Enamine ('2998, '7075) or WuXi ('9087, '4622, '0172) or in-house (PS75) and dissolved 30 minutes prior testing. Available salt forms were used to aid solubility: HCl for '9087 and '7075, TFA for '4622, '0172, and PS75. '9087, '4622 and '0172 were resuspended in 20% Kolliphor (Sigma-Aldrich; cat. #C5135) for s.c./i.p. injections. '2998, '7075, and PS75 were resuspended in 20% cyclodextran (Sigma-Aldrich; cat. #H107) for s.c./i.p. injections. Atipamezole (Cayman Chemical Company; cat. #9001181) and Dexmedetomidine (Cayman Chemical Company; cat. #15581) were resuspended with NaCl 0.9% (Teknova; cat.

#S5819) for s.c./i.p. injections. **‘9087** was formulated with 40% Captisol (Carbosynth; cat. #OC15979) for p.o. dosing.

### *Behavioral analyses*

For all behavioral tests, the experimenter was always blind to treatment. Animals were first habituated for 1 hour in Plexiglas cylinders and then tested 30 minutes after subcutaneous injection of the  $\alpha_{2A}$ AR compounds. The  $\alpha_{2A}$ AR antagonist atipamezole (2 mg/kg, i.p.) was injected 15 minutes prior to s.c. injection of the  $\alpha_{2A}$ AR agonists. The mechanical (Von Frey), thermal (Hargreaves, hotplate and tail flick) and ambulatory (rotarod) tests were conducted as described previously (98). Hindpaw mechanical thresholds were determined with von Frey filaments using the updown method (99). Hindpaw thermal sensitivity was measured with a radiant heat source (Hargreaves) or a 55 °C hotplate. For the tail flick assay, sensitivity was measured by immersing the tail into a 50 °C water bath for both WT and D79N mutant mice. 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.



### *Complete Freund's Adjuvant (CFA)*

The CFA model of chronic inflammation was induced as described previously (100). Briefly, CFA (Sigma) was diluted 1:1 with saline and vortexed for 30 minutes. When fully suspended, we injected 20  $\mu$ L of CFA into one hindpaw. Heat thresholds were measured before the injection (baseline) and 3 days after using the Hargreaves test.

### *Constipation assay*

Mice had access to food and water ad libitum prior to the test. On the test day, mice received an i.p. injection of a solution (100  $\mu$ L) containing saline, 10 mg/kg morphine, 30  $\mu$ g/kg dexmedetomidine, or 5 mg/kg '9087 and then individually placed in a clean cage, with no access to food or water. Fecal pellets were collected and counted every hour, up to 6 hours.

### *Body weight measurement*

The body weights were measured before, 24 hours, and 48 hours after mice received an i.p. injection of a solution (100  $\mu$ L) containing dexmedetomidine (30  $\mu$ g/kg) or '9087 (5, 10, or 20 mg/kg).

### *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 brain distribution for '9087, '2998, '4622, '7075, PS75, and CSF distribution for '7075, PS75, '9087, and '4622, were measured following a 10 mg/kg (i.p.) dose. Plasma and brain samples were also collected for '9087

following 10 mg/kg i.v. and 30 mg/kg p.o. (oral) dose to determine oral bioavailability. In each compound study, nine time points (5, 15, 30, 60, 120, 240, 360, 480 and 1440 min) were collected, each of the time point treatment group included 3 animals. There was also a control group of one animal. In the ‘9087, ‘7075, ‘4622 studies, male C57BL/6N mice were used, for PS75 CD-1 mice, and for ‘2998 male Balb/cAnN mice. For all compound studies the animals were randomly assigned to the treatment groups before the pharmacokinetic study; all animals were fasted for 4 h before dosing. For injections, ‘9087 was dissolved in Captisol – water (40%:60%, w/v), ‘4622 was dissolved in a 20% Kolliphor HS – physiological saline solution, and ‘7075, PS75, and ‘2998 were dissolved in a 20% 2-HPBCD – aqueous solution. The batches of working formulations were prepared 10 minutes prior to the *in vivo* study.

Mice were injected i.p. with 2,2,2- tribromoethanol at 150 mg/kg prior to drawing CSF and blood. CSF was collected under a stereomicroscope from cisterna magna using 1 mL syringes. Blood collection was performed from the orbital sinus in microtainers containing K3EDTA. Animals were sacrificed by cervical dislocation after the blood samples collection. Blood samples were centrifuged 10 min at 3000 rpm. Brain samples (right lobe) were weighed and transferred into 1.5 mL tubes. 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 (IS) solution. After mixing by pipetting and centrifuging for 4 min at 6000 rpm, 4 µL of each supernatant was injected into the LC-MS/MS system. Solutions of internal standards were used to quantify compounds in the plasma samples. Brain samples (weight 200 mg ± 1 mg) were homogenized with 800 µL of an

internal stock solution using zirconium oxide beads (115 mg  $\pm$  5 mg) in a 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 (2  $\mu$ L) were mixed with 40  $\mu$ L of an internal stock solution. After mixing by pipetting and centrifuging for 4 min at 6,000 rpm, 5  $\mu$ L 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 plasma and brain samples were determined using high performance liquid chromatography/tandem mass spectrometry (HPLC-MS/MS). Data acquisition and system control was performed using Analyst 1.5.2 software (AB Sciex, Canada). The concentrations of the test compound below the lower limit of quantitation (LLOQ = 10 ng/mL for plasma, 20 ng/g for brain and 5 ng/mL for CSF samples) were designated as zero. 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  $T_{1/2}$  calculations.

Additional pharmacokinetic experiments were performed by Sai Life Sciences (Hyderabad, India) in accordance with the Sai Study Protocol SAIDMPK/PK-22-04-0340. Brain distribution of dexmedetomidine was measured following a 30  $\mu$ g/kg i.p. dose, using normal saline 0.9% as its vehicle. Plasma distributions were also collected for **PS75** following 10 mg/kg i.v. and 30 mg/kg p.o. (oral) dosing to determine oral bioavailability; both doses were formulated in 20% v/v H $\beta$ CD in saline. Testing was done in C57BL/6 mice. For **PS75**, 24 mice were divided into 4 groups: 9 mice for i.v. dosing of the compound, 3 mice for i.v. dosing of the vehicle, 9 mice for p.o. dosing

with the compound, and 3 mice for p.o. vehicle dosing; sparse sampling of three mice/time point for compound treated groups and 1 mouse/time point for vehicle groups was performed. For dexmedetomidine, 36 mice were included and split into two groups: 3 mice/time point for compound dosing, and 1 mouse/time point for vehicle only dosing. For **PS75**, blood samples (60  $\mu$ L) were collected under light isoflurane anesthesia (Surgivet®) from retro orbital plexus from a set of 3 mice at 0.083, 0.25, 0.5, 1, 2, 4, 6, 8, and 24 hr. Immediately after blood collection, plasma was harvested by centrifugation at 4000 rpm, 10 min at 4 °C. For dexmedetomidine, brain samples were collected at the same time points indicated above. Animals were sacrificed at respective time-points and brain samples were isolated and homogenized in ice-cold phosphate buffer saline (pH 7.4). Total homogenate volume was three times the tissue weight. Samples were stored at -70 °C until bioanalysis. All samples were processed for analysis by protein precipitation method and analyzed with fit-for-purpose LC-MS/MS method (LLOQ = 3.61 ng/mL for plasma for **PS75**, LLOQ = 0.86 ng/mL for brain for dexmedetomidine). The pharmacokinetic parameters were estimated using the non-compartmental analysis tool of Phoenix® WinNonlin software (v. 8.0).

### **Statistical analyses**

Data from functional experiments of adrenergic and D<sub>2long</sub> receptors were analyzed applying the algorithms for four parameter non-linear regression implemented in Prism 8.0 (GraphPad, San Diego, CA) to get dose-response curves representing EC<sub>50</sub> and E<sub>max</sub> values. Mean values were derived by summarizing the results from each individual experiment to provide EC<sub>50</sub>  $\pm$  s.e.m. and E<sub>max</sub>  $\pm$  s.e.m. (or s.d. where indicated). Additional statistical analyses for **Fig. 2.4**, **Fig. 2.S4**, **Fig. 2.S16A**, **Fig. 2.S16C-E**, **Fig. 2.S18** were performed with GraphPad Prism 9.0 (GraphPad Software Inc., San Diego). Data reported are means  $\pm$  s.e.m. or, in **Fig. 2.4** and **Fig. 2.S18**, single data points

with means  $\pm$  s.e.m. Experiments of the compounds in the *in vivo* neuropathic, inflammatory, hot-plate, tail flick, and rotarod models were evaluated using unpaired two-tailed Student's t-test or one-way ANOVA with Dunnett's multiple comparison post-hoc test to determine differences between groups. Experiments for body weight and constipation were analyzed with a two-way ANOVA. Details of the analyses, including groups compared in statistical sets, number of animals per group, and p-values can be found in the figure legends.

### **Data and material availability**

All data are available in the main text, supplementary materials, in the listed Protein Data Bank (PDB) files, Electron Microscopy Data Bank (EMDB) files, or at [https://github.com/efink14/ADRA2AR\\_docking\\_results](https://github.com/efink14/ADRA2AR_docking_results). The 3D cryo-EM density maps of **9087**- $\alpha_{2A}AR-G_{oA}$  and **4622**- $\alpha_{2A}AR-G_{oA}$  generated in this study have been deposited with accession code EMD-32331 and EMD-32342, respectively. The coordinates of **9087**- $\alpha_{2A}AR-G_{oA}$  and **4622**- $\alpha_{2A}AR-G_{oA}$  have been deposited with accession PDB code 7W6P and 7W7E, respectively. The identities of compounds docked in this study are freely available from ZINC15 and ZINC20 databases (<http://zinc15.docking.org>, <http://zinc20.docking.org>) and active compounds may be purchased from Enamine and WuXi Appetec. or are available from the authors. The docking results, including ZINC number, SMILES, and docking score, are located at [https://github.com/efink14/ADRA2AR\\_docking\\_results](https://github.com/efink14/ADRA2AR_docking_results). DOCK3.7 is freely available for non-commercial research (<http://dock.compbio.ucsf.edu/DOCK3.7/>). A web-based version is freely available to all (<http://blaster.docking.org/>). The biosensors used for generating the data in Tables 2.S3 and 2.S4 and Fig. 2.S6, 2.S7, and 2.S8 are protected by a patent but are available from M.B.

for noncommercial research without restrictions under a regular academic Material Transfer Agreement with the Université de Montréal.

## 2.8 Acknowledgements

We thank Brian Kobilka for sharing the structure of PDB 6K41 before publication. We thank T. Tummino and J. Lyu for reading this manuscript. We also thank WuXi Appetec. for providing information on compounds sourced from the WuXi GalaXi virtual library. We gratefully acknowledge OpenEye Software for Omega and related tools, and Schrodinger LLC for the Maestro package.

This work is supported by DARPA grant HR0011-19-2-0020 (B.K.S., A.I.B., J.J.I., M.P.J.), DFG Grant GRK 1910 (P.G.), and by US NIH grants R35GM122481 (B.K.S.), US NIH grant R01GM133836 (J.J.I.), US R35 NS097306 (A.I.B), Open Philanthropy (A.I.B) and Facial Pain Research Foundation (A.I.B), and by CIHR Foundation grant FN-148431 (M.B.). Y.D. is supported by a grant from Science, Technology and Innovation Commission of Shenzhen Municipality (Project JCYJ20200109150019113), and in part by the Kobilka Institute of Innovative Drug Discovery and Shenzhen Key Lab (ZDSYS20190902093417963). X.P.H is supported by NIMH PDSP (HHSN-271-2018-00023-C), directed by Dr. Bryan Roth. M.B. holds the Canada Research Chair in Signal Transduction and Molecular Pharmacology.

## 2.9 Author Contributions

E.A.F. conducted the docking screens with input from B.K.S. Ligand optimization performed by E.A.F. and P.S. with input from M.F.S, H.H., B.K.S and PG. H.H. performed all binding and functional assays and analyses for adrenergic and dopamine receptors with input from D.W. J.X. determined the '9087- $\alpha_{2A}$ AR-G<sub>0A</sub> and '4622- $\alpha_{2A}$ AR-G<sub>0A</sub> structures and made  $\alpha_{2A}$ AR mutations assisted by G.C. and Z.L., with supervision from Y.D. J.M.B. performed and analyzed the *in vivo* pharmacology experiments assisted by V.C., supervised and co-analyzed by A.I.B. C.A. tested select compounds in the panel of G protein and  $\beta$ -arrestin subtypes and receptor internalization with supervision from M.B. S.G. modeled compound '7075 and PS75. X.H. tested compounds in GPCRome and hERG assays. M.F.S. performed contact area calculation, P.S. synthesized bespoke compounds and performed *pKa* and analytical testing with supervision of P.G. D.W. performed ELISA experiments. C.M.W. tested compounds for  $\mu$ OR activity. C.K. performed permeability calculations with supervision of M.P.J. Y.S.M. supervised compound synthesis of Enamine compounds purchased from ZINC15 database and 12 billion catalog, assisted by N.A.T. J.J.I. built the ZINC15 ultra-large libraries. B.K.S., P.G., A.I.B., and Y.D. supervised the project. E.A.F. wrote the paper with contributions from J.M.B. and J.X., input from all other authors, and primary editing from B.K.S. and P.G. B.K.S., P.G. and A.I.B. conceived the project.



## 2.10 References

1. A. S. Yekkirala, D. P. Roberson, B. P. Bean, C. J. Woolf, Breaking barriers to novel analgesic drug development. *Nat. Rev. Drug Discov.* **16**, 545–564 (2017).
2. L. Manchikanti, S. Helm, B. Fellows, J. W. Janata, V. Pampati, J. S. Grider, M. V. Boswell, Opioid epidemic in the United States. *Pain Physician.* **15**, ES9-38 (2012).
3. J. Strang, N. D. Volkow, L. Degenhardt, M. Hickman, K. Johnson, G. F. Koob, B. D. L. Marshall, M. Tyndall, S. L. Walsh, Opioid use disorder. *Nat. Rev. Dis. Primer.* **6**, 3 (2020).
4. K. Rasmussen, D. A. White, J. B. Acri, NIDA’s medication development priorities in response to the Opioid Crisis: ten most wanted. *Neuropsychopharmacology* **44**, 657–659 (2019).
5. A. B. Malmberg, L. R. Hedley, J. R. Jasper, J. C. Hunter, A. I. Basbaum, Contribution of  $\alpha_2$  receptor subtypes to nerve injury-induced pain and its regulation by dexmedetomidine. *Br. J. Pharmacol.* **132**, 1827–1836 (2001).
6. P. P. Lakhani, L. B. MacMillan, T. Z. Guo, B. A. McCool, D. M. Lovinger, M. Maze, L. E. Limbird, Substitution of a mutant  $\alpha_{2a}$ -adrenergic receptor via “hit and run” gene targeting reveals the role of this subtype in sedative, analgesic, and anesthetic-sparing responses in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9950–9955 (1997).
7. H. Buerkle, T. L. Yaksh, Pharmacological evidence for different alpha 2-adrenergic receptor sites mediating analgesia and sedation in the rat. *Br. J. Anaesth.* **81**, 208–215 (1998).
8. M. Anttila, J. Penttilä, A. Helminen, L. Vuorilehto, H. Scheinin, Bioavailability of dexmedetomidine after extravascular doses in healthy subjects. *Br. J. Clin. Pharmacol.* **56**, 691–693 (2003).

9. D. Yuan, Z. Liu, J. Kaindl, S. Maeda, J. Zhao, X. Sun, J. Xu, P. Gmeiner, H.-W. Wang, B. K. Kobilka, Activation of the  $\alpha_{2B}$  adrenoceptor by the sedative sympatholytic dexmedetomidine. *Nat. Chem. Biol.* **16**, 507–512 (2020).
10. J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O’Meara, T. Che, E. Alga, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
11. T. Sterling, J. J. Irwin, ZINC 15 – Ligand discovery for everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
12. J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, R. A. Sayle, ZINC20—A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).
13. R. M. Stein, H. J. Kang, J. D. McCorvy, G. C. Glatfelter, A. J. Jones, T. Che, S. Slocum, X.-P. Huang, O. Savych, Y. S. Moroz, B. Stauch, L. C. Johansson, V. Cherezov, T. Kenakin, J. J. Irwin, B. K. Shoichet, B. L. Roth, M. L. Dubocovich, Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609–614 (2020).
14. A. Alon, J. Lyu, J. M. Braz, T. A. Tummino, V. Craik, M. J. O’Meara, C. M. Webb, D. S. Radchenko, Y. S. Moroz, X.-P. Huang, Y. Liu, B. L. Roth, J. J. Irwin, A. I. Basbaum, B. K. Shoichet, A. C. Kruse, Structures of the  $\sigma_2$  receptor enable docking for bioactive ligand discovery. *Nature* **600**, 759–764 (2021).
15. C. Gorgulla, A. Boeszoermyeni, Z.-F. Wang, P. D. Fischer, P. W. Coote, K. M. Padmanabha Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner, H. Arthanari, An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).

16. A. A. Sadybekov, A. V. Sadybekov, Y. Liu, C. Iliopoulos-Tsoutsouvas, X.-P. Huang, J. Pickett, B. Houser, N. Patel, N. K. Tran, F. Tong, N. Zvonok, M. K. Jain, O. Savych, D. S. Radchenko, S. P. Nikas, N. A. Petasis, Y. S. Moroz, B. L. Roth, A. Makriyannis, V. Katritch, Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022).
17. A. Manglik, H. Lin, D. K. Aryal, J. D. McCorvy, D. Dengler, G. Corder, A. Levit, R. C. Kling, V. Bernat, H. Hübner, X.-P. Huang, M. F. Sassano, P. M. Giguère, S. Löber, Da Duan, G. Scherrer, B. K. Kobilka, P. Gmeiner, B. L. Roth, B. K. Shoichet, Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185–190 (2016).
18. M. Schuller, G. J. Correy, S. Gahbauer, D. Fearon, T. Wu, R. E. Díaz, I. D. Young, L. Carvalho Martins, D. H. Smith, U. Schulze-Gahmen, T. W. Owens, I. Deshpande, G. E. Merz, A. C. Thwin, J. T. Biel, J. K. Peters, M. Moritz, N. Herrera, H. T. Kratochvil, QCRG Structural Biology Consortium, A. Aimon, J. M. Bennett, J. Brandao Neto, A. E. Cohen, A. Dias, A. Douangamath, L. Dunnett, O. Fedorov, M. P. Ferla, M. R. Fuchs, T. J. Gorrie-Stone, J. M. Holton, M. G. Johnson, T. Krojer, G. Meigs, A. J. Powell, J. G. M. Rack, V. L. Rangel, S. Russi, R. E. Skyner, C. A. Smith, A. S. Soares, J. L. Wierman, K. Zhu, P. O'Brien, N. Jura, A. Ashworth, J. J. Irwin, M. C. Thompson, J. E. Gestwicki, F. von Delft, B. K. Shoichet, J. S. Fraser, I. Ahel, Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).
19. A. Rudling, R. Gustafsson, I. Almlöf, E. Homan, M. Scobie, U. Warpman Berglund, T. Helleday, P. Stenmark, J. Carlsson, Fragment-based discovery and optimization of enzyme inhibitors by docking of commercial chemical space. *J. Med. Chem.* **60**, 8160–8169 (2017).

20. A. Levit Kaplan, R. T. Strachan, J. M. Braz, V. Craik, S. Slocum, T. Mangano, V. Amabo, H. O'Donnell, P. Lak, A. I. Basbaum, B. L. Roth, B. K. Shoichet, Structure-based design of a chemical probe set for the 5-HT<sub>5A</sub> Serotonin Receptor. *J. Med. Chem.* **65**, 4201–4217 (2022).
21. C. de Graaf, A. J. Kooistra, H. F. Vischer, V. Katritch, M. Kuijter, M. Shiroishi, S. Iwata, T. Shimamura, R. C. Stevens, I. J. P. de Esch, R. Leurs, Crystal structure-based virtual screening for fragment-like ligands of the human histamine H<sub>1</sub> Receptor. *J. Med. Chem.* **54**, 8195–8206 (2011).
22. R. Kiss, B. Kiss, Á. Könczöl, F. Szalai, I. Jelinek, V. László, B. Noszál, A. Falus, G. M. Keserü, Discovery of novel human histamine H<sub>4</sub> receptor ligands by large-scale structure-based virtual screening. *J. Med. Chem.* **51**, 3145–3153 (2008).
23. N. Patel, X. P. Huang, J. M. Grandner, L. C. Johansson, B. Stauch, J. D. McCorvy, Y. Liu, B. Roth, V. Katritch, Structure-based discovery of potent and selective melatonin receptor agonists. *eLife* **9**, e53779 (2020).
24. C. J. Langmead, S. P. Andrews, M. Congreve, J. C. Errey, E. Hurrell, F. H. Marshall, J. S. Mason, C. M. Richardson, N. Robertson, A. Zhukov, M. Weir, Identification of novel adenosine A<sub>2A</sub> receptor antagonists by virtual screening. *J. Med. Chem.* **55**, 1904–1909 (2012).
25. S. Wang, D. Wacker, A. Levit, T. Che, R. M. Betz, J. D. McCorvy, A. J. Venkatakrishnan, X.-P. Huang, R. O. Dror, B. K. Shoichet, B. L. Roth, D<sub>4</sub> dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* **358**, 381–386 (2017).
26. R. G. Coleman, M. Carchia, T. Sterling, J. J. Irwin, B. K. Shoichet, Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992 (2013).

27. K. Altosaar, P. Balaji, R. A. Bond, D. B. Bylund, S. Cotecchia, D. Devost, V. A. Doze, D. C. Eikenburg, S. Gora, E. Goupil, R. M. Graham, T. Hébert, J. P. Hieble, R. Hills, S. Kan, G. Machkalyan, M. C. Michel, K. P. Minneman, S. Parra, D. Perez, R. Sleno, R. Summers, P. Zylbergold, Adrenoceptors (version 2019.4) in the IUPHAR/BPS Guide to Pharmacology Database. *IUPHARBPS Guide Pharmacol. CITE*. **2019** (2019), doi:10.2218/gtopdb/F4/2019.4.
28. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
29. H. S. Blaxall, T. J. Murphy, J. C. Baker, C. Ray, D. B. Bylund, Characterization of the alpha-2C adrenergic receptor subtype in the opossum kidney and in the OK cell line. *J. Pharmacol. Exp. Ther.* **259**, 323–329 (1991).
30. P. B. M. W. M. Timmermans, P. A. Van Zwieten,  $\alpha_2$  adrenoceptors: classification, localization, mechanisms and targets for drugs. *J. Med. Chem.* **25**, 1389–1401 (1982).
31. B. V. Clineschmidt, L. M. Flataker, E. P. Faison, D. R. Haubrich, alpha-methyldopa reduces locomotor activity in rats via its metabolite, alpha-methylnorepinephrine, acting on alpha 2-adrenoceptors. *Arch. Int. Pharmacodyn. Ther.* **244**, 231–243 (1980).
32. S. Gu, M. S. Smith, Y. Yang, J. J. Irwin, B. K. Shoichet, Ligand strain energy in large library docking. *J. Chem. Inf. Model.* **61**, 4331–4341 (2021).
33. V. Isberg, C. de Graaf, A. Bortolato, V. Cherezov, V. Katritch, F. H. Marshall, S. Mordalski, J.-P. Pin, R. C. Stevens, G. Vriend, D. E. Gloriam, Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).

34. M. Congreve, S. P. Andrews, A. S. Doré, K. Hollenstein, E. Hurrell, C. J. Langmead, J. S. Mason, I. W. Ng, B. Tehan, A. Zhukov, M. Weir, F. H. Marshall, Discovery of 1,2,4-triazine derivatives as adenosine A<sub>2A</sub> antagonists using structure based drug design. *J. Med. Chem.* **55**, 1898–1903 (2012).
35. S. Bhattacharya, S. E. Hall, H. Li, N. Vaidehi, Ligand-stabilized conformational states of human  $\beta_2$  adrenergic receptor: insight into G-protein-coupled receptor activation. *Biophys. J.* **94**, 2027–2042 (2008).
36. D. R. Weiss, S. Ahn, M. F. Sassano, A. Kleist, X. Zhu, R. Strachan, B. L. Roth, R. J. Lefkowitz, B. K. Shoichet, Conformation guides molecular efficacy in docking screens of activated  $\beta$ -2 adrenergic G protein coupled receptor. *ACS Chem. Biol.* **8**, 1018–1026 (2013).
37. M. G. Eason, H. Kurose, B. D. Holt, J. R. Raymond, S. B. Liggett, Simultaneous coupling of  $\alpha_2$ -adrenergic receptors to two G-proteins with opposing effects. Subtype-selective coupling of  $\alpha_2C10$ ,  $\alpha_2C4$ , and  $\alpha_2C2$  adrenergic receptors to G<sub>i</sub> and G<sub>s</sub>. *J. Biol. Chem.* **267**, 15795–15801 (1992).
38. C. Avet, A. Mancini, B. Breton, C. Le Gouill, A. S. Hauser, C. Normand, H. Kobayashi, F. Gross, M. Hogue, V. Lukashova, S. St-Onge, M. Carrier, M. Héroux, S. Morissette, E. B. Fauman, J.-P. Fortin, S. Schann, X. Leroy, D. E. Gloriam, M. Bouvier, Effector membrane translocation biosensors reveal G protein and  $\beta$ arrestin coupling profiles of 100 therapeutically relevant GPCRs. *eLife* **11**, e74101 (2022).
39. Y. Namkung, C. Le Gouill, V. Lukashova, H. Kobayashi, M. Hogue, E. Khoury, M. Song, M. Bouvier, S. A. Laporte, Monitoring G protein-coupled receptor and  $\beta$ -arrestin trafficking in live cells using enhanced bystander BRET. *Nat. Commun.* **7**, 12178 (2016).

40. J. Xu, S. Cao, H. Hübner, D. Weikert, G. Chen, Q. Lu, D. Yuan, P. Gmeiner, Z. Liu, Y. Du, Structural insights into ligand recognition, activation, and signaling of the  $\alpha_{2A}$  adrenergic receptor. *Sci. Adv.* **8**, eabj5347 (2022).
41. L. Qu, Q. Zhou, Y. Xu, Y. Guo, X. Chen, D. Yao, G. W. Han, Z.-J. Liu, R. C. Stevens, G. Zhong, D. Wu, S. Zhao, Structural basis of the diversity of adrenergic receptors. *Cell Rep.* **29**, 2929-2935.e4 (2019).
42. F. M. Heydenreich, M. Marti-Solano, M. Sandhu, B. K. Kobilka, M. Bouvier, M. M. Babu, <https://www.biorxiv.org/content/10.1101/2021.09.14.460253v1> (2021).
43. M. Masureel, Y. Zou, L.-P. Picard, E. van der Westhuizen, J. P. Mahoney, J. P. G. L. M. Rodrigues, T. J. Mildorf, R. O. Dror, D. E. Shaw, M. Bouvier, E. Pardon, J. Steyaert, R. K. Sunahara, W. I. Weis, C. Zhang, B. K. Kobilka, Structural insights into binding specificity, efficacy and bias of a  $\beta_2$ AR partial agonist. *Nat. Chem. Biol.* **14**, 1059–1066 (2018).
44. M. Stanek, L.-P. Picard, M. F. Schmidt, J. M. Kaindl, H. Hübner, M. Bouvier, D. Weikert, P. Gmeiner, Hybridization of  $\beta$ -Adrenergic agonists and antagonists confers G protein bias. *J. Med. Chem.* **62**, 5111–5131 (2019).
45. W. K. Kroeze, M. F. Sassano, X.-P. Huang, K. Lansu, J. D. McCorvy, P. M. Giguère, N. Sciaky, B. L. Roth, PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).
46. H. Kodaira, H. Kusuhara, T. Fujita, J. Ushiki, E. Fuse, Y. Sugiyama, Quantitative evaluation of the impact of active efflux by P-glycoprotein and breast cancer resistance protein at the blood-brain barrier on the predictability of the unbound concentrations of drugs in the brain using cerebrospinal fluid concentration as a surrogate. *J. Pharmacol. Exp. Ther.* **339**, 935–944 (2011).

47. S. D. Shields, W. A. Eckert, A. I. Basbaum, Spared nerve injury model of neuropathic pain in the mouse: a behavioral and anatomic analysis. *J. Pain* **4**, 465–470 (2003).
48. R. Virtanen, J. M. Savola, V. Saano, Highly selective and specific antagonism of central and peripheral alpha 2-adrenoceptors by atipamezole. *Arch. Int. Pharmacodyn. Ther.* **297**, 190–204 (1989).
49. L. B. MacMillan, L. Hein, M. S. Smith, M. T. Piascik, L. E. Limbird, Central hypotensive effects of the  $\alpha_{2a}$ -adrenergic receptor subtype. *Science* **273**, 801–803 (1996).
50. P. P. Lakhiani, D. M. Lovinger, L. E. Limbird, Genetic evidence for involvement of multiple effector systems in alpha 2A-adrenergic receptor inhibition of stimulus-secretion coupling. *Mol. Pharmacol.* **50**, 96–103 (1996).
51. B. P. Ceresa, L. E. Limbird, Mutation of an aspartate residue highly conserved among G-protein-coupled receptors results in nonreciprocal disruption of  $\alpha_2$ -adrenergic receptor-G-protein interactions. A negative charge at amino acid residue 79 forecasts  $\alpha_{2A}$ -adrenergic receptor sensitivity to allosteric modulation by monovalent cations and fully effective receptor/G-protein coupling. *J. Biol. Chem.* **269**, 29557–29564 (1994).
52. J. C. Hunter, D. J. Fontana, L. R. Hedley, J. R. Jasper, R. Lewis, R. E. Link, R. Secchi, J. Sutton, R. M. Eglon, Assessment of the role of  $\alpha_2$ -adrenoceptor subtypes in the antinociceptive, sedative and hypothermic action of dexmedetomidine in transgenic mice. *Br. J. Pharmacol.* **122**, 1339–1344 (1997).
53. S. A. Metz, J. B. Halter, R. Paul Robertson, Induction of defective insulin secretion and impaired glucose tolerance by clonidine selective stimulation of metabolic alpha-adrenergic pathways. *Diabetes* **27**, 554–562 (1978).



54. S. S. Lichtenstein, C. Marinescu, S. F. Leibowitz, Chronic infusion of norepinephrine and clonidine into the hypothalamic paraventricular nucleus. *Brain Res. Bull.* **13**, 591–595 (1984).
55. L. S. Stone, J. P. German, K. F. Kitto, C. A. Fairbanks, G. L. Wilcox, Morphine and clonidine combination therapy improves therapeutic window in mice: synergy in antinociceptive but not in sedative or cardiovascular effects. *PLoS One* **9**, e109903 (2014).
56. D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).
57. E. Von Moo, K. Harpsøe, A. S. Hauser, I. Masuho, H. Bräuner-Osborne, D. E. Gloriam, K. A. Martemyanov, Ligand-directed bias of G protein signaling at the dopamine D<sub>2</sub> receptor. *Cell Chem. Biol.*, S2451945621003147 (2021).
58. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).
59. J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).
60. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).
61. B. J. Bender, S. Gahbauer, A. Lutten, J. Lyu, C. M. Webb, R. M. Stein, E. A. Fink, T. E. Balius, J. Carlsson, J. Irwin, B. K. Shoichet, A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799–4832 (2021).

62. R. M. Stein, Y. Yang, T. E. Balius, M. J. O'Meara, J. Lyu, J. Young, K. Tang, B. K. Shoichet, J. J. Irwin, Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model.* **61**, 699–714 (2021).
63. M. M. Mysinger, D. R. Weiss, J. J. Ziarek, S. Gravel, A. K. Doak, J. Karpiak, N. Heveker, B. K. Shoichet, B. F. Volkman, Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5517–5522 (2012).
64. B. K. Shoichet, A. R. Leach, I. D. Kuntz, Ligand solvation in molecular docking. *Proteins* **34**, 4–16 (1999).
65. K. Gallagher, K. Sharp, Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **75**, 769–776 (1998).
66. M. M. Mysinger, B. K. Shoichet, Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).
67. T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey, M. P. Jacobson, Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **128**, 14073–14080 (2006).
68. S. S. F. Leung, J. Mijalkovic, K. Borrelli, M. P. Jacobson, Testing physical models of passive membrane permeation. *J. Chem. Inf. Model.* **52**, 1621–1636 (2012).
69. M. P. Jacobson, G. A. Kaminski, R. A. Friesner, C. S. Rapp, Force field validation using protein side chain prediction. *J. Phys. Chem. B* **106**, 11673–11680 (2002).

70. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
71. O. H. Lowry, N. J. Rosebrough, A. L. Farr, R. J. Randall, Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275 (1951).
72. Y. Cheng, W. H. Prusoff, Relationship between the inhibition constant ( $K_I$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
73. X. M. Guan, T. S. Kobilka, B. K. Kobilka, Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide. *J. Biol. Chem.* **267**, 21995–21998 (1992).
74. J. Quoyer, J. M. Janz, J. Luo, Y. Ren, S. Armando, V. Lukashova, J. L. Benovic, K. E. Carlson, S. W. Hunt, M. Bouvier, Pepducin targeting the C-X-C chemokine receptor type 4 acts as a biased agonist favoring activation of the inhibitory G protein. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E5088-5097 (2013).
75. C. Galés, J. J. J. Van Durm, S. Schaak, S. Pontier, Y. Percherancier, M. Audet, H. Paris, M. Bouvier, Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nat. Struct. Mol. Biol.* **13**, 778–786 (2006).
76. D. Möller, A. Banerjee, T. C. Uzuneser, M. Skultety, T. Huth, B. Plouffe, H. Hübner, C. Alzheimer, K. Friedland, C. P. Müller, M. Bouvier, P. Gmeiner, Discovery of G protein-biased dopaminergics with a pyrazolo[1,5-a]pyridine substructure. *J. Med. Chem.* **60**, 2908–2929 (2017).

77. H. Liu, J. Hofmann, I. Fish, B. Schaake, K. Eitel, A. Bartuschat, J. Kaindl, H. Rampp, A. Banerjee, H. Hübner, M. J. Clark, S. G. Vincent, J. T. Fisher, M. R. Heinrich, K. Hirata, X. Liu, R. K. Sunahara, B. K. Shoichet, B. K. Kobilka, P. Gmeiner, Structure-guided development of selective M3 muscarinic acetylcholine receptor antagonists. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12046–12050 (2018).
78. J. Xu, Y. Hu, J. Kaindl, P. Risel, H. Hübner, S. Maeda, X. Niu, H. Li, P. Gmeiner, C. Jin, B. K. Kobilka, Conformational complexity and dynamics in a muscarinic receptor revealed by NMR spectroscopy. *Mol. Cell* **75**, 53-65.e7 (2019).
79. X. Liu, J. Kaindl, M. Korczynska, A. Stöbel, D. Dengler, M. Stanek, H. Hübner, M. J. Clark, J. Mahoney, R. A. Matt, X. Xu, K. Hirata, B. K. Shoichet, R. K. Sunahara, B. K. Kobilka, P. Gmeiner, An allosteric modulator binds to a conformational hub in the  $\beta_2$  adrenergic receptor. *Nat. Chem. Biol.* **16**, 749–755 (2020).
80. S. Maeda, A. Koehl, H. Matile, H. Hu, D. Hilger, G. F. X. Schertler, A. Manglik, G. Skiniotis, R. J. P. Dawson, B. K. Kobilka, Development of an antibody fragment that stabilizes GPCR/G-protein complexes. *Nat. Commun.* **9**, 3712 (2018).
81. X. Huang, L. Zhang, Z. Wen, H. Chen, S. Li, G. Ji, C. Yin, F. Sun, Amorphous nickel titanium alloy film: a new choice for cryo electron microscopy sample preparation. *Prog. Biophys. Mol. Biol.* **156**, 3–13 (2020).
82. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
83. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

84. W. Jiang, F. Guo, Z. Liu, A graph theory method for determination of cryo-EM image focuses. *J. Struct. Biol.* **180**, 343–351 (2012).
85. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
86. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
87. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
88. W. J. Allen, R. C. Rizzo, Implementation of the hungarian algorithm to account for ligand symmetry and similarity in structure-based design. *J. Chem. Inf. Model.* **54**, 518–529 (2014).
89. C. A. Flanagan, "GPCR-radioligand binding assays" in *Methods in Cell Biology* (Elsevier, 2016), vol. 132, pp. 191–215.
90. R. N. DeHaven, D. L. DeHaven-Hudkins, Characterization of opioid receptors. *Curr. Protoc. Pharmacol.* **8** (2000), doi:10.1002/0471141755.ph0104s08.
91. E. C. Hulme, M. A. Trevethick, Ligand binding assays at equilibrium: validation and interpretation: equilibrium binding assays. *Br. J. Pharmacol.* **161**, 1219–1237 (2010).
92. X.-P. Huang, T. Mangano, S. Hufeisen, V. Setola, B. L. Roth, Identification of human *Ether-à-go-go* related gene modulators by three screening platforms in an academic drug-discovery setting. *ASSAY Drug Dev. Technol.* **8**, 727–742 (2010).
93. C. Hiller, R. C. Kling, F. W. Heinemann, K. Meyer, H. Hübner, P. Gmeiner, Functionally selective dopamine D<sub>2</sub>/D<sub>3</sub> receptor agonists comprising an enyne moiety. *J. Med. Chem.* **56**, 5130–5141 (2013).

94. J. Hellmann, M. Drabek, J. Yin, J. Gunera, T. Pröll, F. Kraus, C. J. Langmead, H. Hübner, D. Weikert, P. Kolb, D. M. Rosenbaum, P. Gmeiner, Structure-based development of a subtype-selective orexin 1 receptor antagonist. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 18059–18067 (2020).
95. L. Di, C. Keefer, D. O. Scott, T. J. Strelevitz, G. Chang, Y.-A. Bi, Y. Lai, J. Duckworth, K. Fenner, M. D. Troutman, R. S. Obach, Mechanistic insights from comparing intrinsic clearance values between human liver microsomes and hepatocytes to guide drug design. *Eur. J. Med. Chem.* **57**, 441–448 (2012).
96. G. Scherrer, N. Imamachi, Y.-Q. Cao, C. Contet, F. Mennicken, D. O'Donnell, B. L. Kieffer, A. I. Basbaum, Dissociation of the opioid receptor mechanisms that control mechanical and heat pain. *Cell* **137**, 1148–1159 (2009).
97. A. Muralidharan, A. Samoshkin, M. Convertino, M. H. Piltonen, P. Gris, J. Wang, C. Jiang, R. Klares, A. Linton, R. Ji, W. Maixner, N. V. Dokholyan, J. S. Mogil, L. Diatchenko, Identification and characterization of novel candidate compounds targeting 6- and 7-transmembrane  $\mu$ -opioid receptor isoforms. *Br. J. Pharmacol.* **178**, 2709–2726 (2021).
98. C. G. Alvaro, J. M. Braz, M. Bernstein, K. A. Hamel, V. Craik, H. Yamanaka, A. I. Basbaum, Hippocalcin-like 4, a neural calcium sensor, has a limited contribution to pain and itch processing. *PLoS One* **15**, e0226289 (2020).
99. S. R. Chaplan, F. W. Bach, J. W. Pogrel, J. M. Chung, T. L. Yaksh, Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Methods* **53**, 55–63 (1994).
100. Y. Q. Cao, P. W. Mantyh, E. J. Carlson, A.-M. Gillespie, C. J. Epstein, A. I. Basbaum, Primary afferent tachykinins are required to experience moderate to intense pain. *Nature* **392**, 390–394 (1998).

### **Gloss to Chapter 3**

In the midst of working on Chapter 2 research towards applications in the ongoing opioid epidemic, a world-wide pandemic began. It was scary, we all went home for a long, long time, and there was a lot unknown about life. One thing that was clear very early on, is that we as a field were equipped to start trying to work towards SARS-CoV-2 therapeutics. The Shoichet lab joined many labs around the world in SARS-CoV-2 research with many viral proteins or even human proteins in mind. Here, we took on the challenge of the understudied viral non-structural protein 14 (nsp14).

In contrast to Chapter 2, here we knew very little about SARS-CoV-2 and nsp14. There were hints of knowledge from earlier coronavirus strains, but nothing too concrete. Isha, another co-first author in the lab, and I dived into many different approaches to virtual screening for nsp14 in a ‘all but the kitchen sink’ approach. Here, I was able to learn how different subsets of the virtual chemical libraries could be used, and even how different types of protein-ligand interactions could be exploited. This would prove to be a useful skill as it was applied also in Chapter 4 research. Overall, we tried ‘normal’ noncovalent docking, fragment docking with very small molecules sampling parts of the overall protein site of interest, and also covalent docking with molecules able to react and form formal bonds with the protein.

We learned all types of approaches were useful and successfully discovered nsp14 inhibitors from each. However, in contrast to Chapter 2, here we yielded overall many fewer successful compounds because we knew so little starting the project. This reemphasized the importance of trying multiple strategies if the likelihood of success was already slimmer than an ‘average’ project.

From there, work began, and will continue, to improve compound inhibition profiles and cellular antiviral activity with many great labs, including the Vedadi, Fujimori, Jin Gastía-Sastre and Ott labs. Just as in Chapter 2, the collaborative effort in a dire public health situation was immensely rewarding to not just learn from the other scientists, but to contribute to the collective good in a very isolating time.



## Chapter 3

### Structure-based discovery of inhibitors of the SARS-CoV-2 Nsp14 N7-methyltransferase

#### Contributing Authors

Isha Singh<sup>†,1</sup>, Fengling Li<sup>†,2</sup>, Elissa Fink<sup>†,1,3</sup>, Irene Chau<sup>2</sup>, Alice Li<sup>2,4</sup>, Annía Rodríguez-Hernández<sup>5</sup>, Francisco J. Zapatero-Belinchón<sup>6</sup>, Mario Rodríguez<sup>7,8</sup>, Kanchan Devkota<sup>2</sup>, Zhijie Deng<sup>9</sup>, Kris White<sup>7,8</sup>, Xiaobo Wan<sup>1</sup>, Nataliya A. Tolmachova<sup>10,11</sup>, Yurii S. Moroz<sup>12,13</sup>, John Gross<sup>1</sup>, H. Ümit Kaniskan<sup>9</sup>, Melanie Ott<sup>6,14,15,16</sup>, Adolfo Gastía-Sastre<sup>7,8,14,17,18</sup>, Jian Jin<sup>9,14</sup>, Danica Galonić Fujimori<sup>1,5,14</sup>, John J. Irwin<sup>1,14\*</sup>, Masoud Vedadi<sup>2,4,14\*</sup> and Brian K. Shoichet<sup>1,14\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA.

<sup>2</sup>Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada.

<sup>3</sup>Graduate Program in Biophysics, University of California San Francisco, San Francisco, CA, USA.

<sup>4</sup>Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

<sup>5</sup>Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA.

<sup>6</sup>Gladstone Institutes, San Francisco, CA, USA.

<sup>7</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>8</sup>Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>9</sup>Mount Sinai Center for Therapeutics Discovery, Departments of Pharmacological Sciences, Oncological Sciences and Neuroscience, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>10</sup>Enamine Ltd., Kyiv, Ukraine.

<sup>11</sup>Institute of Bioorganic Chemistry and Petrochemistry, National Ukrainian Academy of Science, Kyiv Ukraine

<sup>12</sup>National Taras Shevchenko University of Kyiv, Kyiv, Ukraine

<sup>13</sup>Chemspace, Riga, Latvia

<sup>14</sup>QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA

<sup>15</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

<sup>16</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA

<sup>17</sup>Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>18</sup>The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

† Contributed equally.

\*Corresponding authors: [bshoichet@gmail.com](mailto:bshoichet@gmail.com) (B.K.S.); [m.vedadi@utoronto.ca](mailto:m.vedadi@utoronto.ca) (M.V.); [jir322@gmail.com](mailto:jir322@gmail.com) (J.J.I.)

### 3.1 Abstract

An under-explored target for SARS-CoV-2 is non-structural protein 14 (nsp14), a crucial enzyme for viral replication that catalyzes the methylation of *N*7-guanosine of the viral RNA at 5'-end; this enables the virus to evade the host immune response by mimicking the eukaryotic post-transcriptional modification mechanism. We sought new inhibitors of the S-adenosyl methionine (SAM)-dependent methyltransferase (MTase) activity of nsp14 with three large library docking strategies. In the first strategy, up to 1.1 billion make-on-demand (“tangible”) lead-like molecules were docked against the enzyme’s SAM site, seeking reversible inhibitors. On de novo synthesis and testing, three emerged with IC<sub>50</sub> values ranging from 6 to 43 μM, each with novel chemotypes. Structure-guided optimization and *in vitro* characterization supported their non-covalent mechanism. In a second strategy, docking a library of 16 million tangible fragments revealed nine new inhibitors with IC<sub>50</sub> values ranging from 28 to 275 μM, with the most potent having a ligand efficiency (LE) of 0.30 kcal/heavy-atom. In a third strategy, a new library of 25 million tangible electrophiles was docked to covalently modify Cys387 in the SAM binding site. Seven inhibitors emerged with IC<sub>50</sub> values ranging from 3.5 to 39 μM, the most potent being a reversible aldehyde. Initial optimization of a second series yielded a 7 μM acrylamide inhibitor. Three inhibitors characteristic of the new series were tested for selectivity against 30 human protein and RNA MTases, with one showing moderate selectivity and one showing high selectivity. Overall, 32 inhibitors encompassing 11 chemotypes had IC<sub>50</sub> values <50 μM and 6 inhibitors in 4 chemotypes had IC<sub>50</sub> values <10 μM. The inhibitors are among the first non-SAM-like inhibitors of nsp14, and constitute multiple starting points for optimizing towards antiviral activity.

### 3.2 Introduction

The Covid-19 pandemic has inspired a search for targets whose inhibition would combat the virus. Fruits of such efforts have been the development of Paxlovid (1), an inhibitor of the major protease (MPro) of SARS-CoV-2, of Molnupiravir (2,3), a disruptor of viral RNA polymerization, and the introduction of Remdesivir (4,5), an RNA-dependent RNA polymerase (RdRp) inhibitor first developed to treat Ebola virus. These targets are well-precedented in antiviral research, with successful drugs treating analogous enzymes for HIV, HCV, RSV, HBV, HCMV, HSV, HPV, and human influenza virus (6-9), among others, and these SARS-CoV-2 enzymes have been the focus of enormous efforts among many groups (1,4,5,10-13). Other SARS-CoV-2 enzymes have attracted less attention, likely because there is less precedence for their targeting as antivirals. Nevertheless, enzymes like the macrodomain (14,15) and the papain like protease of nsp3 (16), and the MTases nsp10-nsp16 complex (17) and nsp14 play key roles in the virulence of SARS-CoV-2 (18-21). While they have little precedence as antiviral drug targets, they seem attractive as novel enzymes for antiviral drug discovery.

Among these, the nsp14 SAM-dependent MTase seemed attractive. This enzyme catalyzes the methylation of the *N7* position of the terminal guanine of viral RNA, forming a cap-0 structure similar to those in eukaryotic mRNA, which are required for translation (18,22-26). Subsequently, nsp10-nsp16 complex methylates the 2'-O of the cap ribose to form cap-1 on the 3' end. The capping of viral RNA by nsp14 evades the host innate immune response to viral RNA, while ensuring efficient ribosome binding and engagement of the host-translational complex. Deletion of nsp14 is thought to eliminate viral virulence, confirming its importance and potential status as a SARS-CoV-2 drug target (18,19).

While SAM-dependent MTases have little precedence in antiviral chemotherapy, these enzymes have long been targeted in cancer chemotherapy (27-29). The binding determinants of these enzymes have been explored, especially in the SAM site, and several inhibitory analogs of the co-factor are available (30). This has supported the determination of the crystal structure of nsp14 complexed with s-adenosyl homocysteine (SAH) (23), along with other structures (23,31), and the development of an enzyme inhibition assay (32,33). The latter revealed relatively potent SAM-like inhibitors. However, most of these were relatively large and charged, likely reducing permeability and bioavailability, and as SAM analogs likely to have activities against other MTases, especially of human Class I MTases (33). This makes them problematic as leads to new antivirals.

Building off the structural and enzymatic work, we sought to discover novel scaffolds, dissimilar to that of the SAM-like inhibitors previously investigated, that would complement the nsp14 structure well but with better physical properties. We adopted a structure-based docking approach, where large libraries of “tangible”, make-on-demand molecules were fit into the SAM binding site of nsp14. Those that fit well (34-38) were prioritized for synthesis and testing. Essentially three libraries were docked: one composed of up to 1.1 billion lead-like molecules (39), one composed of 16 million fragment-like molecules, and a newly-constructed library of 25 million electrophiles, seeking those that could covalently modify the active-site Cys387 of nsp14 (**Figure 3.1**). Inhibitors emerged from all three campaigns, and subsequent structure-based optimization led to several classes of inhibitors binding in the low  $\mu\text{M}$  range. Methodologically, it was interesting that the fragment screened revealed perhaps the most diverse set of compounds, and the set most useful

for characterizing the binding site. This has also been seen for other SARS-CoV-2 targets such as Mpro (40) and Mac1 (14,15) and is a point to which we will return.

### 3.3 Results

#### Ultra-large scale docking against nsp14 identifies novel inhibitors

Due to the lack of SARS-CoV-2 nsp14 protein structure when this project began, we initially used the *N7*-MTase domain of SARS-CoV-1 nsp14 (PDB ID 5C8S) (23) for retrospective control calculations that helped us to validate the binding of known ligands. The SAM binding site of SARS-CoV-1 nsp14 was used without any modifications as the active site residues are conserved in both SARS-CoV-1 and SARS-CoV-2 *N7*-MTase domain of nsp14 (**Figure 3.1A**). These control calculations helped us to confirm that we could preferentially dock known MTase adenosyl group-containing compounds (SAM, SAH, and Sinefungin) and other known MTase inhibitors including LLY283 (41), BMS-compd7f (42), and Epz04777 (41-45) in favorable geometries with high ranks versus 300 property matched decoys (46,47). With the recent determination of the cryo-EM structure of the nsp10-nsp14 complex (PDB ID 7N0B) (31), the retrospective control calculations were also validated in the SAM binding pocket of *N7*-MTase domain of nsp14 of SARS-CoV-2. Comparison of the C $\alpha$  carbons of SARS-CoV-2 *N7*-MTase domain of nsp14 with the SARS-CoV-1 *N7*-MTase domain yielded an RMSD of 0.938 Å. The SAM binding site of SARS-CoV-2 nsp14 exhibited major differences in the position of side chain of amino acid residues Arg289, Asp352 and Gln354 with respect to the SARS-CoV-1 nsp14 (**Figure 3.1A**).

Seeking non-covalent inhibitors, we first docked over 680 million molecules, mostly in the “lead-like” range of the ZINC20 database (e.g., molecular weight  $\leq$  350 amu, cLogP values  $\leq$  3.5), against the SAM site of nsp14 from SARS-CoV-1 (PDB ID 5C8S) (23), as the structure of the enzyme from SARS-CoV-2 was unavailable at that time. Each library molecule was sampled for complementarity to the SAH site, with an average of 3438 orientations sampled, and for each

orientation about 187 conformations—over  $3.6 \times 10^{14}$  ligand configurations were sampled in the site in 121,018 core hours (about 5 days on 1000 cores). Seeking only novel chemotypes, molecules topologically similar to known SAM analogs were discarded. Compounds remaining were clustered based on ECFP4 fingerprints to identify unique chemotypes. The cluster representatives—most of them—were prioritized for interactions with Trp292, Gly333, Asn334, Asp352, Ala353, Phe367, Tyr368 and Val389 using LUNA (48). Molecules with strained conformations were deprioritized (38). Of the remaining molecules, the best scoring 5000 were visually inspected for key interactions and for unfavorable features, such as uncomplemented polar groups buried in the active site, using Chimera (49). Ultimately, 93 molecules, each in a different scaffold, were de novo synthesized and tested for enzyme inhibition at 30 and 50  $\mu\text{M}$  concentrations, measuring the transfer of [ $^3\text{H}$ ]-methyl from the SAM methyl donor onto the cap structure of an RNA substrate (GpppAC<sub>4</sub>) (see Methods). Those molecules that inhibited by more than 50% were considered active. Of the 93 molecules tested, only **ZINC475239213** (**'9213**) was active, a hit rate of 1% (hit rate = number-inhibited/number-physically-tested). This molecule had an IC<sub>50</sub> of 20  $\mu\text{M}$  in concentration-response (**Figure 2.2A**, middle panel). In its docked pose, the base-like moiety of **'9213** hydrogen-bonds with backbone amides of Ala353, Phe367 and Tyr368, while more distal parts of the molecule hydrogen-bond with Gln354 and Lys336 (**Figure 2.2A**, right panel). Van der Waals and stacking interactions are also apparent in the docked pose; overall these interactions resemble those observed among SAM and established SAM-like inhibitors but are made with different ligand groups.

With the determination of the cryo-EM structure of the SARS-CoV-2 nsp10-nsp14 complex (PDB ID 7N0B) (23), and the development of a larger “tangible” ZINC22 (50) library of 1.1 billion



molecules, we launched a second docking screen. The same retrospective control calculations were performed to optimize docking parameters, leading to similar sampling and calculation times. Following the same prioritization strategy as before, but seeking different chemotypes, 72 diverse molecules, were de novo synthesized and experimentally tested for inhibition of nsp14 (see Methods). Two inhibitors were found, **ZINC730084824** ('**4824**) with an IC<sub>50</sub> of 43 μM and **ZINC61142882** ('**2882**) with an IC<sub>50</sub> of 6 μM (**Figure 3.2B-C**), a hit rate of 2.7 %. The origins of the low hit rates for these two initial screens, and strategies to improve upon them, will be considered below.

### **Optimization of the non-covalent compounds**

To improve the affinity of the three non-covalent active docking hits, we sought analogs among the 20 billion tangible molecules that have been enumerated in a version of the REAL database (<http://enamine.net/compound-collections/real-compounds/real-space-navigator>), using substructure and similarity searches in the SmallWorld (<http://sw.docking.org>) and Arthor (<http://arthor.docking.org>) search engines (NextMove Software, Cambridge UK) (51). Conservative analogs were prioritized, their structures and physical properties were calculated, and then docked into the nsp14 SAM binding site. Analogs that docked to interact via  $\pi$ - $\pi$  stacking with Phe367 or Trp292, and that appeared to hydrogen-bond with Tyr368, Ala353, Asp352 and Asp388, were prioritized. Overall, 12, 20 and 36 analogs of '**9213**, '**4824** and '**2882**, respectively, were synthesized and tested for enzyme inhibition (**Figure 3.3**). The affinities of 1 of the 36 analogs of '**2882** was same as the initial hit, with an IC<sub>50</sub> of 6 μM for compound '**1988** (**Figure 3.3C**, **Table 3.S1**). The docked pose of '**1988** suggests a new hydrogen bond with Asp352 while conserving previous hydrogen bonds with Tyr368 and Ala353 (**Figure 3.3C**). Two '**9213** analogs,

**Z5347169163** (**'9163**, IC<sub>50</sub> 15 μM) and **ZINC001342858621** (**'8621**, IC<sub>50</sub> 19 μM), also had similar affinities (**Figure 3.3A, Table 3.S2**). For **'4824**, modest 2-fold improvement was observed for analogs **ZINC000916131631** (**'1631**, IC<sub>50</sub> 25 μM) and **Z5347186947** (**'6947**, IC<sub>50</sub> 19 μM) (**Figure 3.3B, Table 3.S3**). Even here, the SAR was revealing. While little improvement was seen over the parent **'4824** or **'9213**, for instance, many of the analogs tested remained relatively potent, with IC<sub>50</sub> values often below 40 μM (**Table 3.S1, Table 3.S2, Table 3.S3**). Moreover, replacement of the Michael acceptor vinyl group of **'4824** in analogs **'6947** and **Z5347186943** (**'6943**, IC<sub>50</sub> 40 μM), and removing the nitrile in analogs **'1631** and **'6947**, left molecules that remained active. In **'9213**, this was also the case with the removal of the nitrile group in analogs **'9163** and **'8621**. These analogs support the idea that docking hits **'9213** and **'4824** are acting via a non-covalent mechanism, as modeled. Assessment of reversibility by jump dilution also suggest that inhibition by **'4824** is not due to the nitrile electrophile (**Supplementary Figure 3.1**). **'1988** also appeared to be a non-covalent inhibitor in rapid dilution experiments (**Supplementary Figure 3.1**). In addition, **'1988** and **'4824** showed SAM- and RNA-competitive pattern of inhibition (**Supplementary Figure 3.2**).

### **Docking 16 million fragment-like molecules**

With only 3 inhibitor scaffolds discovered by lead-like docking, we thought to step back and interrogate the site with fragment-based docking. Fragment screens explore more chemical space than a larger lead-like library (*14,52,53*), which may be helpful for an under-explored site where warheads and key residue interactions have not been characterized. With the proviso that they have lower affinities, fragments also have higher hit rates in empirical (*54*) and docking screens (*52,53,55*) than do lead-like molecules, providing a richer tiling of the binding site by ligand

functional groups. Indeed, a strategy of fragment-docking to characterize the binding site followed by large-library lead-like docking was effective against another under-studied SARS-CoV-2 enzyme, Mac1 (14,15). Accordingly, from the 16 million molecule fragment-like set (e.g., molecular weight  $\leq 250$ , cLogP  $\leq 2.5$ ) in ZINC22, we targeted the full SARS-CoV-2 (PDB 7N0B) SAM binding site, the adenine portion of that site, and the SAM-tail region by modifying the DOCK pseudoatoms or “matching spheres” used, in three independent campaigns (**Figure 3.4A**) (Methods) (56). Overall, 14,406,946, 14,124,978, and 14,908,652 million molecules were scored, respectively. For each, the top-ranked 300,000 fragments were filtered as above, and the remaining fragments were clustered by topological similarity. Top-ranking cluster heads were visually inspected in Chimera (49) for favorable interactions, prioritizing those in the adenine site campaign for hydrogen bonds to Tyr368 and Ala353, and hydrophobic interactions with Phe367 (48). For the SAM-tail docking screen, interactions with Gly333 were prioritized, with additional interactions were selected for such as Gln313 and Asn386. For fragments docked against the entire SAM binding site, a combination of these interaction criteria were used. Ultimately 68 fragments were prioritized, of which 53 were successfully synthesized (78% fulfilment rate) (see Methods).

Of these 53, 9 fragments had IC<sub>50</sub> values less than 300  $\mu\text{M}$  (LE 0.13 to 0.30 kcal/heavy atoms), with the most potent fragment, **‘0683** having an IC<sub>50</sub> of 28  $\mu\text{M}$  (**Figure 3.4, Supplementary Figure 3.3**). As with the larger lead-like inhibitors (above), **‘0683** was a competitive and presumably reversible inhibitor of both SAM and RNA binding, with an LE of 0.30 (**Supplementary Figure 3.2**). In their docked poses, the fragment inhibitors spanned the full SAM site, with eight docking in the adenine-site and one docking in the SAM-tail site (**Figure 3.4B**). The hits in the adenine site dock to hydrogen bond with Tyr368, Phe367, and Ala353, often

mimicking interactions of the adenine of SAM/SAH but with different functional groups and diverse chemotypes (**Figure 3.4, Supplementary Figure 3.4**). The fragment docking in the SAM-tail site, **6066**, hydrogen bonds with residues such as Gly333 and Gln313 in its docked pose; its activity was lower than in the more tightly defined adenine site, with an  $IC_{50}$  of 275  $\mu$ M (**Figure 3.4C, Figure 3.4E, Supplementary Figure 3.3**).

### **Curation of 25 million aldehyde and acrylamide electrophiles for covalent docking**

In a final strategy, we sought potential covalent electrophiles that could react with the active site Cys387 of the nsp14 MTase. Such covalent docking has been successful in campaigns that targeted catalytic serine and non-catalytic, active site cysteine and lysine residues in enzymes such as  $\beta$ -lactamase, Jak kinases (57), eIF4e (58), M<sup>Pro</sup> (59) and targets such as RSK2 and MSK1 (60). These earlier campaigns had been limited to several hundred thousand electrophiles, largely from “in-stock” libraries. With the advent of the ultra-large tangible libraries, we thought to curate a larger set of electrophiles, focusing on aldehydes and acrylamides. Searching smarts patterns allowed us to build databases of 7.3 million aldehydes and 17.7 million acrylamides. We compared our aldehyde and acrylamide libraries to those that can be found in other in-stock or physical screening libraries, including the UCSF Small Molecule Discovery Center (SMDC) (61), Molecular Libraries Small Molecule Repository of the NIH (MLSMR) (62), and the in-stock set curated in ZINC20 (51). By total numbers, the aldehyde library is 196- to 10,000-fold larger than the number of aldehydes in the other libraries, and the number of scaffolds is also much larger at 252- to 3,600-fold (**Table 3.1**). For acrylamides, there are 811- to 465,000-fold more in the new library than in the public and in-stock sets, encompassing 250- to 58,000-fold more scaffolds. For the acrylamides, there are on average 13 molecules per scaffold in the new libraries, compared to 2-4

per scaffold in the other libraries. For aldehydes, the databases are comparable with our new library averaging 7 molecules per scaffold and 2-9 molecules per scaffold in the other databases. Given the rising interest in covalent-based inhibitors (60,63), we have made these 25 million electrophiles openly available via <http://covalent2022.docking.org> in both 2D and DOCKoalvent 3D format (Methods), along with ZINC and Enamine codes for ready acquisition.

### **Covalent inhibitors from 25 million docking screen against Cys387**

Using three nsp14 docking parameterizations (above and Methods), 17.7 million acrylamides and 6.2 million aldehydes were docked against the SAM site adjacent to Cys387, using DOCK3.7 DOCKoalvent (57,58). Molecules were docked to form a covalent adduct with Cys387. Those with non-covalent DOCK3.7 scores < 0 kcal/mol were further filtered for internal strain (38), stranded hydrogen bond donors and acceptors, and molecules hydrogen bonding with either Tyr368, Ala353, or Gly333 (48). Lastly, 33,156 molecules across the setups were clustered for topological similarity, and 9,591 molecules were prioritized for visual inspection in Chimera (49). Overall, 92 molecules were purchased for *de novo* synthesis. Of 61 aldehydes and 31 acrylamides, 47 and 26 were successfully synthesized, respectively, a 79% fulfilment rate (see Methods). On experimental testing, hits were defined as having at least 50% inhibition at 100  $\mu$ M. For the aldehydes, 5 compounds were active of 51 tested, a hit rate of 10%; 5 had IC<sub>50</sub> values ranging 3.5  $\mu$ M to 39  $\mu$ M (**Figure 3.5, Supplementary Figure 3.5, Supplementary Figure 3.6**). The most potent were ‘4975, ‘1911, ‘1875, and ‘1889 with IC<sub>50</sub> values of 3.5  $\mu$ M, 8  $\mu$ M, 11  $\mu$ M, and 12  $\mu$ M, respectively (**Figure 3.5A**). Per filtering criteria, these most potent aldehydes hydrogen bond with Ala 353 or Gly333 and form additional hydrogen bonds in the pocket, including with Asn388, Arg310, Gln313, or Trp385, in their docked poses (**Figure 3.5D, Supplementary Figure 3.6**). Of

the acrylamides, 2 compounds of 26 tested had >50% inhibition at 100  $\mu$ M, for a hit rate of 8%. (**Figure 3.5C, Supplementary Figure 3.5**). The most potent were **acryl42**, with an  $IC_{50}$  of 32  $\mu$ M, and **acryl41** at 39  $\mu$ M. **Acryl42** docks hydrogen bonds with Ala353 and Gly333 in its docked pose, while **acryl41** is a longer molecule extending farther than SAM/SAH, making hydrogen bonds with Gly333 and Gly313 in its docked pose (**Supplementary Figure 3.6**).

In early optimization of **acryl42**, analog **acryl42\_10** was 4.5-fold more potent at 7  $\mu$ M with the addition of a methoxy (**Figure 3.5C**). Adding a hydroxyl in the same place in analog **acryl42\_11** resulted in an inactive analog, indicating the methoxy could be adding hydrophobic contacts, opposed to additional hydrogen bonds with the protein (**Table 3.S4**). We tested the importance of the free amide of the acrylamide warhead with methylation of analog **acryl42\_5**; the analog was inactive and could be because of rigidification or loss of a modeled hydrogen bond with the proximal Ala353.

We evaluated the covalent mechanism for the most potent covalent docking hits, aldehyde **'1911** and acrylamides **acryl40**, **acryl42**, and the **acryl42** analog, **acryl42\_10**, first by mass spectrometry analysis. **'1911** did not increase the molecular mass of nsp14 (**Supplementary Figure 3.7**), likely reflecting the reversible binding of aldehydes to cysteines (Cys387 of Nsp14). In rapid dilution enzymatic experiments, **'1911**, incubated at high concentrations, showed little residual inhibition when diluted below its  $IC_{50}$ , further supporting a reversible covalent mechanism (**Supplementary Figure 3.1**). While, **acryl40** did not form a measurable adduct by mass spectrometry, **acryl42** and its analog **acryl42\_10** did do so, supporting a covalent inhibition mechanism (**Supplementary Figure 3.7**). We also changed the acrylamide warhead to the saturated propanamide group in

compound **ZD160-68** resulting in no enzymatic inhibition, which furthered the support of acryl42 acting through covalent inhibition (**Figure 3.5C**). Overall, **acryl\_42** and its analog, **acryl42\_10**, appear to be irreversible covalent inhibitors, while **'1911** appears to be a reversible covalent inhibitor. We expect that **acryl\_40** is also acting as a covalent inhibitor but note that further mechanistic study of these classes is warranted.

### **Selectivity against human protein and RNA methyltransferases.**

Three of the most potent inhibitors were counter-screened against a panel of 30 human SAM-dependent MTases. Compounds were tested for inhibition of the enzymes at 10  $\mu\text{M}$ , then selected for  $\text{IC}_{50}$  determination if higher than 50% inhibition was observed. The non-covalent, 6  $\mu\text{M}$  lead-like inhibitor, **'1988**, showed only modest selectivity, inhibiting nine enzymes more than 50% with  $\text{IC}_{50}$  values ranging from 4 to 26  $\mu\text{M}$  (**Figure 3.6**). The apparently reversible covalent, 8  $\mu\text{M}$  **'1911** had much better selectivity, inhibiting only two histone MTases G9a and G9a-like protein (GLP) with  $\text{IC}_{50}$  values of 30  $\mu\text{M}$  and 14  $\mu\text{M}$ , respectively. The most selective compound was **'4824** with no inhibition greater than 50% for any enzyme in the panel.

### **Evaluating antiviral activity.**

Potent compounds were also evaluated in a viral infectivity assay. HeLa-ACE2 cells were pre-incubated with varying concentrations of inhibitors **'1911** and **'1988**, and then infected with SARS-CoV-2 using an assay that monitors build-up of spike protein (5,64,65). The reversible aldehyde covalent inhibitor **'1911** ( $\text{IC}_{50} = 8 \mu\text{M}$ ) did not show antiviral activity up to a concentration of 100  $\mu\text{M}$ —further study was blocked by cytotoxicity (**Figure 3.7**). Meanwhile, the antiviral activity of the non-covalent **'1988** (enzymatic  $\text{IC}_{50} = 6 \mu\text{M}$ ) was modest with an

antiviral IC<sub>50</sub> of 133 μM (**Figure 3.7, Supplementary Figure 3.8**). This low activity may reflect potential instability of the **'1988** ester, or its low cell permeability, which in MDCK cells was measured to be 1.4 x 10<sup>-6</sup> cm/s (**Table 3.S5**), among other factors. We also tested non-covalent lead-line compounds **'9163** (**'9213** analog), **'1988** (**'2882** analog), and **'1631** (**'4824** analog), covalent compounds **acryl42\_10** (**acryl42** analog), **acryl42**, **'4975**, **'1889**, and **'1911**, and fragment **'0683**, alongside sinefungin in cell viability assays in A549-ACE2 cells (**Figure 3.7, Supplementary Figure 3.8**). Sinefungin was not cytotoxic with a CC<sub>50</sub> of >1000 μM. Compounds **'9163**, **'1889**, and **'1631** had high cell toxicities, whereas the fragment **'0683**, non-covalent lead-like compound **'1988**, and covalent compounds **acryl42** and **acryl42-10** had favorable cell viabilities. The viability of **'1911** in both the A549-ACE2 and HeLa-ACE2 cells was favorable. While 19 series of nsp14 inhibitors have emerged from this effort, each a different scaffold, it is clear that further optimization of all, for both affinity and physical properties, will be necessary to advance them towards lead status for potential SARS-CoV-2 therapeutics.



### 3.4 Discussion

From this study emerge among the first nsp14 inhibitors unrelated to SAM, either topologically or by physical properties. Overall, 23 non-covalent, lead-like inhibitors across three scaffolds were found with IC<sub>50</sub> values less than 50 μM, providing SAR for additional optimization (**Figure 3.2**, **Figure 3.3**, **Table 3.S1**, **Table 3.S2**, **Table 3.S3**). Additional characterization and structure-based optimization demonstrated their competitive, non-covalent mechanism of action against nsp14 (**Supplementary Figure 3.1**, **Supplementary Figure 3.2**). The most active covalent inhibitors were the initial aldehyde docking hits, with IC<sub>50</sub> values ranging 3.5 to 12 μM, and the acrylamide analog **acryl42\_10** with an IC<sub>50</sub> of 7 μM, all modeled to modify Cys387 of nsp14 (**Figure 3.5**). Finding these depended on developing new tangible libraries of 25 million electrophiles—these have been made publicly available for community use (<https://covalent2022.docking.org>) (**Table 3.1**). Another nine families of inhibitors were revealed from docking a library of 16 million tangible fragments (**Figure 3.4**). While affinities were naturally lower than the best of the lead-like inhibitors, several fragments had mid-μM IC<sub>50</sub> values, and the four most potent had LEs 0.26 to 0.30 kcal/HAC. Taken together, 19 new chemotypes were found; of these, 11 had members with IC<sub>50</sub> values <50 μM.

SARS-CoV-2 nsp14 inhibitors described to date are SAM analogs (33,66,67) or fragments with extensive water networks (68). While the SAM analogs are widely-studied, they typically suffer from low cell-permeability, owing to their size and ionization state, and they suffer from low selectivity, owing to their high similarity to the shared co-factor of this large family of MTases. Conversely, the new molecules described here are smaller and mostly uncharged, and topologically unrelated to SAM (**Table 3.2**). These properties may support optimization for cell

permeability and bioavailability, and for selectivity. From counter-screening 30 SAM-dependent human protein and RNA MTases, **‘4824** was nsp14 selective, and **‘1911** only hit two very closely related MTases (G9a and GLP). **‘1988** showed less selectivity with significant inhibitory effect on 9 out of 30 enzymes on tested selectivity panel. Similarly, the previously reported SARS-CoV-2 nsp14 inhibitors also showed significant inhibitory effect on human protein arginine methyltransferases (PRMTs) and BCDIN3D, an RNA MTase (33). However, the substantial selectivity of **‘4824** against the most likely human off-targets is promising.

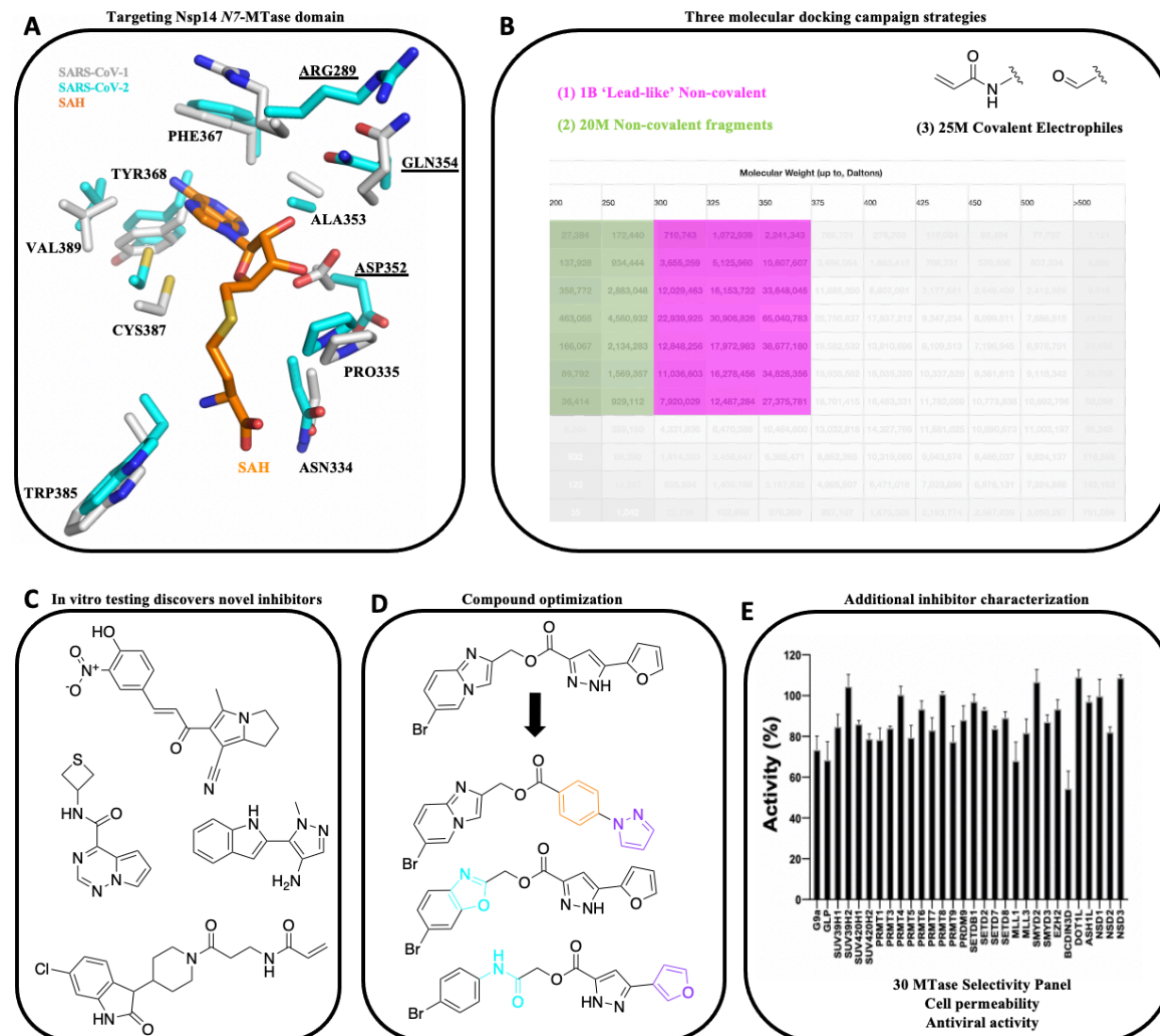
Apart from the particular inhibitors that were found, the strategy employed here may hold lessons for campaigns against the less-studied SARS-CoV-2 targets. Unlike enzymes for which investigational drug had been developed via similarities with other viruses, such as MPro or RdRp (2-5,69), the MTase of nsp14 had little inhibitor precedence on which to draw. Moreover, as a SAM-dependent enzyme with many related human enzymes, chemical novelty was important. Thus, as may be true with many SARS-CoV-2 targets, we could not leverage knowledge from previous chemical series other than SAM analogs. The lack of chemical precedence meant that these screens had a bootstrapping element to them—a small number of successes in early campaigns enabled us to optimize subsequent ones, contributing to improved hit-rates and affinities. We do note that our most informative screens—against the 16 million tangible fragments—occurred late in the campaign. Whereas there may still be skepticism about fragment docking, our own experience, not only here but also against the SARS-2 enzyme Mac1 (14,15) and in earlier studies (52,53,55), is that fragment docking can reveal multiple chemotypes with high-ligand efficiency and fidelity to subsequently determined crystal structures. Were we to begin again, we might have started with the fragment screen, leveraging the interactions it revealed

for campaigns against the larger, lead-like libraries. Such an approach may be useful against other understudied viral targets.

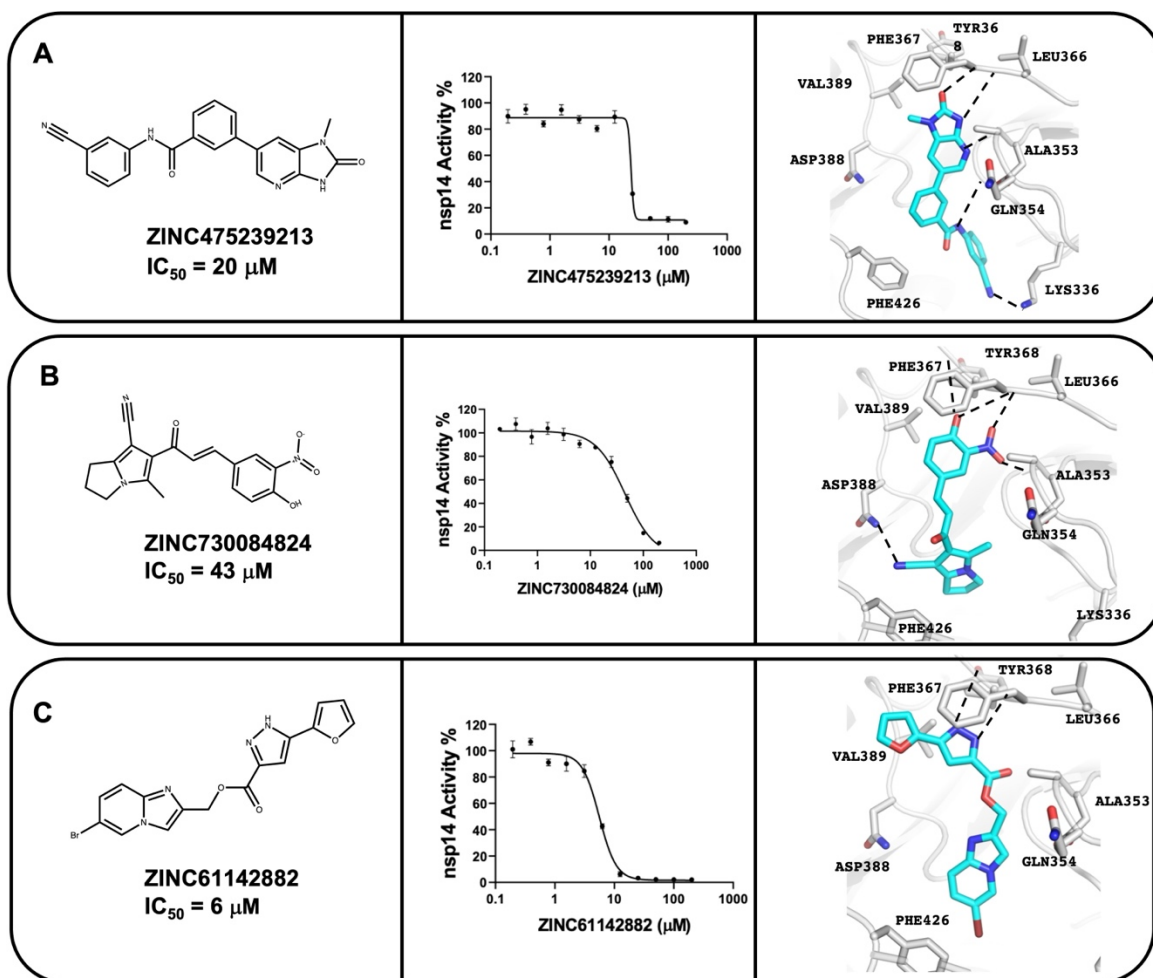
Certain caveats merit airing. Our most potent inhibitors are low- $\mu\text{M}$ , weaker than the most potent of the SAM analogs previously characterized for nsp14, the best of which inhibited in the 100 nM range (33,66,67). **'1911** needs additional characterization of its reversible covalent mechanism of inhibition, limited here by its reversibility in mass spectrometry analysis, and low- $\mu\text{M}$  activity in the rapid dilution experiments. Of the two molecules tested for viral inhibition in HeLa-ACE2 cells, only one showed modest anti-viral activity, with an  $\text{IC}_{50}$  value of 133  $\mu\text{M}$ , despite precedence of antiviral nsp14 inhibitors (70) (**Figure 3.7**). This may reflect cell-toxicities that occur before antiviral activity thresholds, barriers to cell-permeability (**Table 3.S5**), or simply lack of potency. Understanding these liabilities, and optimizing against them, will inform future compound advancement.

These caveats should not obscure the key observation of this study, the discovery of new 19 families of nsp14 inhibitors. These new inhibitors are not only diverse among themselves, but they do not resemble the SAM-related molecules previously described for nsp14 either topologically or by physical properties. They represent both non-covalent and covalent families, as well as multiple fragments that tile the binding site, in sites typically adjacent to and partly overlapping with the modeled sites of the larger, lead-like inhibitors. With the ongoing pandemic, they are being made openly available without restriction in the hopes that they may support a broad attack on this key but understudied target for antiviral drug discovery against Covid-19.

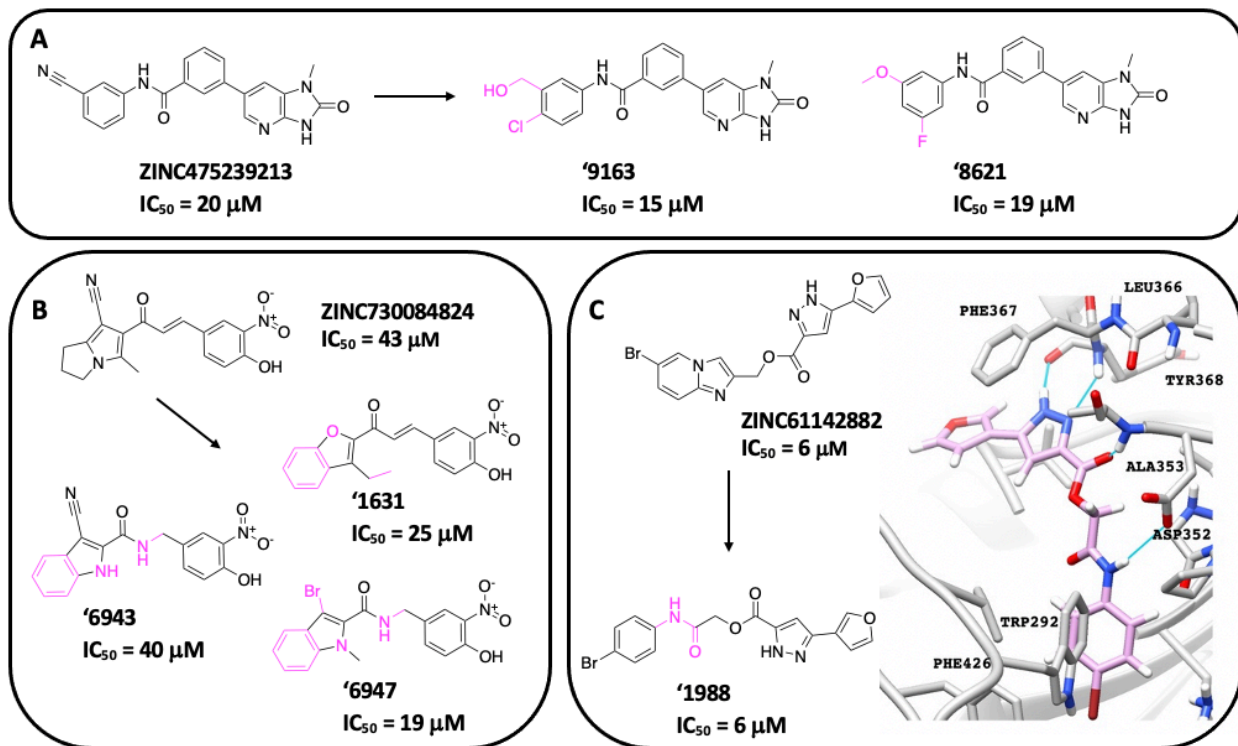
### 3.5 Figures



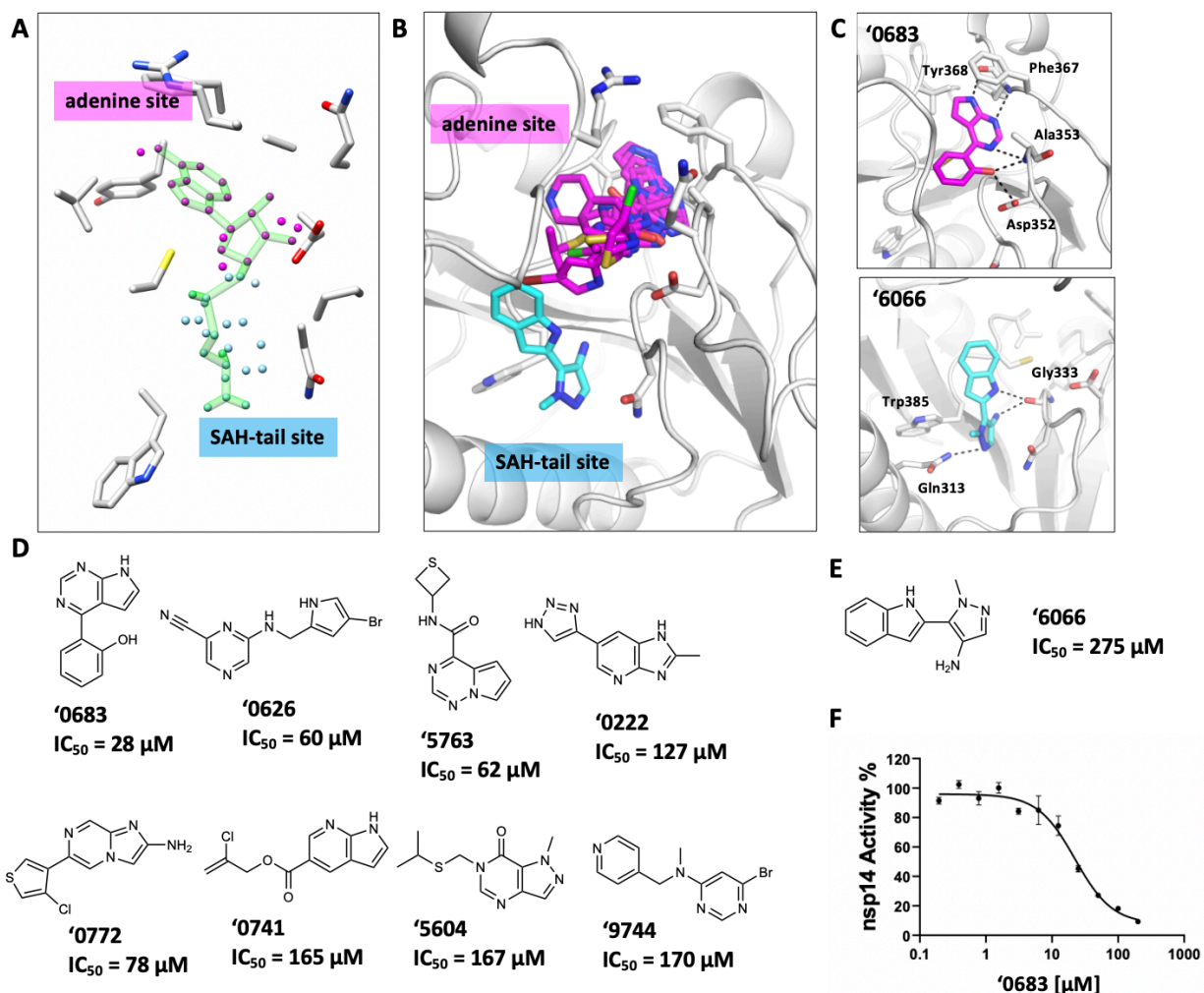
**Figure 3.1 Workflow for inhibitor discovery against N7-MTase domain of nsp14 using molecular docking.** (A) SARS-CoV-1 and -2 nsp14 MTase domains targeted with (B) three molecule subsets in molecular docking: lead-like non-covalent, fragment non-covalent, and acrylamide and aldehyde covalent electrophiles. (C) Diverse inhibitors discovered from each docking strategy followed by (D) compound optimization to improve potencies. (E) Best inhibitors evaluated for additional properties including MTase selectivity and antiviral efficacy.



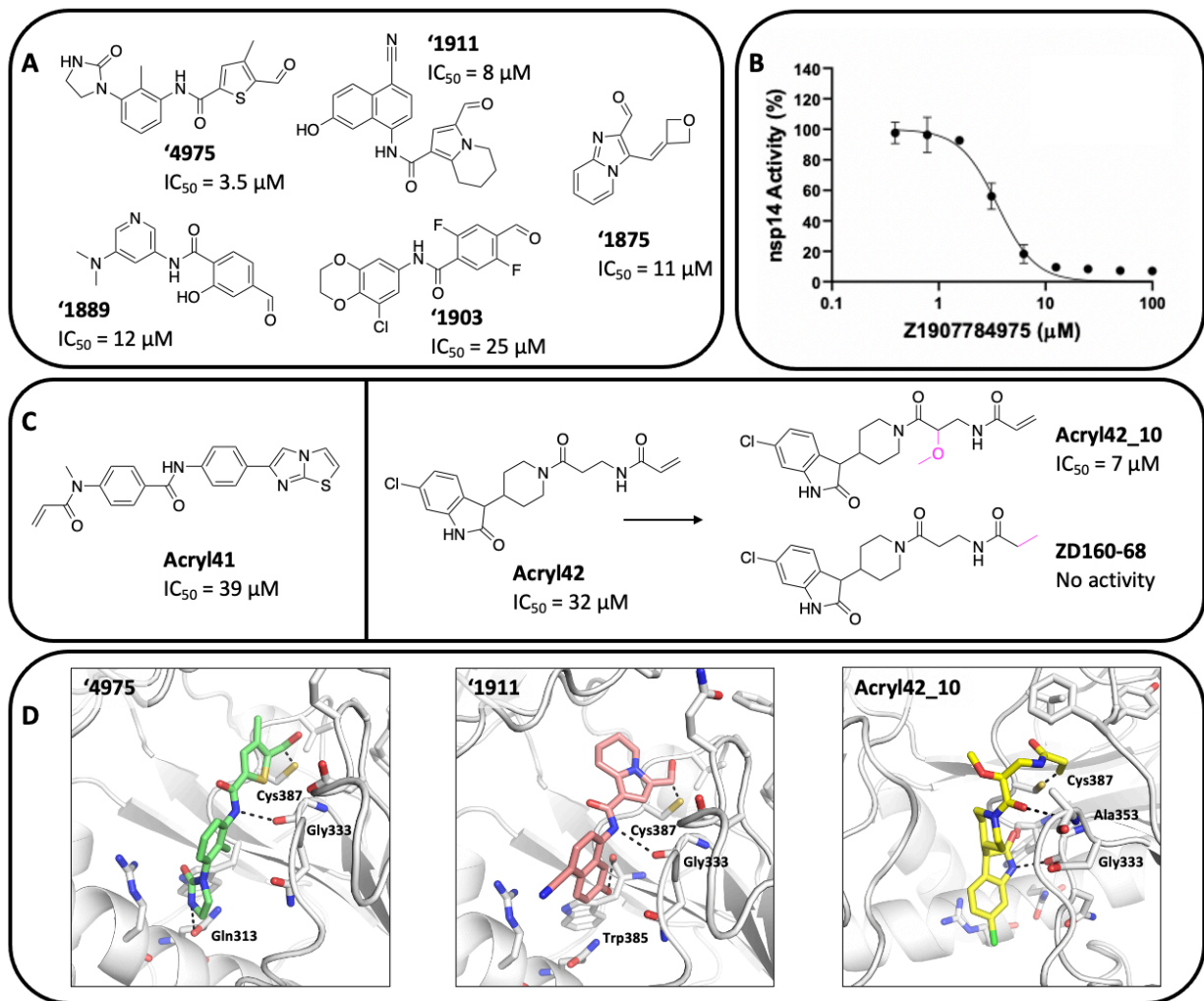
**Figure 3.2. Ultra-large scale docking identifies three nsp14 inhibitors with novel chemical scaffolds.** 2D chemical structures, concentration-dependent nsp14 MTase inhibition, and docked poses are represented for compounds ZINC475239213, ZINC730084824 and ZINC61142882 in panels A, B and C, respectively. SARS-CoV nsp14 and inhibitors are shown in gray and cyan carbons, respectively, and hydrogen bonds are shown as black dashed lines. The experiments were performed in triplicate.



**Figure 3.3. Hit optimization of the non-covalent compounds '9213, '4824, '2882.** 2D chemical structures of the parent hit and corresponding analogs with chemical changes represented in pink. (A) '9213 analogs with the nitrile removed have similar IC<sub>50</sub> values indicating non-covalent mechanism of action. (B) '4824 analogs with the nitrile or vinyl group removed have similar or more potent IC<sub>50</sub> values. (C) The '2882 analog '1988 is just as potent with opening of the bicyclic group. The '1988 docked pose (magenta carbons) is shown in SARS-CoV-2 nsp14 (grey carbons). The experiments were performed in triplicate.

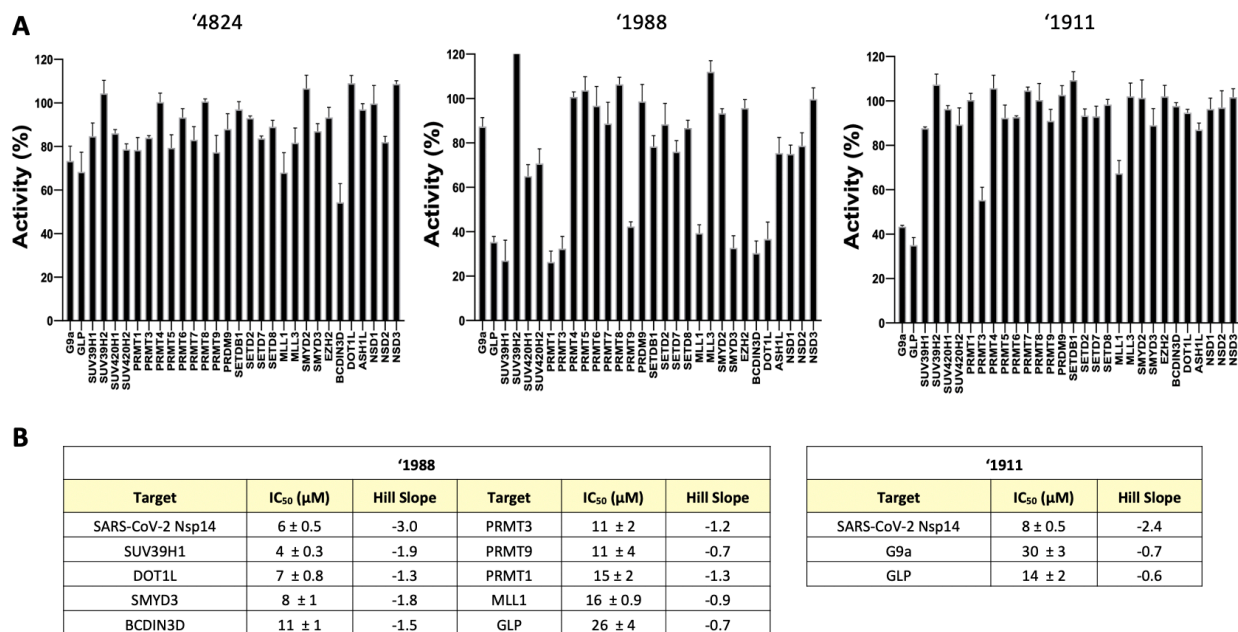


**Figure 3.4. Fragment inhibitors from 16M docking screen.** (A) Three sets of pseudoatoms, which define where ligands are sampled in the binding site (“spheres”) used in the docking screens (56) including the adenine-site spheres (pink), SAM-tail site spheres (blue), and a superset of both. SARS-CoV-2 nsp14 (grey carbons) with SAH (green carbons). (B) Overlay of all 9 fragment docking hits in the SAM binding site of SARS-CoV-2 (tan carbons). Docked poses for adenine-site inhibitors shown (pink carbons) and SAM-tail site inhibitors (cyan carbons). (C) Docked poses of the best adenine-site fragment ‘0683, and the SAM-tail site fragment ‘6066. (D) Eight adenine-site fragment hits shown with their respective IC<sub>50</sub> values. (E) The SAM-tail site fragment hit ‘6066. (F) Concentration response curve of ‘0683 in the N7-MTase inhibitory activity assay. For D and E, IC<sub>50</sub> values derived from concentration-response curves shown in Supplementary Figure 3. The experiments were performed in triplicate.

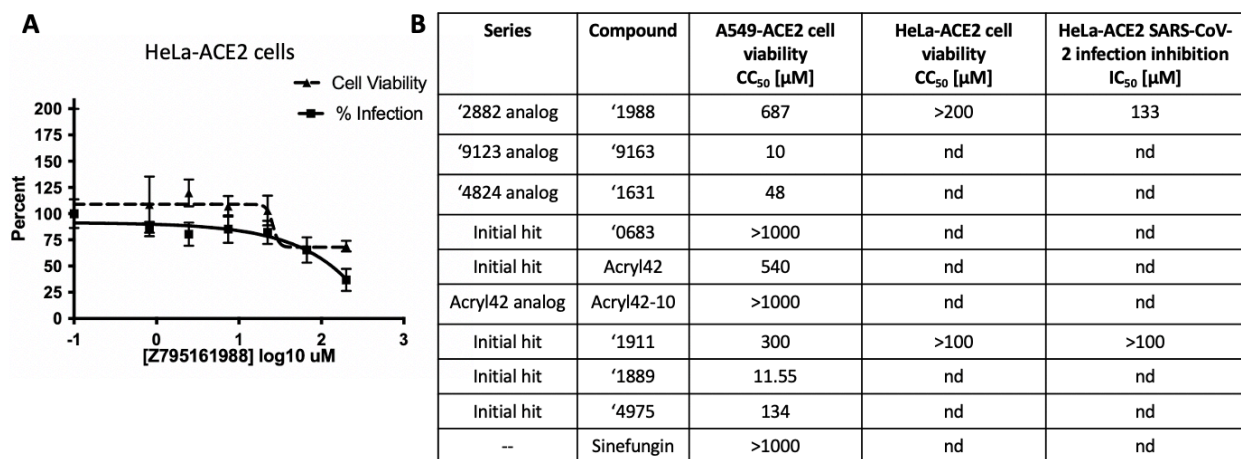


**Figure 3.5. Docking 25 million electrophiles reveals aldehyde and acrylamide inhibitors. (A)** Aldehyde docking hits. **(B)** Concentration-response curve for the most potent aldehyde '4975 in the *N7*-MTase inhibitory activity assay. **(C)** Acrylamide docking hits **acryl41** and **acryl42**, with analog **acryl42\_10** and inactive analog **ZD160-68**. **(D)** Docked poses of '4975, '1911, and modeled pose of analog **acryl42\_10**. The experiments were performed in triplicate and additional concentration-response curves found in Supplementary Figure 5.

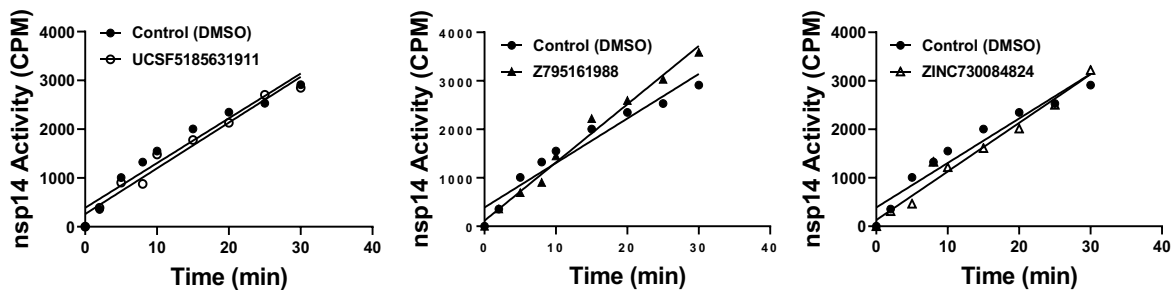




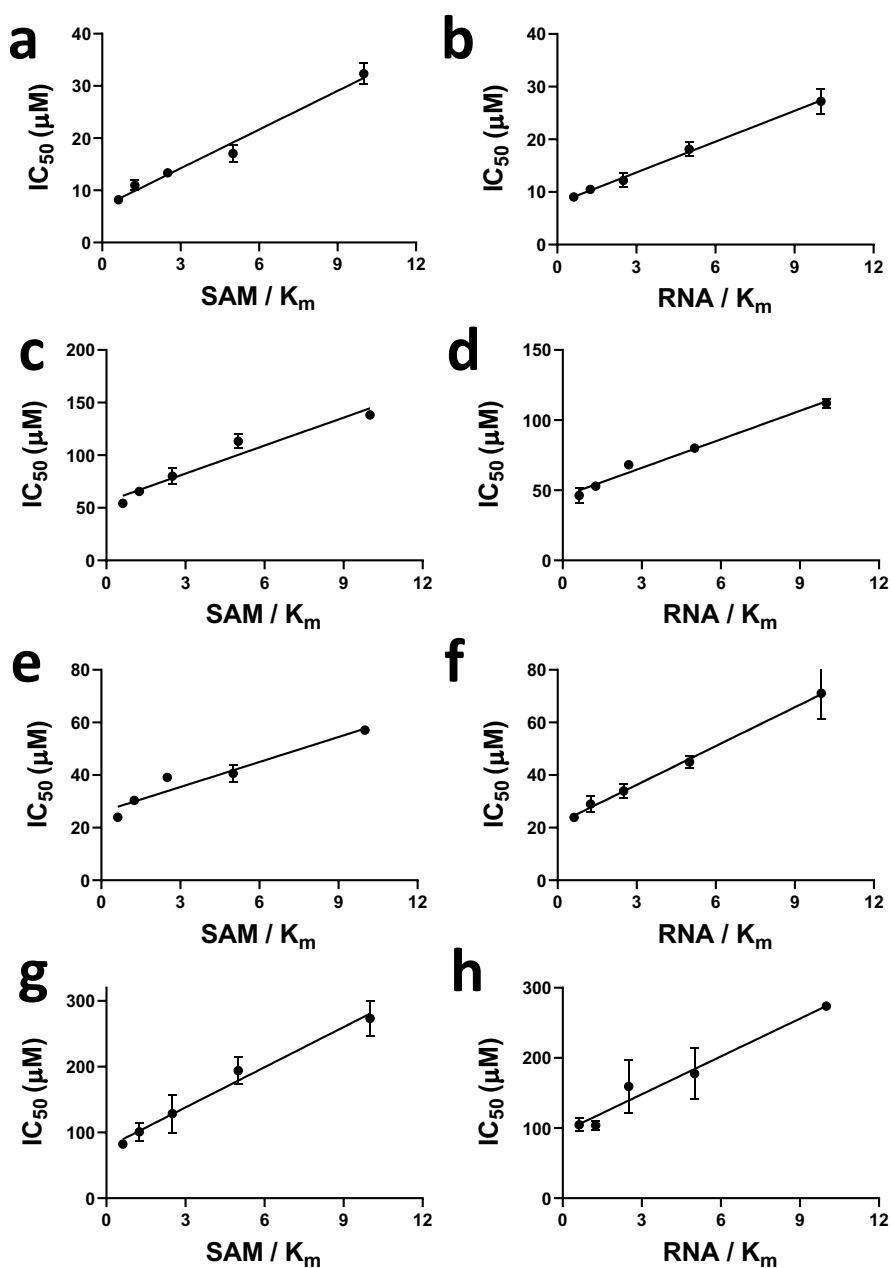
**Figure 3.6. MTase selectivity of docking-derived inhibitors.** (A) Compounds were tested against a panel of 30 SAM-dependent human protein and RNA MTases. Those with > 50% inhibition were prioritized for (B) IC<sub>50</sub> determination. The experiments were performed in triplicate.



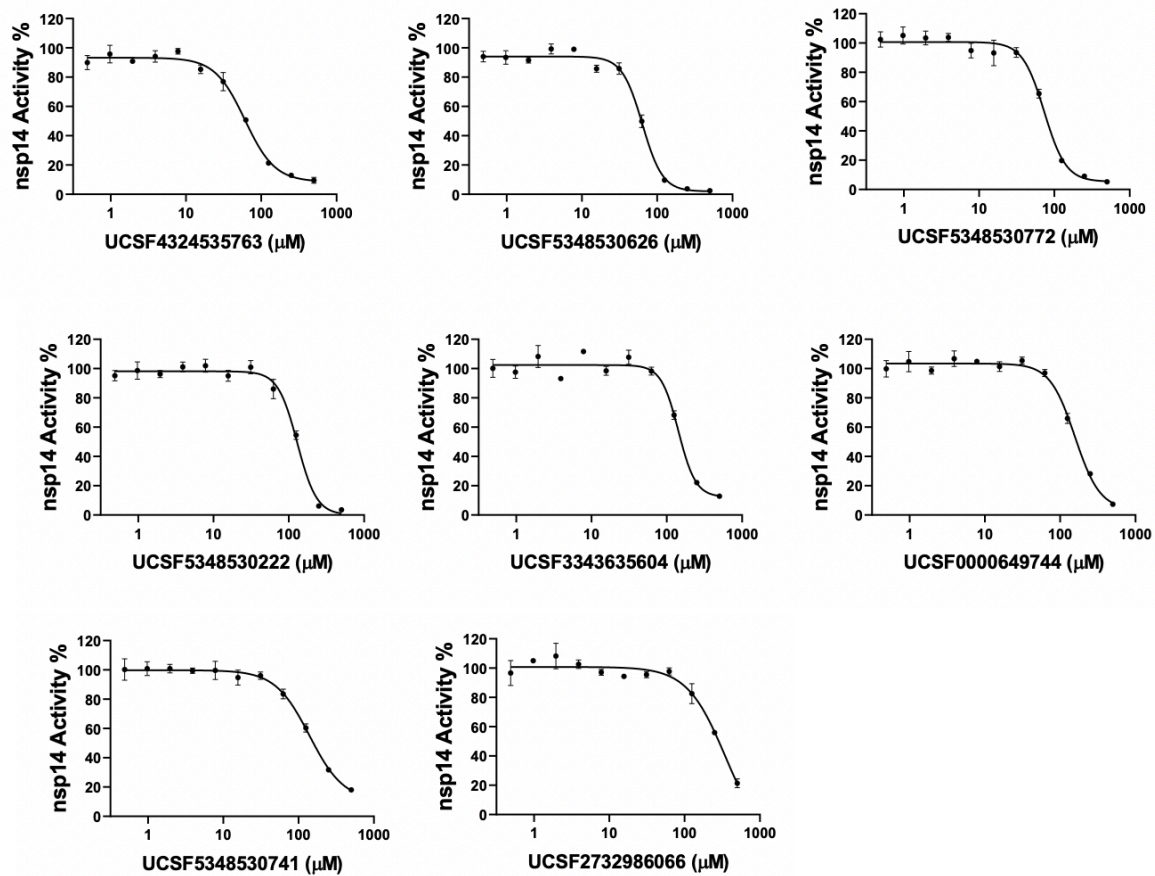
**Figure 3.7. Assessing the antiviral activity of nsp14 inhibitors. (A)** HeLa-ACE2 cell-based antiviral assay used to evaluate inhibition of SARS-CoV-2 infection and cytotoxicity of '1988. **(B)** Summary of cell viability and antiviral activities of inhibitors in HeLa-ACE2 and A549-ACE2 cells. The experiments were performed in triplicate. CC<sub>50</sub>, cytotoxicity concentration reducing cell viability by 50%. IC<sub>50</sub>, antiviral inhibitory concentration by 50%. nd, not determined.



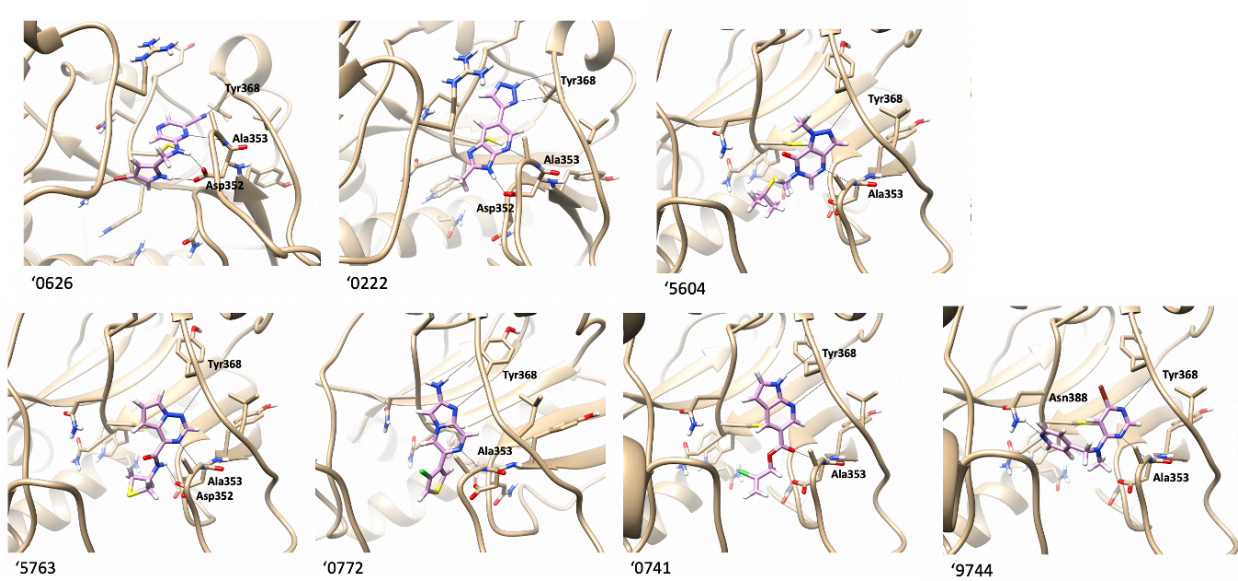
**Figure 3.S1. Assessment of reversibility of inhibition by rapid dilution.** The recovery of nsp14 activity after incubation with each inhibitor (‘1911, ‘1988, ‘4824) and rapid dilution was monitored. Nsp14 at 100-fold higher concentration than what required for activity measurement (1.5 nM) was incubated with each compound at 10-fold of  $IC_{50}$  value concentration for 1h at room temperature. Reaction mixtures were then rapidly diluted 100-fold into the assay buffer containing substrate RNA and SAM, and recovery of the nsp14 activity was monitored.



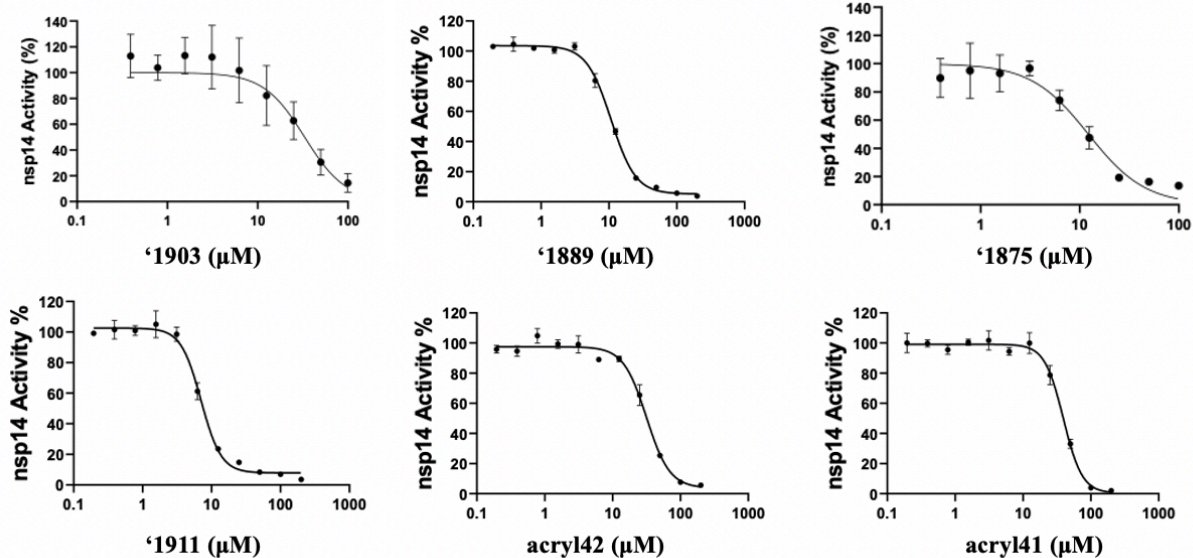
**Figure 3.S2. Mechanism of action of inhibitors.** Compounds (a, b) ‘1988, (c, d) ‘4824, (e, f) ‘1911, and (g, h) ‘0683 were evaluated for their inhibition of SAM and RNA binding. (a, c, e, g) IC<sub>50</sub> values were determined at a fixed concentration of RNA substrate (0.25 μM; 5xK<sub>m</sub>) and varying concentrations of SAM (up to 2.5 μM/10x K<sub>m</sub>) and (b, d, f, h) varying concentrations of RNA (up to 0.5 μM; 10x K<sub>m</sub>) and fixed 1.25 μM (5xK<sub>m</sub>) of SAM. The experiments were performed in triplicate. Linear increase in IC<sub>50</sub> values as the concentration of substrate is increased, indicated a competitive pattern of inhibition.



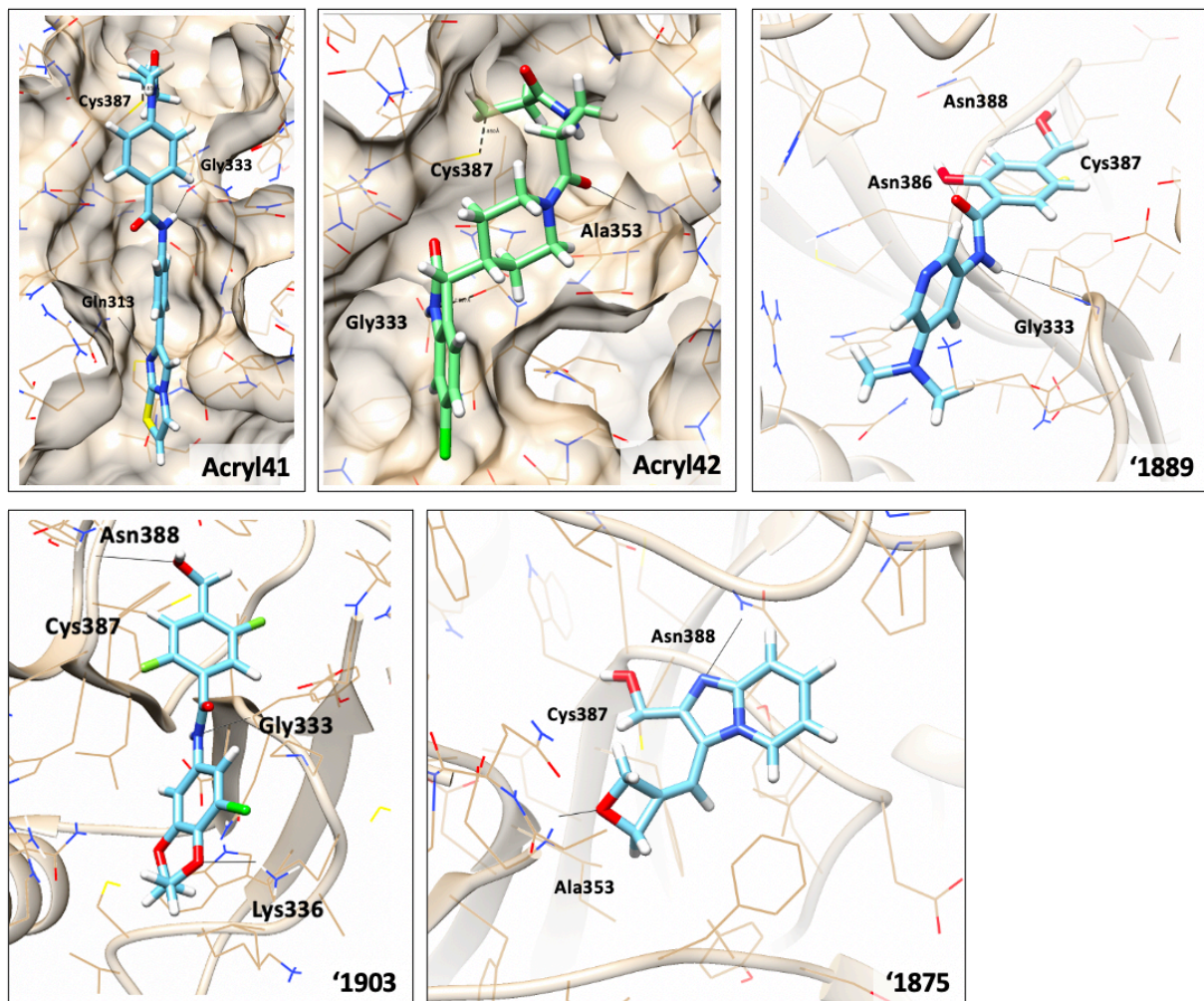
**Figure 3.S3. Concentration-response curves of non-covalent fragment hits.** Docking hits shown with their respective concentration-response curves in the *N7*-MTase inhibitory activity assay and docked poses. The experiments were performed in triplicate.



**Figure 3.S4. Docked poses of non-covalent fragment hits.**

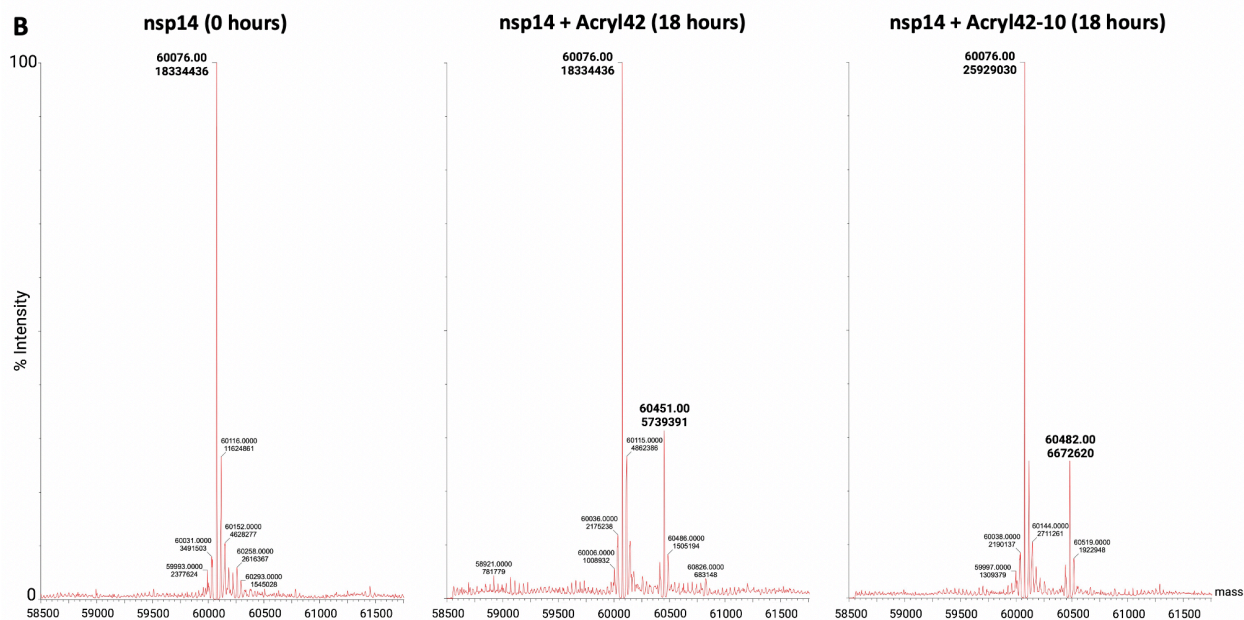
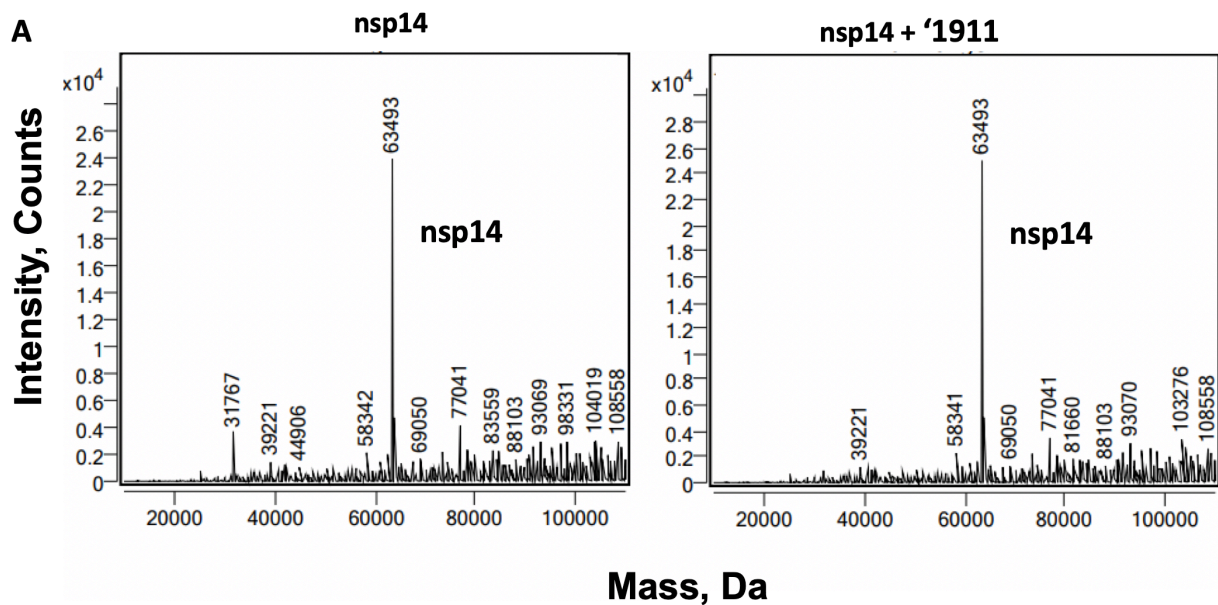


**Supplementary Figure 3.S5. Concentration-response curves of covalent hits.** Docking hits shown with their respective concentration-response curves in the *N7*-MTase inhibitory activity assay and docked poses. The experiments were performed in triplicate.

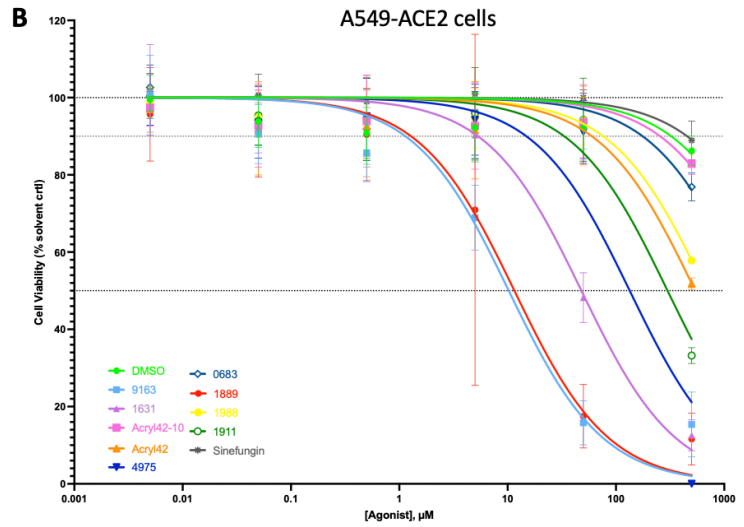
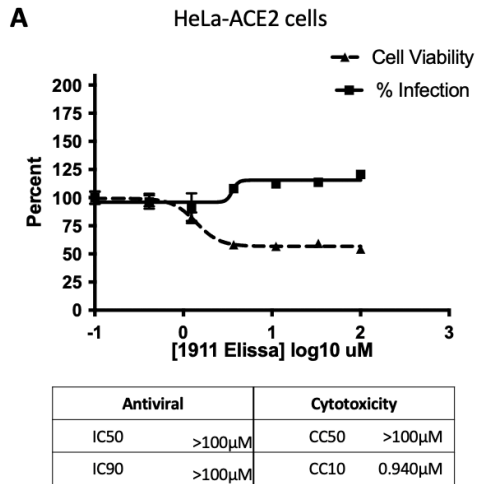


Supplementary Figure 3.S6. Docked poses of covalent hits.





**Supplementary Figure 3.S7. Evaluating possible covalent binding of compound '1911, acryl42, and acryl42-10 to nsp14 by mass spectrometry.** (A) Nsp14 was incubated in the absence (left) and presence (right) of compound 1911 at 20x molar excess for 2 hours at room temperature ( $22\pm 1$  °C). Both samples were analyzed by mass spectrometry. No change in molecular mass of nsp14 upon incubation with compound 1911 confirmed that it is not an irreversible covalent inhibitor. (B) LC-MS observed mass and peak intensity data of nsp14 with no compound (left), in the presence of 500  $\mu$ M acryl42 (middle), or 50  $\mu$ M acryl42-10 (right). The peak at 60451 corresponds to nsp14 adduct with acryl42 (middle; fraction adduct 0.238) and the 60482 peak corresponds to the adduct of nsp14 with acryl42-10 (right; fraction adduct 0.205) both incubated at room temperature for 18 hours.



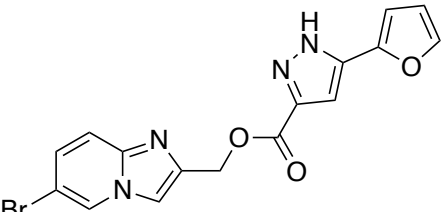
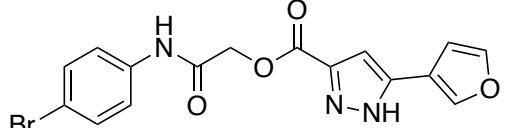
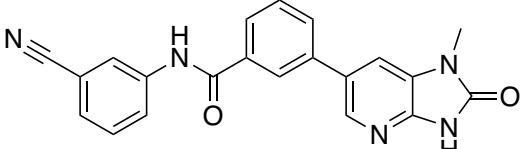
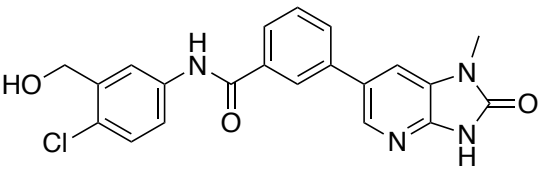
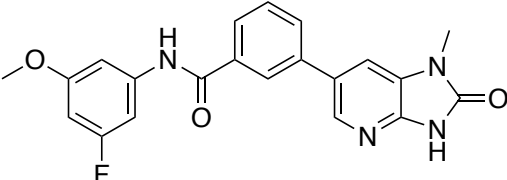
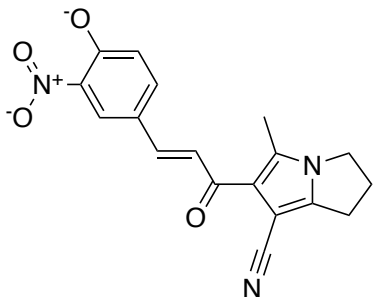
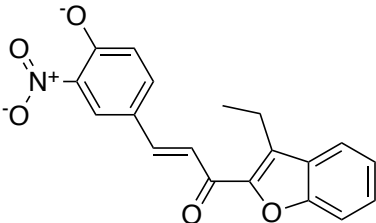
**Supplementary Figure 3.S8. Antiviral and cell viability experiments for nsp14 inhibitors.** (A) HeLa-ACE2 cell viability and antiviral assay. (B) Inhibitors and sinefungin in A549-ACE2 cell viability assay. Data shown were performed in triplicate.

### 3.6 Tables

**Table 3.1. Expanded DOCKoalvent electrophile databases.**

Database		Aldehydes		Acrylamides	
Name	Size	# molecules	# BM Scaffolds	# molecules	# BM Scaffolds
Covalent2022.docking.org	31,000,000,000	6,197,526	848,830	17,680,357	1,404,874
ZINC20 In-Stock	7,517,254	31,554	3,373	21,798	5,612
UCSF SMDC	690,125	615	235	38	24
MLSMR	406,098	908	407	63	27

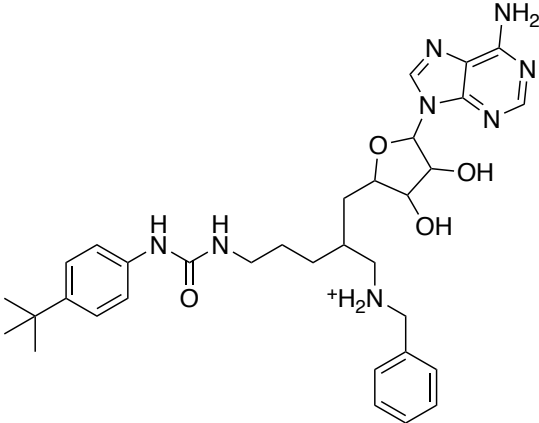
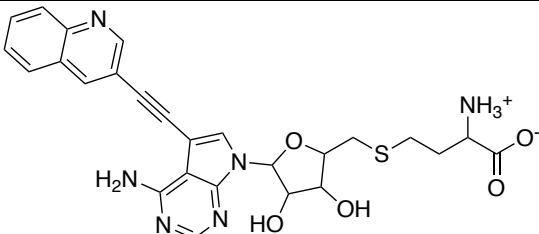
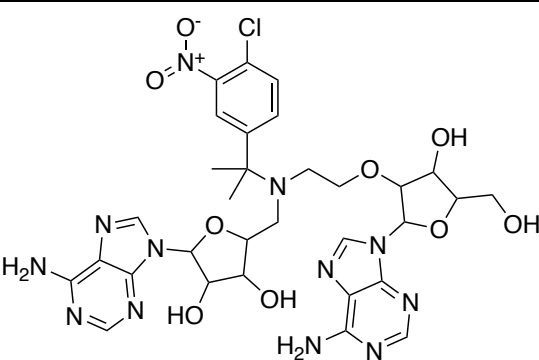
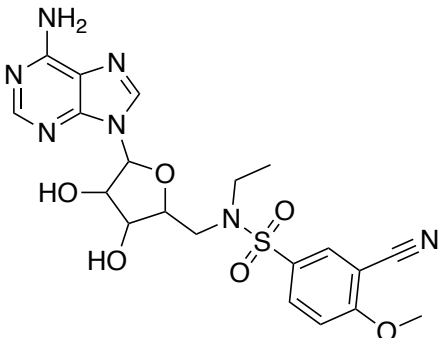
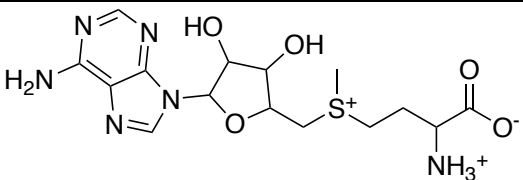
**Table 3.2. Docking inhibitors are novel, non-SAM like chemotypes.**

Compound name	IC <sub>50</sub> ( $\mu$ M)	2D chemical structure	Molecular Weight cLogP # charged groups	Similarity to SAM (TC)
ZINC61142882 (2882)	6		386.0 3.44 0	0.09
Z795161988	4		389.0 3.23 0	0.09
ZINC475239213 (9213)	20		369.12 3.05 0	0.09
Z5347169163	15		408.10 3.33 0	0.11
ZINC0013428586 21	19		392.13 3.33 0	0.09
ZINC730084824 (4824)	43		336.11 2.49 3	0.08
ZINC0009161316 31	25		336.09 3.87 3	0.11

Compound name	IC <sub>50</sub> ( $\mu$ M)	2D chemical structure	Molecular Weight cLogP # charged groups	Similarity to SAM (TC)
Z5347186953	20		387.01 3.49 1	0.09
Z5347186947	19		402.01 2.85 3	0.12
Z5347186943	40		335.08 1.95 3	0.09
*0683	28		221.07 2.33 0	0.10
*0626	60		277.00 2.05 0	0.10
*5763	62		234.06 0.57 0	0.04

Compound name	IC <sub>50</sub> (μM)	2D chemical structure	Molecular Weight cLogP # charged groups	Similarity to SAM (TC)
'0772	78		250.01 2.69 0	0.10
'4975	3.5		343.10 2.96 0	0.09
'1911	8		359.13 3.62 0	0.08
'1889	12		285.11 1.92 0	0.10
acryl42	32		375.13 2.31 0	0.09
acryl41	39		402.12 4.46 0	0.08
ald1903	25		353.03 3.45 0	0.10
acryl42-10	7		405.15 1.93 0	0.12

Compound name	IC <sub>50</sub> (μM)	2D chemical structure	Molecular Weight cLogP # charged groups	Similarity to SAM (TC)
SS148 <sup>a</sup>	0.07		408.12 -3.01 2	0.40
WZ16 <sup>a</sup>	0.19		396.20 -4.58 3	0.58
DS0464 <sup>a</sup>	1.1		473.18 0.30 0	0.51
JL2756A1 <sup>a</sup>	0.26		514.23 -3.58 2	0.48
SD0466 <sup>a</sup>	3.4		465.22 0.64 0	0.48
MTTR025495 <sup>a</sup>	17		297.09 -0.61 0	0.57

Compound name	IC <sub>50</sub> ( $\mu$ M)	2D chemical structure	Molecular Weight cLogP # charged groups	Similarity to SAM (TC)
compound8 <sup>a</sup>	95		617.36 2.70 1	0.46
compound16 <sup>a</sup>	0.004		534.17 -0.94 2	0.36
compound_d5 <sup>a</sup>	0.6		756.25 -0.11 2	0.40
compund_25 <sup>a</sup>	0.019		489.14 -0.38 0	0.44
SAM	--		399.15 -3.97 3	1.0

<sup>a</sup> Activities reported from literature (33,66,67).



**Table 3.S1. Compound optimization for '2882.**

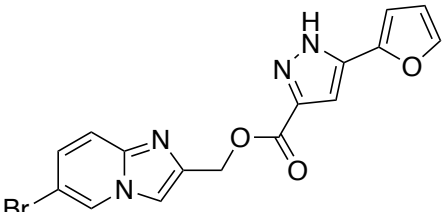
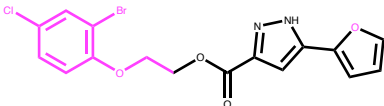
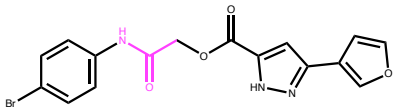
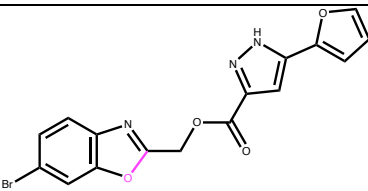
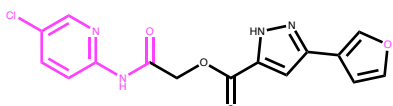
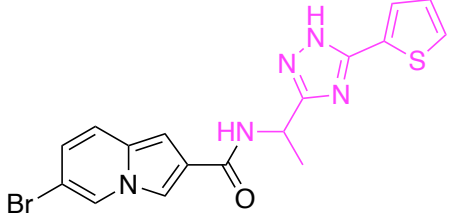
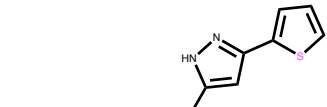
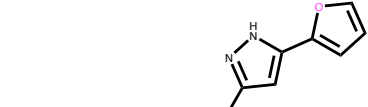

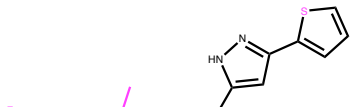
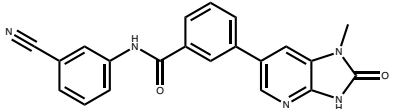
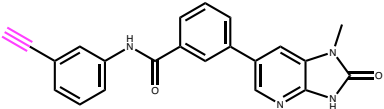
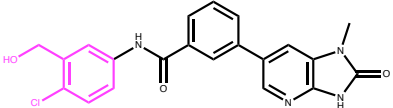
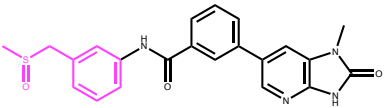
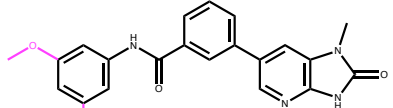
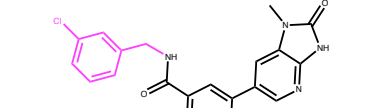
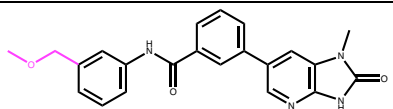
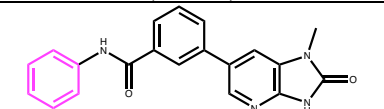
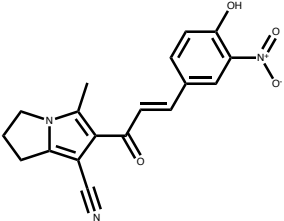
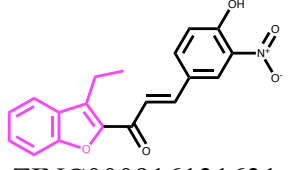
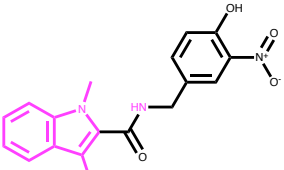
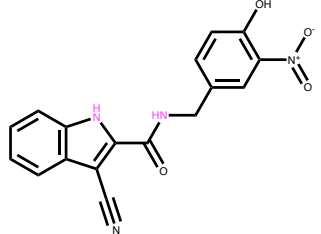
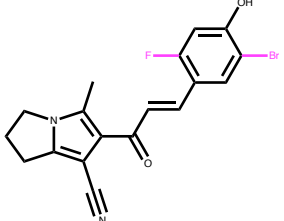
2D Chemical Structure	IC <sub>50</sub> ( $\mu$ M)	2D Chemical Structure	IC <sub>50</sub> ( $\mu$ M)
 <p><b>Initial hit ZINC61142882 ('2882)</b></p>	6	 <p>ZINC002325862871 ('2817)</p>	18
 <p>Z795161988 ('1988)</p>	6	 <p>Z1724303092 ('3092)</p>	21
 <p>ZINC000919327864 ('7864)</p>	7	 <p>Z5347191013 ('1013)</p>	30
 <p>ZINC001141402437 ('2437)</p>	15	 <p>ZINC002325862682 ('2682)</p>	38
 <p>ZINC000918899102 ('9102)</p>	17	 <p>ZINC001444146032 ('6032)</p>	28

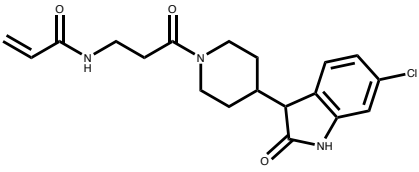
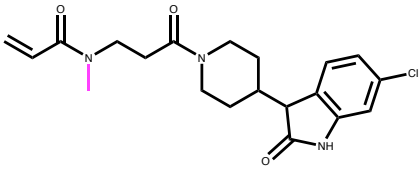
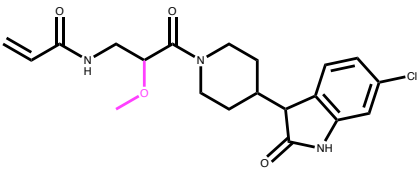
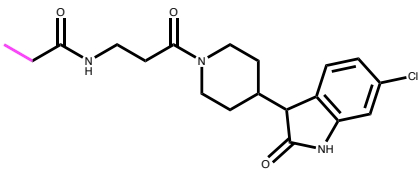
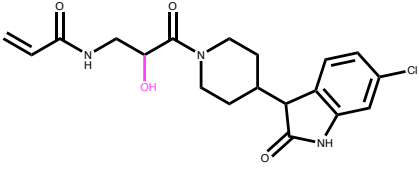
Table 3.S2. Compound optimization for '9213.

2D Chemical Structure	IC <sub>50</sub> (μM)	2D Chemical Structure	IC <sub>50</sub> (μM)
 <p><b>Initial hit ZINC475239213 ('9213)</b></p>	20	 <p>ZINC771823888 ('3888)</p>	27
 <p>Z5347169163 ('9163)</p>	15	 <p>ZINC000475232808 ('2808)</p>	31
 <p>ZINC001342858621 ('8621)</p>	19	 <p>ZINC475241593 ('1593)</p>	44
 <p>ZINC000475240670 ('0670)</p>	20	 <p>ZINC000355153196 ('3196)</p>	47

**Table 3.S3. Compound optimization for '4824.**

2D Chemical Structure	IC <sub>50</sub> (μM)	2D Chemical Structure	IC <sub>50</sub> (μM)
 <p><b>Initial hit</b> <b>ZINC730084824</b> <b>('4824)</b></p>	43	 <p><b>ZINC000916131631</b> <b>('1631)</b></p>	25
 <p><b>Z5347186947</b> <b>('6947)</b></p>	19	 <p><b>Z5347186943</b> <b>('6943)</b></p>	40
 <p><b>Z5347186953</b> <b>('6953)</b></p>	20		

**Table 3.S4. Compound optimization for acryl42.**

2D Chemical Structure	IC <sub>50</sub> (μM)	2D Chemical Structure	IC <sub>50</sub> (μM)
 <p style="text-align: center;"><b>Acryl42</b></p>	32	 <p style="text-align: center;">Acryl42_5</p>	NI
 <p style="text-align: center;">Acryl42_10</p>	7	 <p style="text-align: center;">ZD160-68</p>	NI
 <p style="text-align: center;">Acryl42_11</p>	>200		

NI, no inhibition

**Table 3.S5. Permeability of selected inhibitors and reference compounds.**

A-B and B-A permeability data in MDR1-MDCKII cells									
Test compound	P <sub>app</sub> <sup>a</sup> (AB), 10 <sup>-6</sup> cm/s				P <sub>app</sub> <sup>a</sup> (BA), 10 <sup>-6</sup> cm/s				Efflux ratio <sup>b</sup>
	1	2	Mean	SD	1	2	Mean	SD	SD
Atenolol	1.0	1.3	1.2	0.2	nd	nd	nd	nd	nd
Propranolol	24.3	21.3	22.8	2.1	14.6	12.3	13.5	1.6	0.6
Ketoprofen	17.3	19.0	18.2	1.2	16.4	16.0	16.2	0.3	0.9
Digoxin	0.6	0.7	0.6	0.0	10.3	9.6	10.0	0.5	15.4
Z795161988	1.4	1.3	<b>1.4</b>	0.1	0.5	0.4	0.5	0.0	0.3
Z5185631911	1.2	1.5	<b>1.3</b>	0.2	0.5	0.4	0.4	0.0	0.3
Z5185631889	1.1	1.1	<b>1.1</b>	0.0	0.3	0.3	0.3	0.0	0.2
Z5348530683	15.9	16.8	<b>16.4</b>	0.7	13.2	10.9	12.1	1.6	0.7

<sup>a</sup> Apparent permeability

<sup>b</sup> Efflux ratio is expressed as the quotient of P<sub>app</sub>(BA) to P<sub>app</sub>(AB)  
nd, not determined

### 3.7 Materials and Methods

#### Non-covalent ultra-large scale docking.

*N7*-MTase domain of SARS-CoV nsp14 (PDB ID 5C8S) (12) and the *N7*-MTase domain of SARS-CoV2 nsp14 from cryo-EM structure of the nsp10-nsp14 complex (PDB ID 7N0B) were used in two docking campaigns of >680 million “lead-like” molecules from the ZINC20 database (<http://zinc20.docking.org>) (51), and the ZINC22 >1.1 billion “lead-like” molecules (<http://files.docking.org/zinc22>), respectively, using DOCK3.7 (37). Forty-five matching spheres or local hot-spots generated from the crystal pose of SAM/SAH were used in the binding site for superimposing pre-generated flexible ligands and the different poses were scored by summing the different energies including; van der Waals interaction energies, Poisson-Boltzmann-based electrostatic interaction, and Generalized-Born/Surface Area-based ligand desolvation energies (35,36). Receptor atoms were protonated with Reduce (71), and partial atomic charges were calculated using united-atom AMBER force field (72). AMBER atom-types were also used for Poisson-Boltzmann electrostatic potential energy grids using QNIFFT (73), CHEMGRID (74) was used for calculating van der Waals potential grids, and SOLVMAP (35) was used to calculate the Generalized-Born/Surface Area grids for ligand desolvation.

The docking setup was optimized for its ability to enrich known MTase adenosyl group-containing compounds (SAM, SAH, and Sinefungin) and other known MTase inhibitors including Lly283, BMS-compd7f, and Epz04777 (41-45), in favorable geometries with high complementarity versus a set of property matched decoys (46,47). About 50 decoys were generated for each ligand that had similar chemical properties to known ligands but were different topologically. The best optimized docking setup was evaluated for enrichment of ligands over decoys using log-adjusted

area under the curve (logAUC values) (46,47). All docked ligands were protonated with Marvin (version 15.11.23.0, ChemAxon, 2015; <https://www.chemaxon.com>) at pH 7.4, rendered into 3D with Corina (v.3.6.0026, Molecular Networks GmbH; <https://www.mn-am.com/products/corina>), and conformationally sampled using Omega (v.2.5.1.4, OpenEye Scientific Software; <https://www.eyesopen.com/omega>). Before docking the lead-like libraries, an ‘extrema set’ (46,75) of 61,687 molecules was docked in the optimized system to ensure that the molecules with correct physical properties were enriched.

Overall, in the prospective screen, each library molecule was sampled in about 3438 orientations, on average about 187 conformations were sampled over five days on 1000 cores. The top-ranking 300,000 molecules were filtered for novelty using ECFP4-based Tanimoto coefficient ( $T_c < 0.35$ ) against known inhibitors of MTases. The remaining molecules were then clustered into related groups using an ECFP4-based  $T_c$  of 0.4. From the top 10,000 novel chemotypes, molecules with  $>2 \text{ kcal mol}^{-1}$  internal strain (38) were excluded and the remaining candidates were visually inspected for best docked poses with favorable interactions with the SARS-CoV2 active site. Ultimately, overall 165 molecules were selected for de novo synthesis and testing.

### **Non-covalent optimization**

Analogs for docking hits ‘2882’, ‘9213’, and ‘4824’ were queried in Arthor and SmallWorld 1.4 and 12 billion make-on-demand libraries (<http://sw.docking.org>, <http://arthor.docking.org>), the latter primarily containing Enamine REAL compounds (<http://enamine.net/compound-collections/real-compounds/real-space-navigator>). The resulting analogs were further filtered based on  $T_c > 0.4$  and docked to the N7-MTase domain of SARS-CoV2 nsp14. Compounds were also designed by

modifying 2D structure and custom synthesis by Enamine Ltd. (Kyiv, Ukraine). The docked poses were visually inspected for compatibility with the site and prioritized analogs were synthesized and tested for each series, respectively ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### **Fragment docking**

The optimized docking setup from the SARS-CoV-2 second non-covalent lead-like screen described above was used. Three different screens were run with different matching spheres (56) – those in the adenine-site, SAM-tail site, or all matching spheres (**Figure 3.4A**), with 15,738,235 docked and 14,406,946 scored, 15,738,278 docked and 14,124,978 scored, and 16,299,173 docked and 14,908,652 scored, respectively. Each setup was analyzed separately until visualization in Chimera (49) – the top 300,000 ranked poses were filtered for having torsional strain less than 7 REU total, and single strain of 2.5 REU (38), less than 2 stranded hydrogen bond donors, less than 4 stranded hydrogen bond acceptors, and greater than 1 hydrogen bond to Tyr368, Ala353, or Gly333 (48). Remaining molecules were visually inspected for having favorable interactions. In total, 68 compounds were selected for purchasing, 50 from Enamine and 18 from WuXi, and overall, 53 were successfully synthesized for a fulfilment rate of 82%.

### **Covalent database curation**

SMARTS patterns for aldehydes or acrylamides ([CX3H1](=O)[#6] and ([CD1]=[CD2]-C(=O)-[NX3]), respectively) were searched in Enamine REAL databases, finding 20 million acrylamides and 6 million aldehydes. The DOCKoValent 3D files were generated as previously described (58-60). Briefly, the electrophiles were converted to their transition state product and a dummy atom was placed indicating to the docking algorithm which atom should be modeled covalently bound



to the sulfur of the cysteine. Both 2D structures and 3D DOCK covalent files are now publicly available at <http://covalent2022.docking.org>. To compare to other public molecule databases, we used the ZINC20 in-stock set (51), the MLSMR library (62) and the UCSF SMDC library (61), and searched the same SMARTS patterns for acrylamides and aldehydes. The number of chemotypes were determined by Bemis-Murcko clustering (76).

### **Covalent docking and compound optimization**

The optimized docking setup from the first SARS-CoV-1 lead-like screen described above was used, with differences being which residues have been hyper-polarized (75) (Tyr368, Tyr368 and Ala353, or Tyr368, Ala353, and Gly333, referred to as 1-HP, 2-HP and 3-HP, respectively). For the acrylamide screen against 1-HP, molecules with docked scores less than 0 were selected for filtering (top 341,000); those with internal torsional strain less than total strain of 6.5 REU and single strain of 2 REU (38), molecules with less than 2 stranded hydrogen bond donors and less than 4 stranded hydrogen bond acceptors were prioritized. Molecules were also selected that formed at least one hydrogen bond to Tyr368, Ala353 or Gly333 using LUNA (48) leaving 2,423. After clustering for chemical similarity, 533 were visually inspected in Chimera (49). For the 2-HP setup, molecules with scores less than 0 (top 440,661) were filtered using the same criteria with 2,961 molecules remaining, comprising of 622 clusters that were visually inspected. For the 3-HP setup, no molecules passed the strain, IFP, and hydrogen bond filter and were not considered further. Visual inspection prioritized molecules with the same criterion as above. Lastly, selected compounds from both 1-HP and 2-HP setups were clustered to select unique chemotypes, and 31 were purchased. Synthesis was successful for 26 for a fulfillment rate of 84%.

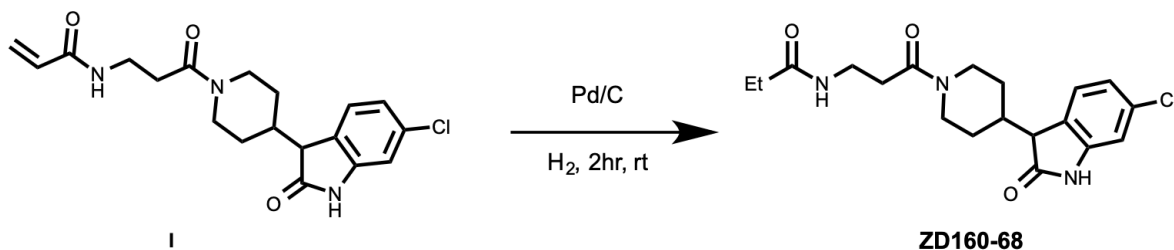
For the aldehydes in the 1-HP setup, the top 894,979 compounds (dock score less than 0) were filtered to prioritized as the acrylamides were above, with clustering for chemical similarity leaving 1,340 for visual filtering. For the 2-HP setup, the top 1,494,350 were filtered to 3,548, and 3-HP setup of top 1,494,345 to 3,548 for visual inspection. Compounds were prioritized for the same interactions as the acrylamides, and finally 61 aldehydes were selected. Synthesis was successful for 47 of these for a fulfilment rate of 77%, and an overall covalent fulfilment rate of 79%.

Acryl42 analogs acryl42\_5, acryl42\_11 and acryl42\_10 was designed off the acryl42 2D chemical structure, and synthesized by Enamine; ZD160-68 was designed to test the activity of the acrylamide warhead. The modeled pose of acryl42\_10 was performed in Maestro (version 2021-2, Schrödinger, Inc.) by manually changing the acryl42 docked pose to acryl42\_10 and minimizing the nsp14-acryl42\_10 complex using the Protein Preparation Wizard protocol.

### Make-on-demand synthesis

Compounds were purchased from Enamine. Purities of active molecules were at least 90% and typically above 95% ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### Synthesis of ZD160-68



### **N-(3-(4-(6-chloro-2-oxoindolin-3-yl)piperidin-1-yl)-3-oxopropyl)propionamide (ZD160-68)**

To a solution of compound **I** (5.6 mg, 0.015 mmol) in THF (2 mL), was added Pd/C (10%, 5 mg). The mixture was stirred under H<sub>2</sub> atmosphere for 2 hours followed by filtering. The filtrate was collected and purified by prep-HPLC to yield ZD160-68, white 1 solid (4 mg, 70% yield). <sup>1</sup>HNMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 7.32 – 7.21 (m, 1H), 7.03 (dt, *J* = 8.0, 2.2 Hz, 1H), 6.91 (d, *J* = 2.2 Hz, 1H), 4.61 (t, *J* = 15.8 Hz, 1H), 4.00 (t, *J* = 13.9 Hz, 1H), 3.50 (q, *J* = 2.6 Hz, 1H), 3.42 (q, *J* = 6.3, 5.8 Hz, 2H), 3.11 – 3.00 (m, 1H), 2.63 – 2.52 (m, 3H), 2.36 (d, *J* = 12.7 Hz, 1H), 2.24 – 2.10 (m, 2H), 1.73 – 1.59 (m, 2H), 1.49 (dq, *J* = 17.2, 5.7, 4.3 Hz, 2H), 1.14 – 1.08 (m, 3H). **MS (ESI) *m/z***: [M+H]<sup>+</sup> calcd for C<sub>19</sub>H<sub>24</sub>ClN<sub>3</sub>O<sub>3</sub> 378.8; found 378.3.

### **Nsp14 expression and purification**

For activity assays, nsp14 was expressed and purified as previously described (33). The codon-optimized gene was also sub-cloned in a modified pET28b with 6x-Histidine and SUMO tag at the N-terminus. Nsp14 was expressed in *E. coli* Rosetta2(DE3) PlysS cells, growing in terrific broth at 37°C, induced at 18°C with 0.4 mM IPTG at OD 600nm of 1.2 for 18 h. Cell pellets were recovered and stored at -80°C.

For purification, cells were suspended in lysis buffer containing 50 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% v/v glycerol, 5 mM MgCl<sub>2</sub>, 1 mM TCEP pH 8.1 supplemented with EDTA-free protease cocktail inhibitor tablets (Thermo Scientific). Cells were disrupted by sonication and lysate centrifuged at 16,000 rpm for 30 min. Nsp14 was purified using a 5 mL HisTrap HP column, washed with 20 column volumes of lysis buffer with additional 20 and 30 mM imidazole, and eluted with buffer containing 500 mM imidazole. Protein fractions were exchanged to 50mM

HEPES, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, and 1 mM TCEP pH 8.0, and incubated overnight at 4°C with SenP1 protease at a 1:100 mass ratio. SUMO tag was removed using a MonoQ 10/100 column, pre-equilibrated in 50 mM HEPES, 20 mM NaCl, 5 mM MgCl<sub>2</sub>, and 1 mM TCEP pH 8.0. Nsp14 was in the unbound fraction. As a final step, the protein was purified using a size exclusion column s200 16/600 in the same buffer for SenP1 digestion. Purest fractions were pulled together, flash frozen, and stored at -80°C until needed.

### **Enzyme Inhibition**

The inhibitory effect of compounds on the methyltransferase activity of SARS-CoV-2 nsp14 was assessed using a previously developed radiometric assay (33).

### **Jump dilution**

The recovery of nsp14 activity after incubation with each inhibitor and rapid dilution was monitored. Nsp14 at 100-fold higher concentration than what required for activity measurement (1.5 nM) was incubated with each compound at 10-fold of IC<sub>50</sub> value concentration for 1h at room temperature. Reaction mixtures were then rapidly diluted 100-fold into the assay buffer containing substrate RNA and SAM, and recovery of the nsp14 activity was monitored.

### **Mechanism Of Action**

IC<sub>50</sub> values were determined at a fixed concentration of RNA substrate (0.25 μM; 5xK<sub>m</sub>) and varying concentrations of SAM (up to 2.5 μM/10x K<sub>m</sub>), and varying concentrations of RNA (up to 0.5 μM; 10x K<sub>m</sub>) at fixed 1.25 μM (5xK<sub>m</sub>) of SAM. Linear increase in IC<sub>50</sub> values as the

concentration of substrate is increased, indicated a competitive pattern of inhibition as described by (77).

#### **Assessment of covalent binding by LC-MS.**

To form the protein–ligand (‘1911) complex, nsp14 was incubated with 20 molar excess of compound for 2hr. at room temperature (20°C) before adding MS running buffer (0.1% FA). The resulting samples were separated on a HPLC column with 5-95% acetonitrile in water as eluent. The MS data were analyzed using an Agilent LC/MSD Time-of-Flight Mass Spectrometer equipped with an electrospray ionization source.

For compounds acryl42 and acryl42-10, an aliquot of pure nsp14 enzyme was thawed on ice, centrifuged for 10 min at 14,000 rpm, diafiltrated, and concentrated to 40  $\mu$ M in 50 mM HEPES pH 8.0, 20 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM TCEP. Then, 1  $\mu$ M nsp14 was incubated alone or in the presence of 500  $\mu$ M acryl42, or 50  $\mu$ M acryl42-10 at room temperature in 100  $\mu$ L aliquots. Time points were taken at 0, 2, 4, 8, and 18 h. For each time point, 1 $\mu$ L of the mix was injected into a Xevo G2-XS QToF Quadrupole Time of Flight mass spectrometer (Waters) using a solution of 0.05% formic acid at room temperature. Collected spectra from 700 to 1400 m/z were used to determine protein mass using MaxEnt with a 58500 to 62000 Da range and 1 Da/channel resolution.

## **Compound Selectivity**

Selectivity assays were performed as previously described (78). Compounds were tested at 10  $\mu$ M in triplicate using radiometric assays. Enzymes with >50% inhibition were prioritized for concentration-response curves for IC<sub>50</sub> determination.

## **Antiviral and cytotoxicity assays**

For compounds '1911 and '1988, two thousand (2,000) HeLa-ACE2 cells were seeded into 96-well plates and incubated for 24 h. Two hours before infection, the media was replaced with a new media containing the compound of interest, including a DMSO control. Plates were then transferred into the biosafety level 3 (BSL-3) facility and 1,000 PFU (MOI = 0.25) of SARS-CoV-2 was added, bringing the final compound concentration to those indicated. SARS-CoV-2/WA1 variant was used as indicated. Plates were then incubated for 48 h. Infectivity was measured by the accumulation of viral NP protein in the nucleus of the HeLa-ACE2 cells (fluorescence accumulation). Percent infection was quantified as  $((\text{Infected cells}/\text{Total cells}) - \text{Background}) \times 100$ , and the DMSO control was then set to 100% infection for analysis. Cytotoxicity was also performed at matched concentrations using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected HeLa-ACE2 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

Cell viability for compounds '9163 ('9213 analog), '1631 ('4824 analog), acryl42-10 (acryl42 analog), acryl42, '4975, '0683, '1889, '1988, '1911, and sinefungin (Sigma-Aldrich #S8559) was measured using Cell Titer-Glo® Luminescent cell viability assay's standard protocol (Promega,

Madison, WI, USA). Briefly,  $2 \times 10^4$  A549-ACE2 cells/well were seeded in 96-well plates and incubated at 37°C and 5 % CO<sub>2</sub>. The next day compounds were added to the cells at the corresponding concentrations and further incubated for 48 h. Prior luciferase measurement, cells were equilibrated for 30 min at room temperature before addition of a 1:1 ratio of CellTiter-Glo® Reagent and cell culture media. Mixture was incubated for 10 min in the dark while constant shaking to allow cell lysis. Luciferase measurement was performed in a Tecan Infinite® 200 PRO (Tecan Group Ltd., Männedorf, Switzerland) with a 0.5 sec integration time.

### **Statistical analyses.**

Data was analyzed using Prism 8.0 (GraphPad, San Diego, CA). For dose response curves, data was fitted to the four-parameter logistic equation.

### **Data availability**

The identities of compounds docked in this study are freely available from ZINC15, ZINC20, ZINC22, and covalent2002 databases (<http://zinc15.docking.org>, <http://zinc20.docking.org>, <http://files.docking.org/zinc22>, <http://covalent2022.docking.org>). Active compounds may be purchased from Enamine or are available from the authors. All other data are available from the corresponding authors on request.

### **Code availability**

DOCK3.7 is freely available for non-commercial research (<http://dock.compbio.ucsf.edu/DOCK3.7/>), as is DOCKcovalent. Web-based versions are freely

usable by all (<http://blaster.docking.org/>). The ultra-large libraries used here are freely available (<http://zinc15.docking.org>, <http://zinc20.docking.org>, <http://files.docking.org/zinc22>).



### 3.8 Acknowledgements

We thank OpenEye Software for Omega and additional tools, and Schrodinger LLC for the Maestro package. We thank Prof. Kevan Shokat and Dr. Qinheng Zheng for access to the Xevo G2-XS QToF Quadrupole Time of Flight mass spectrometer. We would like to also thank Structural Genomics Consortium (SGC) for their support. SGC is a registered charity (no: 1097737) that receives funds from Bayer AG, Boehringer Ingelheim, Bristol Myers Squibb, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking [EUbOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer and Takeda. This work was supported by the US NIH R35GM122481 and DARPA HR0011-19-2-0020 (B.K.S.), NIH GM071896 and GM133836 (J.J.I.), and NIH U19AI171110 (D.G.F, B.K.S, M.V., J.J., M.O.). This work utilized the NMR Spectrometer Systems at Mount Sinai acquired with funding from National Institutes of Health SIG grants 1S10OD025132 and 1S10OD028504. M.O. received support from NIH U19AI171110 and the James B. Pendleton Charitable Trust.

### 3.9 Author Contributions

I.S. and E.A.F. performed docking screens, with input from B.K.S. Ligand optimization was conducted by I.S. and E.A.F. with input from B.K.S. E.A.F. and J.J.I. created the covalent libraries with assistance from X.W. F.L. performed nsp14 enzymatic assays, Jump dilution and MOA experiments, and '1911 LCMS. IC performed selectivity assays. A.L, and K.D. contributed to enzymatic assays. M.V., designed experiments, reviewed data and supervised. Additional LCMS testing was performed by A.R.-H. with supervision by D.G.F. M.R. and K.W. tested compounds in HeLa-ACE2 antiviral and cytotoxicity studies with supervision by A.G.-S. Cell viability in A549-ACE2 cells was performed by F.Z.B., supervised by M.O. J.G. assisted in experimental design. Y.S.M. supervised Enamine compound synthesis purchased, assisted by N.A.T. J.J.I. built the ZINC15, ZINC22 ultra-large libraries. Synthesis of ZD160-68 was done by Z.D. with input and supervision by H.U.K. and J.J. B.K.S. and M.V. supervised the project with chemoinformatic input from J.J.I. I.S. and E.A.F. wrote the paper with input from the other authors and primary editing from M.V. and B.K.S. M.V. and B.K.S. conceived the project.

### 3.10 References

1. Owen, D. R., Allerton, C. M. N., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., Boras, B., Cardin, R. D., Carlo, A., Coffman, K. J., Dantonio, A., Di, L., Eng, H., Ferre, R., Gajiwala, K. S., Gibson, S. A., Greasley, S. E., Hurst, B. L., Kadar, E. P., Kalgutkar, A. S., Lee, J. C., Lee, J., Liu, W., Mason, S. W., Noell, S., Novak, J. J., Obach, R. S., Ogilvie, K., Patel, N. C., Pettersson, M., Rai, D. K., Reese, M. R., Sammons, M. F., Sathish, J. G., Singh, R. S. P., Steppan, C. M., Stewart, A. E., Tuttle, J. B., Updyke, L., Verhoest, P. R., Wei, L., Yang, Q., and Zhu, Y. (2021) An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **374**, 1586-1593
2. Sheahan, T. P., Sims, A. C., Zhou, S., Graham, R. L., Pruijssers, A. J., Agostini, M. L., Leist, S. R., Schäfer, A., Dinnon, K. H., 3rd, Stevens, L. J., Chappell, J. D., Lu, X., Hughes, T. M., George, A. S., Hill, C. S., Montgomery, S. A., Brown, A. J., Bluemling, G. R., Natchus, M. G., Saindane, M., Kolykhalov, A. A., Painter, G., Harcourt, J., Tamin, A., Thornburg, N. J., Swanstrom, R., Denison, M. R., and Baric, R. S. (2020) An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci Transl Med* **12**
3. Wahl, A., Gralinski, L. E., Johnson, C. E., Yao, W., Kovarova, M., Dinnon, K. H., Liu, H., Madden, V. J., Krzystek, H. M., De, C., White, K. K., Gully, K., Schäfer, A., Zaman, T., Leist, S. R., Grant, P. O., Bluemling, G. R., Kolykhalov, A. A., Natchus, M. G., Askin, F. B., Painter, G., Browne, E. P., Jones, C. D., Pickles, R. J., Baric, R. S., and Garcia, J. V. (2021) SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature* **591**, 451-457

4. Siegel, D., Hui, H. C., Doerffler, E., Clarke, M. O., Chun, K., Zhang, L., Neville, S., Carra, E., Lew, W., Ross, B., Wang, Q., Wolfe, L., Jordan, R., Soloveva, V., Knox, J., Perry, J., Perron, M., Stray, K. M., Barauskas, O., Feng, J. Y., Xu, Y., Lee, G., Rheingold, A. L., Ray, A. S., Bannister, R., Strickley, R., Swaminathan, S., Lee, W. A., Bavari, S., Cihlar, T., Lo, M. K., Warren, T. K., and Mackman, R. L. (2017) Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo[2,1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. *J Med Chem* **60**, 1648-1661
5. Gordon, C. J., Tchesnokov, E. P., Woolner, E., Perry, J. K., Feng, J. Y., Porter, D. P., and Götte, M. (2020) Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* **295**, 6785-6797
6. Rehman, S., Ashfaq, U. A., and Javed, T. (2011) Antiviral drugs against hepatitis C virus. *Genet Vaccines Ther* **9**, 11
7. De Clercq, E. (2007) The design of drugs for HIV and HCV. *Nat Rev Drug Discov* **6**, 1001-1018
8. De Clercq, E. (2009) The history of antiretrovirals: key discoveries over the past 25 years. *Rev Med Virol* **19**, 287-299
9. De Clercq, E., and Li, G. (2016) Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **29**, 695-747
10. Luttens, A., Gullberg, H., Abdurakhmanov, E., Vo, D. D., Akaberi, D., Talibov, V. O., Nekhotiaeva, N., Vangeel, L., De Jonghe, S., Jochmans, D., Krambrich, J., Tas, A., Lundgren, B., Gravenfors, Y., Craig, A. J., Atilaw, Y., Sandström, A., Moodie, L. W. K., Lundkvist, Å., van Hemert, M. J., Neyts, J., Lennerstrand, J., Kihlberg, J., Sandberg, K.,

- Danielson, U. H., and Carlsson, J. (2022) Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc* **144**, 2905-2920
11. Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y., and Wang, J. (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* **30**, 678-692
12. Zhang, C. H., Spasov, K. A., Reilly, R. A., Hollander, K., Stone, E. A., Ippolito, J. A., Liosi, M. E., Deshmukh, M. G., Tirado-Rives, J., Zhang, S., Liang, Z., Miller, S. J., Isaacs, F., Lindenbach, B. D., Anderson, K. S., and Jorgensen, W. L. (2021) Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med Chem Lett* **12**, 1325-1332
13. Unoh, Y., Uehara, S., Nakahara, K., Nobori, H., Yamatsu, Y., Yamamoto, S., Maruyama, Y., Taoda, Y., Kasamatsu, K., Suto, T., Kouki, K., Nakahashi, A., Kawashima, S., Sanaki, T., Toba, S., Uemura, K., Mizutare, T., Ando, S., Sasaki, M., Orba, Y., Sawa, H., Sato, A., Sato, T., Kato, T., and Tachibana, Y. (2022) Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J Med Chem* **65**, 6499-6512
14. Schuller, M., Correy, G. J., Gahbauer, S., Fearon, D., Wu, T., Díaz, R. E., Young, I. D., Carvalho Martins, L., Smith, D. H., Schulze-Gahmen, U., Owens, T. W., Deshpande, I., Merz, G. E., Thwin, A. C., Biel, J. T., Peters, J. K., Moritz, M., Herrera, N., Kratochvil, H. T., Aimon, A., Bennett, J. M., Brandao Neto, J., Cohen, A. E., Dias, A., Douangamath, A., Dunnett, L., Fedorov, O., Ferla, M. P., Fuchs, M. R., Gorrie-Stone, T. J., Holton, J. M.,

- Johnson, M. G., Krojer, T., Meigs, G., Powell, A. J., Rack, J. G. M., Rangel, V. L., Russi, S., Skyner, R. E., Smith, C. A., Soares, A. S., Wierman, J. L., Zhu, K., O'Brien, P., Jura, N., Ashworth, A., Irwin, J. J., Thompson, M. C., Gestwicki, J. E., von Delft, F., Shoichet, B. K., Fraser, J. S., and Ahel, I. (2021) Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci Adv* **7**
15. Gahbauer, S., Correy, G. J., Schuller, M., Ferla, M. P., Umay Doruk, Y., Rachman, M., Wu, T., Diolaiti, M., Wang, S., Neitz, R. J., Fearon, D., Radchenko, D., Moroz, Y., Irwin, J. J., Renslo, A. R., Taylor, J. C., Gestwicki, J. E., von Delft, F., Ashworth, A., Ahel, I., Shoichet, B. K., and Fraser, J. s. (2022) Structure-based inhibitor optimization for the Nsp3 Macrodomain of SARS-CoV-2. *bioRxiv*
16. Shan, H., Liu, J., Shen, J., Dai, J., Xu, G., Lu, K., Han, C., Wang, Y., Xu, X., Tong, Y., Xiang, H., Ai, Z., Zhuang, G., Hu, J., Zhang, Z., Li, Y., Pan, L., and Tan, L. (2021) Development of potent and selective inhibitors targeting the papain-like protease of SARS-CoV-2. *Cell Chem Biol* **28**, 855-865.e859
17. Klima, M., Khalili Yazdi, A., Li, F., Chau, I., Hajian, T., Bolotokova, A., Kaniskan, H., Han, Y., Wang, K., Li, D., Luo, M., Jin, J., Boura, E., and Vedadi, M. (2022) Crystal structure of SARS-CoV-2 nsp10-nsp16 in complex with small molecule inhibitors, SS148 and WZ16. *Protein Sci* **31**, e4395
18. Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., and Guo, D. (2009) Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A* **106**, 3484-3489

19. Jin, Y., Ouyang, M., Yu, T., Zhuang, J., Wang, W., Liu, X., Duan, F., Guo, D., Peng, X., and Pan, J. A. (2022) Genome-Wide Analysis of the Indispensable Role of Non-structural Proteins in the Replication of SARS-CoV-2. *Front Microbiol* **13**, 907422
20. Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R. L., Baric, R. S., Stockwell, T. B., Spiro, D. J., and Denison, M. R. (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog* **6**, e1000896
21. Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., and Decroly, E. (2012) RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci U S A* **109**, 9372-9377
22. Ogando, N. S., Ferron, F., Decroly, E., Canard, B., Posthuma, C. C., and Snijder, E. J. (2019) The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. *Front Microbiol* **10**, 1813
23. Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., and Rao, Z. (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **112**, 9436-9441
24. Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., and Decroly, E. (2010) In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* **6**, e1000863
25. Marcotrigiano, J., Gingras, A. C., Sonenberg, N., and Burley, S. K. (1997) Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951-961

26. Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2011) Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol* **10**, 51-65
27. Talukdar, A., Mukherjee, A., and Bhattacharya, D. (2022) Fascinating Transformation of SAM-Competitive Protein Methyltransferase Inhibitors from Nucleoside Analogues to Non-Nucleoside Analogues. *J Med Chem* **65**, 1662-1684
28. Spannhoff, A., Sippl, W., and Jung, M. (2009) Cancer treatment of the future: inhibitors of histone methyltransferases. *Int J Biochem Cell Biol* **41**, 4-11
29. Ham, M. S., Lee, J. K., and Kim, K. C. (2013) S-adenosyl methionine specifically protects the anticancer effect of 5-FU via DNMTs expression in human A549 lung cancer cells. *Mol Clin Oncol* **1**, 373-378
30. Ferreira de Freitas, R., Ivanochko, D., and Schapira, M. (2019) Methyltransferase Inhibitors: Competing with, or Exploiting the Bound Cofactor. *Molecules* **24**
31. Liu, C., Shi, W., Becker, S. T., Schatz, D. G., Liu, B., and Yang, Y. (2021) Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science* **373**, 1142-1146
32. Li, A. S. M., Li, F., Eram, M. S., Bolotokova, A., Dela Seña, C. C., and Vedadi, M. (2020) Chemical probes for protein arginine methyltransferases. *Methods* **175**, 30-43
33. Devkota, K., Schapira, M., Perveen, S., Khalili Yazdi, A., Li, F., Chau, I., Ghiabi, P., Hajian, T., Loppnau, P., Bolotokova, A., Satchell, K. J. F., Wang, K., Li, D., Liu, J., Smil, D., Luo, M., Jin, J., Fish, P. V., Brown, P. J., and Vedadi, M. (2021) Probing the SAM Binding Site of SARS-CoV-2 Nsp14 In Vitro Using SAM Competitive Inhibitors Guides Developing Selective Bisubstrate Inhibitors. *SLAS Discov* **26**, 1200-1211
34. Meng, E. C., Gschwend, D. A., Blaney, J. M., and Kuntz, I. D. (1993) Orientational sampling and rigid-body minimization in molecular docking. *Proteins* **17**, 266-278



35. Mysinger, M. M., and Shoichet, B. K. (2010) Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573
36. Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W., and Shoichet, B. K. (2002) A model binding site for testing scoring functions in molecular docking. *J Mol Biol* **322**, 339-355
37. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J., and Shoichet, B. K. (2013) Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992
38. Gu, S., Smith, M. S., Yang, Y., Irwin, J. J., and Shoichet, B. K. (2021) Ligand Strain Energy in Large Library Docking. *J Chem Inf Model* **61**, 4331-4341
39. Oprea, T. I. (2002) Chemical space navigation in lead discovery. *Curr Opin Chem Biol* **6**, 384-389
40. Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., Resnick, E., Strain-Damerell, C., Aimon, A., Ábrányi-Balogh, P., Brandão-Neto, J., Carbery, A., Davison, G., Dias, A., Downes, T. D., Dunnett, L., Fairhead, M., Firth, J. D., Jones, S. P., Keeley, A., Keserü, G. M., Klein, H. F., Martin, M. P., Noble, M. E. M., O'Brien, P., Powell, A., Reddi, R. N., Skyner, R., Snee, M., Waring, M. J., Wild, C., London, N., von Delft, F., and Walsh, M. A. (2020) Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat Commun* **11**, 5047
41. Bonday, Z. Q., Cortez, G. S., Grogan, M. J., Antonysamy, S., Weichert, K., Bocchinfuso, W. P., Li, F., Kennedy, S., Li, B., Mader, M. M., Arrowsmith, C. H., Brown, P. J., Eram, M. S., Szewczyk, M. M., Barsyte-Lovejoy, D., Vedadi, M., Guccione, E., and Campbell, R. M. (2018) LLY-283, a Potent and Selective Inhibitor of Arginine Methyltransferase 5, PRMT5, with Antitumor Activity. *ACS Med Chem Lett* **9**, 612-617

42. Huynh, T., Chen, Z., Pang, S., Geng, J., Bandiera, T., Bindi, S., Vianello, P., Roletto, F., Thieffine, S., Galvani, A., Vaccaro, W., Poss, M. A., Trainor, G. L., Lorenzi, M. V., Gottardis, M., Jayaraman, L., and Purandare, A. V. (2009) Optimization of pyrazole inhibitors of Coactivator Associated Arginine Methyltransferase 1 (CARM1). *Bioorg Med Chem Lett* **19**, 2924-2927
43. Daigle, S. R., Olhava, E. J., Therkelsen, C. A., Majer, C. R., Sneeringer, C. J., Song, J., Johnston, L. D., Scott, M. P., Smith, J. J., Xiao, Y., Jin, L., Kuntz, K. W., Chesworth, R., Moyer, M. P., Bernt, K. M., Tseng, J. C., Kung, A. L., Armstrong, S. A., Copeland, R. A., Richon, V. M., and Pollock, R. M. (2011) Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* **20**, 53-65
44. Chory, E. J., Calarco, J. P., Hathaway, N. A., Bell, O., Neel, D. S., and Crabtree, G. R. (2019) Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome. *Mol Cell* **73**, 61-72.e63
45. Bonham, K., Hemmers, S., Lim, Y. H., Hill, D. M., Finn, M. G., and Mowen, K. A. (2010) Effects of a novel arginine methyltransferase inhibitor on T-helper cell cytokine production. *FEBS J* **277**, 2096-2108
46. Stein, R. M., Yang, Y., Balias, T. E., O'Meara, M. J., Lyu, J., Young, J., Tang, K., Shoichet, B. K., and Irwin, J. J. (2021) Property-Unmatched Decoys in Docking Benchmarks. *J Chem Inf Model* **61**, 699-714
47. Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582-6594

48. Fassio, A. V., Shub, L., Ponzoni, L., McKinley, J., O'Meara, M. J., Ferreira, R. S., Keiser, M. J., and de Melo Minardi, R. C. (2022) Prioritizing Virtual Screening with Interpretable Interaction Fingerprints. *J Chem Inf Model*
49. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612
50. Tingle, B., Tang, K., Castanon, J., Gutierrez, J., Khurelbaatar, M., Dandarchuluun, C., Moroz, Y., and Irwin, J. (2022) ZINC-22 - A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *ChemRxiv*
51. Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., and Sayle, R. A. (2020) ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J Chem Inf Model* **60**, 6065-6073
52. Chen, Y., and Shoichet, B. K. (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol* **5**, 358-364
53. Barelier, S., Eidam, O., Fish, I., Hollander, J., Figaroa, F., Nachane, R., Irwin, J. J., Shoichet, B. K., and Siegal, G. (2014) Increasing chemical space coverage by combining empirical and computational fragment screens. *ACS Chem Biol* **9**, 1528-1535
54. Schuffenhauer, A., Ruedisser, S., Marzinzik, A. L., Jahnke, W., Blommers, M., Selzer, P., and Jacoby, E. (2005) Library design for fragment based screening. *Curr Top Med Chem* **5**, 751-762
55. Teotico, D. G., Babaoglu, K., Rocklin, G. J., Ferreira, R. S., Giannetti, A. M., and Shoichet, B. K. (2009) Docking for fragment inhibitors of AmpC beta-lactamase. *Proc Natl Acad Sci U S A* **106**, 7455-7460

56. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161**, 269-288
57. Zautra, A. J., Maxwell, B. M., and Reich, J. W. (1989) Relationship among physical impairment, distress, and well-being in older adults. *J Behav Med* **12**, 543-557
58. Wan, X., Yang, T., Cuesta, A., Pang, X., Balius, T. E., Irwin, J. J., Shoichet, B. K., and Taunton, J. (2020) Discovery of Lysine-Targeted eIF4E Inhibitors through Covalent Docking. *J Am Chem Soc* **142**, 4960-4964
59. Fink, E. A., Bardine, C., Gahbauer, S., Singh, I., White, K., Gu, S., Wan, X., Ary, B., Glenn, I., O'Connell, J., O'Donnell, H., Fajtova, P., Lyu, J., Vigneron, S., Young, N. J., Kondratov, I. S., O'Donoghue, A. J., Moroz, Y., Taunton, J., Renslo, A. R., Irwin, J. J., Garcia-Sastre, A., Shoichet, B. K., and Craik, C. S. (2022) Large library docking for novel SARS-CoV-2 main protease non-covalent and covalent inhibitors. bioRxiv
60. London, N., Miller, R. M., Krishnan, S., Uchida, K., Irwin, J. J., Eidam, O., Gibold, L., Cimermančič, P., Bonnet, R., Shoichet, B. K., and Taunton, J. (2014) Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol* **10**, 1066-1072
61. Arkin, M. R., Ang, K. K., Chen, S., Davies, J., Merron, C., Tang, Y., Wilson, C. G., and Renslo, A. R. (2014) UCSF Small Molecule Discovery Center: innovation, collaboration and chemical biology in the Bay Area. *Comb Chem High Throughput Screen* **17**, 333-342
62. PubChem Molecular Libraries Small Molecule Repository .
63. Boike, L., Henning, N. J., and Nomura, D. K. (2022) Advances in covalent drug discovery. *Nat Rev Drug Discov*
64. Bouhaddou, M., Memon, D., Meyer, B., White, K. M., Rezelj, V. V., Correa Marrero, M., Polacco, B. J., Melnyk, J. E., Ulferts, S., Kaake, R. M., Batra, J., Richards, A. L., Stevenson,

- E., Gordon, D. E., Rojc, A., Obernier, K., Fabius, J. M., Soucheray, M., Miorin, L., Moreno, E., Koh, C., Tran, Q. D., Hardy, A., Robinot, R., Vallet, T., Nilsson-Payant, B. E., Hernandez-Armenta, C., Dunham, A., Weigang, S., Knerr, J., Modak, M., Quintero, D., Zhou, Y., Dugourd, A., Valdeolivas, A., Patil, T., Li, Q., Hüttenhain, R., Cakir, M., Muralidharan, M., Kim, M., Jang, G., Tutuncuoglu, B., Hiatt, J., Guo, J. Z., Xu, J., Bouhaddou, S., Mathy, C. J. P., Gaulton, A., Manners, E. J., Félix, E., Shi, Y., Goff, M., Lim, J. K., McBride, T., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., De Wit, E., Leach, A. R., Kortemme, T., Shoichet, B., Ott, M., Saez-Rodriguez, J., tenOever, B. R., Mullins, R. D., Fischer, E. R., Kochs, G., Grosse, R., García-Sastre, A., Vignuzzi, M., Johnson, J. R., Shokat, K. M., Swaney, D. L., Beltrao, P., and Krogan, N. J. (2020) The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182**, 685-712.e619
65. Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O'Meara, M. J., Monel, B., Vallet, T., White, K. M., Zhang, Z., Alon, A., Schadt, H., O'Donnell, H. R., Lyu, J., Rosales, R., McGovern, B. L., Rathnasinghe, R., Jangra, S., Schotsaert, M., Galarneau, J. R., Krogan, N. J., Urban, L., Shokat, K. M., Kruse, A. C., García-Sastre, A., Schwartz, O., Moretti, F., Vignuzzi, M., Pognan, F., and Shoichet, B. K. (2021) Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science* **373**, 541-547
66. Ahmed-Belkacem, R., Hausdorff, M., Delpal, A., Sutto-Ortiz, P., Colmant, A. M. G., Touret, F., Ogando, N. S., Snijder, E. J., Canard, B., Coutard, B., Vasseur, J. J., Decroly, E., and Debart, F. (2022) Potent Inhibition of SARS-CoV-2 nsp14. *J Med Chem* **65**, 6231-6249
67. Otava, T., Šála, M., Li, F., Fanfrlík, J., Devkota, K., Perveen, S., Chau, I., Pakarian, P., Hobza, P., Vedadi, M., Boura, E., and Nencka, R. (2021) The Structure-Based Design of

SARS-CoV-2 nsp14 Methyltransferase Ligands Yields Nanomolar Inhibitors. *ACS Infect Dis* **7**, 2214-2220

68. Imprachim, N., Yosaatmadja, Y., and Newman, J. A. (2022) Crystal structures and fragment screening of SARS-CoV-2 NSP14 reveal details of exoribonuclease activation and mRNA capping and provide starting points for antiviral drug development. *bioRxiv*
69. Owen, D. R., Allerton, C. M. N., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., Boras, B., Cardin, R. D., Carlo, A., Coffman, K. J., Dantonio, A., Di, L., Eng, H., Ferre, R., Gajiwala, K. S., Gibson, S. A., Greasley, S. E., Hurst, B. L., Kadar, E. P., Kalgutkar, A. S., Lee, J. C., Lee, J., Liu, W., Mason, S. W., Noell, S., Novak, J. J., Obach, R. S., Ogilvie, K., Patel, N. C., Pettersson, M., Rai, D. K., Reese, M. R., Sammons, M. F., Sathish, J. G., Singh, R. S. P., Stepan, C. M., Stewart, A. E., Tuttle, J. B., Updyke, L., Verhoest, P. R., Wei, L., Yang, Q., and Zhu, Y. (2021) An oral SARS-CoV-2 M. *Science* **374**, 1586-1593
70. Basu, S., Mak, T., Ulferts, R., Wu, M., Deegan, T., Fujisawa, R., Tan, K. W., Lim, C. T., Basier, C., Canal, B., Curran, J. F., Drury, L. S., McClure, A. W., Roberts, E. L., Weissmann, F., Zeisner, T. U., Beale, R., Cowling, V. H., Howell, M., Labib, K., and Diffley, J. F. X. (2021) Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp14 RNA cap methyltransferase. *Biochem J* **478**, 2481-2497
71. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735-1747
72. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., DeBolt, S., Ferguson, D., Seibel, G., and Coleman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free

energy calculations to simulate the structural and energetic properties of molecules.

*Computer Physics Communications* **91**, 1-41

73. Gallagher, K., and Sharp, K. (1998) Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys J* **75**, 769-776
74. Meng, E. C., Shoichet, B. K., and Kuntz, I. D. (1992) Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry* **13**, 505-524
75. Bender, B. J., Gahbauer, S., Lutgens, A., Lyu, J., Webb, C. M., Stein, R. M., Fink, E. A., Balias, T. E., Carlsson, J., Irwin, J. J., and Shoichet, B. K. (2021) A practical guide to large-scale docking. *Nat Protoc* **16**, 4799-4832
76. Bemis, G. W., and Murcko, M. A. (1996) The properties of known drugs. 1. Molecular frameworks. *J Med Chem* **39**, 2887-2893
77. Copeland, R. A. (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, Wiley-VCH, Inc.
78. Scheer, S., Ackloo, S., Medina, T. S., Schapira, M., Li, F., Ward, J. A., Lewis, A. M., Northrop, J. P., Richardson, P. L., Kaniskan, H., Shen, Y., Liu, J., Smil, D., McLeod, D., Zepeda-Velazquez, C. A., Luo, M., Jin, J., Barsyte-Lovejoy, D., Huber, K. V. M., De Carvalho, D. D., Vedadi, M., Zaph, C., Brown, P. J., and Arrowsmith, C. H. (2019) A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nat Commun* **10**, 19

## Gloss to Chapter 4

In this chapter, I demonstrate the usefulness of multiple docking strategies and chemical subsets in a pandemic drug discovery project; this time we aimed to interfere with the SARS-CoV-2 main protease ( $M^{\text{Pro}}$ ). This project began before Chapter 3's Nsp14, and actually even before the pandemic hit the U.S. As we watched cases increase and spread across Asia and Europe, scientists published the genome of Covid-19/SARS-CoV-2 and quickly the first protein structures were determined, including  $M^{\text{Pro}}$ . Thankfully  $M^{\text{Pro}}$  exists in other coronaviruses and is decently well studied for its mechanism and inhibitor discovery. Larger pharmaceutical companies could pull from vaults of past projects, and often previous coronaviral work was too focused on  $M^{\text{Pro}}$  inhibitors. Early on, everyone jumped on the same idea to develop SARS-CoV-2  $M^{\text{Pro}}$  inhibitors.

It was helpful to have many groups working at once. We learned a lot from broader strategies in the 'drug discovery' toolbox, such as experimental crystallography screens with different types of inhibitors, and the massive structure-activity-relationship effort developing in real time by the open-source Moonshot Consortium. Our contributions included a goal we have for almost every project: discover multiple new chemical scaffolds that work against our protein of interest. Here we discovered over 35 new chemotypes that can inhibit  $M^{\text{Pro}}$  through non-covalent or covalent mechanisms. Some of these chemical scaffolds have been identified by other groups, but others have not, especially as we continued to seek novelty in a second docking screen 1.5 years after the first. We also show with the crystallography expert, Isha, that the computationally predicted protein-ligand interactions are recapitulated experimentally. Overall, this work showcases a pipeline that may serve useful if we have another pandemic driven drug discovery situation.



## Chapter 4

### Large library docking for novel SARS-CoV-2 main protease non-covalent inhibitors

#### Contributing Authors

Elissa A. Fink<sup>1,2,†</sup>, Conner Bardine<sup>1,3,†</sup>, Stefan Gahbauer<sup>1,†</sup>, Isha Singh<sup>1,†</sup>, Kris White<sup>4,5</sup>, Shuo Gu<sup>1</sup>, Xiaobo Wan<sup>1</sup>, Beatrice Ary<sup>1</sup>, Isabella Glenn<sup>1</sup>, Joseph O’Connell<sup>1</sup>, Henry O’Donnell<sup>1</sup>, Pavla Fajtová<sup>6</sup>, Jiankun Lyu<sup>1</sup>, Seth Vigneron<sup>1</sup>, Nicholas J. Young<sup>1</sup>, Ivan S. Kondratov<sup>7,8</sup>, Anthony J. O’Donoghue<sup>6</sup>, Yurii Moroz<sup>9,10</sup>, Jack Taunton<sup>11</sup>, Adam R. Renslo<sup>1</sup>, John J. Irwin<sup>1</sup>, Adolfo García-Sastre<sup>4,5,12,13,14,15</sup>, Brian K. Shoichet<sup>1,15,\*</sup>, Charles S. Craik<sup>1,15,\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, CA, USA

<sup>2</sup>Graduate Program in Biophysics, University of California-San Francisco, San Francisco, CA, USA

<sup>3</sup>Graduate Program in Chemistry and Chemical Biology, University of California-San Francisco, San Francisco, CA, USA

<sup>4</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>5</sup>Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>6</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California-San Diego, San Diego, CA, USA

<sup>7</sup>Enamine Ltd. Kyïv, Ukraine

<sup>8</sup>V.P. Kukhar Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, Kyïv, Ukraine

<sup>9</sup>National Taras Shevchenko University of Kyïv, Kyïv, Ukraine

<sup>10</sup>Chemspace LLC, Kyïv, Ukraine

<sup>11</sup>Department of Cellular and Molecular Pharmacology, University of California-San Francisco, San Francisco, CA, USA

<sup>12</sup>Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>13</sup>The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>14</sup>Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>15</sup>QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA

† These authors contributed equally.

\* Corresponding author. Email: [bshoichet@gmail.com](mailto:bshoichet@gmail.com) (B.K.S.), [charles.craik@ucsf.edu](mailto:charles.craik@ucsf.edu) (C.S.C.)

## 4.1 Abstract

Antiviral therapeutics to treat SARS-CoV-2 are much desired for the on-going pandemic. A well-precedented viral enzyme is the main protease ( $M^{\text{Pro}}$ ), now targeted by an approved drug and by several investigational drugs. Facing viral resistance, there remains a call for new inhibitor chemotypes; the best strategy to find them remains much debated. Adopting a structure-based approach, we docked 1.2 billion non-covalent lead-like molecules and a new library of 6.5 million electrophiles against the enzyme structure. From these, 29 non-covalent and 11 covalent inhibitors were identified in 37 series, the most potent having an  $IC_{50}$  of 29  $\mu\text{M}$  and 20  $\mu\text{M}$ , respectively. Several series were optimized, resulting in low micromolar inhibitors. Subsequent crystallography confirmed the docking predicted binding modes and may template further optimization. While the new chemotypes may be interesting in themselves, and may aid further optimization of  $M^{\text{Pro}}$  inhibitors for SARS-CoV-2, the modest success rate also reveals weaknesses in our approach for challenging targets like  $M^{\text{Pro}}$  versus others where it has been more successful, and versus other structure-based techniques against  $M^{\text{Pro}}$  itself. These will be considered.

## 4.2 Introduction

SARS-CoV-2 encodes two cysteine proteases that have essential roles in hydrolyzing viral polyproteins into nonstructural proteins, enabling virus replication. The main protease ( $M^{\text{Pro}}$ , also known as 3CL protease) cleaves 11 different sites in viral polyproteins (1,2). While  $M^{\text{Pro}}$  is highly conserved across other coronaviruses such as SARS-CoV-1 and MERS, it has no close human homolog (3-5). This makes it attractive for potential pan-coronavirus targeting, and for selective action.

The therapeutic potential of  $M^{\text{Pro}}$  inhibitors was substantiated by the approval of Paxlovid in December 2021. The treatment combines nirmatrelvir, which covalently inhibits  $M^{\text{Pro}}$ , with ritonavir, which slows nirmatrelvir's metabolism (6). Nirmatrelvir was optimized from PF-00835231, an inhibitor of the SARS-CoV-1  $M^{\text{Pro}}$  developed in response to the 2002 SARS outbreak (7). Meanwhile, other potent  $M^{\text{Pro}}$  inhibitors are advancing through the drug development pipeline. Among them is the orally active  $M^{\text{Pro}}$  inhibitor S-217622 (8), which has entered clinical trials. Other inhibitors show much promise (4,9-17), including a non-covalent  $M^{\text{Pro}}$  inhibitor from the international Covid-19 Moonshot consortium, an advanced pre-clinical candidate (18-20), and more experimental molecules that show promise (21).

Notwithstanding these successes, both the resistance that may be expected to emerge (22,23), and the inevitable liabilities of the early drugs support the discovery of new scaffolds. Accordingly, we targeted the structure of  $M^{\text{Pro}}$  for large library docking, seeking new starting points for lead discovery. Docking a library of over 1.2 billion “tangible” (make-on-demand) lead-like molecules and 6.5 million tangible lead-like electrophiles from Enamine *REAL* space

(<https://enamine.net/compound-collections/real-compounds>) led to M<sup>Pro</sup> inhibitors from 37 scaffolds, with affinities ranging from the low  $\mu\text{M}$  to 200  $\mu\text{M}$ . Crystal structures for eight of the new inhibitors bound to M<sup>Pro</sup> largely confirmed the docking predictions, while cell-based antiviral activity for two of the new inhibitors supports their further optimization (**Fig. 4.1**).

While the multiple chemotypes explored here, and their crystallographic structures, may template further optimization of M<sup>Pro</sup> inhibitors, a feature of these studies was docking hit rates between 7 and 15%, with hits often in the mid- $\mu\text{M}$  range. These rates and affinities are substantially worse than observed in many GPCRs, integral membrane proteins, transporters, and enzymes like  $\beta$ -lactamase (24-33). Meanwhile, the optimized affinities reached here were meaningfully worse than those achieved by approaches such as fragment-based discovery, both against M<sup>Pro</sup> itself and against other SARS-2 targets, like macrodomain (34,35). What we might learn from these differences for future studies will be discussed.

## 4.3 Results

### Assay development and substrate design

M<sup>Pro</sup> is the fifth nonstructural protein (Nsp5) encoded by SARS-CoV-2 and is a homodimeric cysteine protease with a catalytic diad comprised of Cys145 and His41. M<sup>Pro</sup> has a P1 primary specificity determinant of glutamine and a preference for aliphatic residues in the P4 and P2 positions, while alanine and serine are preferred in the P1' position (36) (**Fig. 4.1D**). The catalytic cycle is typical of many cysteine proteases, with the catalytic Cys145 primed by proton transfer to His41 and formation of an acyl enzyme intermediate via nucleophilic attack of Cys145 at the scissile peptide carbonyl function. The thioester intermediate is then hydrolyzed by an attacking water to free the catalytic cysteine and initiate another catalytic cycle (37).

Crucial to inhibitor testing was the design and synthesis of an optimal substrate, as was done previously for SARS CoV M<sup>Pro</sup> (38) (**Fig. 4.1**). The endogenous Nsp substrates of M<sup>Pro</sup> were compiled and a consensus sequence was observed that closely matched the individual sequence of the Nsp7 cleavage site (ATLQAIAS) (**Fig. 4.1B**). This sequence was flanked with an N-terminal Lysine-MCA fluorophore and a C-terminal DNP-quencher. Noting the preference for nonpolar residues at multiple sites, we were concerned that this substrate would have low solubility. Accordingly, two D-Arginines were coupled to the N-terminal Lysine-MCA to increase solubility (**Fig. 4.1A**). This Nsp7-like substrate yielded a favorable  $K_m$  of 12  $\mu\text{M}$  and a  $k_{\text{cat}}/K_m$  of 93,000  $\text{M}^{-1} \text{s}^{-1}$ , 3.5-fold better than that of the commonly used commercial substrate (Nsp4: AVLQSGFR;  $k_{\text{cat}}/K_m = 26,500 \text{ M}^{-1} \text{ s}^{-1}$ ) (2); this substrate was used in all enzyme inhibition assays (**Fig. 4.1C**). This more efficient Nsp7-like peptide is readily synthesized and provides the field with an optimized M<sup>Pro</sup> substrate.

In early proof-of-concept testing, we observed an intolerance of M<sup>Pro</sup> activity to high concentrations of DMSO, introduced when evaluating inhibitors and substrate itself from DMSO stocks, perhaps reflecting oxidation of the catalytic cysteine. The increased solubility of the D-Arginine-modified substrate mitigated the DMSO effect by reducing the volume of DMSO needed in substrate aliquots. In addition, we found that ethanol and acetonitrile were better tolerated by the enzyme, though these solvents have issues with volatility (**Fig. 4.S1A**). These observations highlight the importance of controlling and minimizing the addition of organic solvents in M<sup>Pro</sup> activity assays and provide alternatives when DMSO is unsuitable. We also found that small amounts of non-ionic detergent were crucial for retaining M<sup>Pro</sup> activity in our *in vitro* assays. Removing the 0.05% Tween-20 we used in our assays resulted in no observed substrate cleavage. Activity could be recovered by increasing addition of bovine serum albumin (BSA); these effects may reflect M<sup>Pro</sup> sequestration on by reaction-well polymer and highlight the need of detergent or enzyme stabilizing additives (**Fig. 4.S1B**). We tested three previously reported compounds under our assay conditions. The covalent inhibitor nirmatrelvir had a similar IC<sub>50</sub> as reported (6), and while two non-covalent inhibitors (PET-UNK-29afea89-2 and VLA-UCB-1dbca3b4-15) had IC<sub>50</sub> values 2- to 5-fold higher versus reported values (18), this likely reflects simply different substrate concentrations and K<sub>m</sub> values in the different assays (see Methods). These rates provide a reference for comparing the different inhibitors.

### **Non-covalent docking screen and compound optimization for M<sup>Pro</sup> inhibitors**

Seeking new inhibitors, we began with a SARS-CoV-2 M<sup>Pro</sup> crystal structure in complex with an  $\alpha$ -ketoamide covalent inhibitor (PDB 6Y2G) (39). To define hot-spots for ligand docking in the active site, we modeled a complex of SARS-CoV-2 M<sup>Pro</sup> bound to a non-covalent SARS-CoV M<sup>Pro</sup>

inhibitor (PubChem SID87915542) (41) (non-covalent inhibitor complex crystal structures of the enzyme from SARS-CoV-2 were at that time unavailable). The crystal structure of the non-covalently ligated SARS-CoV M<sup>Pro</sup> (PDB 3V3M) (41) was structurally aligned onto the SARS-CoV-2 structure, the atomic coordinates of the  $\alpha$ -ketoamide inhibitor were replaced with those of the non-covalent SARS-CoV M<sup>Pro</sup> inhibitor SID87915542 ( $IC_{50} = 4.8 \mu\text{M}$ ) (41) and the complex was energy-minimized (Methods). After calibration of the docking parameters (42) (Methods), approximately 225 million neutral molecules, mainly from the lead-like subset of the ZINC15 library (43) (molecular weight (MWT) ranging from 250-350 amu and  $\text{clogP} \leq 4.5$ ) were docked against M<sup>Pro</sup>. Another 110 million molecules with  $350 < \text{MWT} \leq 500$  were docked in a separate screen. Docked molecules were filtered for intramolecular strain (44) and selected for their ability to hydrogen bond with Gly143, His163, or Glu166, and make favorable non-polar contacts with Met49 and Asp187. Ultimately, 220 molecules were prioritized, of which 194 (88%) were successfully synthesized by Enamine. Enzymatically, compounds were first tested at a concentration of 100  $\mu\text{M}$  using the fluorescence-based substrate cleavage assay, and 19 showed >30% inhibition of enzyme activity and were prioritized for full concentration-response curves. Concentration-response experiments revealed 12 molecules with  $IC_{50}$  values  $\leq 300 \mu\text{M}$ , a hit rate of 6% (12 hits/194 molecules tested);  $IC_{50}$ s ranged from 97 to 291  $\mu\text{M}$  (Table 4.1, Fig. 4.S2.1, Fig. 4.S2.2).

As DMSO had been observed to lower enzyme activity, the actives, initially tested from 10 mM DMSO stocks, were re-tested against M<sup>Pro</sup> from 30 mM acetonitrile (ACN) or ethanol (EtOH) stocks. Eleven compounds showed clear dose-response with  $IC_{50}$  values ranging from 30 to 200  $\mu\text{M}$ . Although covalent docking was not employed in this campaign, we noted three initial docking



hits (**ZINC338540162**:  $IC_{50}[\text{ACN}] = 30 \mu\text{M}$ , **ZINC271072260**:  $IC_{50}[\text{ACN}] = 143 \mu\text{M}$  and **ZINC795258204**:  $IC_{50}[\text{DMSO}] = 177 \mu\text{M}$ ) could, in principle, inhibit  $M^{\text{Pro}}$  covalently as they contain warheads (nitrile, aldehyde) known to react with catalytic cysteines. Several initial docking hits were tested for colloidal aggregation using dynamic light scattering (DLS) and off-target counter screens against malate dehydrogenase (MDH) and AmpC  $\beta$ -lactamase (45,46) (**Fig. 4.S3**). In DLS experiments, some scattering higher than  $10^6$  is observed indicating potential aggregation. While a few compounds *e.g.*, ‘**3312** showed unspecific inhibition of MDH, off-target activities were reversed by addition of 0.01% Triton X-100. As the  $M^{\text{Pro}}$  enzymatic assay is run with 0.05% Tween-20, an even stronger disruptor of colloidal aggregation than 0.01% Triton-X 100, we deemed the weak aggregation of these compounds irrelevant to their activity on  $M^{\text{Pro}}$ .

We focused on four initial hits (**ZINC346371112**:  $IC_{50}[\text{ACN}] = 98 \mu\text{M}$ , **ZINC301553312**:  $IC_{50}[\text{EtOH}] = 63 \mu\text{M}$ , **ZINC813360541**:  $IC_{50}[\text{ACN}] = 90 \mu\text{M}$  and **ZINC553840273**:  $IC_{50}[\text{ACN}] = 88 \mu\text{M}$ ) for structure-based optimization. We used the SmallWorld search engine (NextMove Software, Cambridge UK) (47) to identify purchasable analogs of these inhibitors within a 12 billion compound version of the REAL library (<https://enamine.net/compound-collections/real-compounds/real-space-navigator>), docking each analog into the  $M^{\text{Pro}}$  structure to assess complementarity. Between 10-20 analogs of each of the four inhibitors were selected for testing in the initial round of optimization (**Fig. 4.2**). For two initial hits, ‘**0541** and ‘**0273**, more potent analogs were identified in two to three rounds of this analog-by-catalog approach. The ‘**0273** analogs **Z4924562413** and **Z4946671001** had  $IC_{50}$  values of  $13 \mu\text{M}$  and  $5 \mu\text{M}$ , respectively (**Fig. 4.2A**). Analog of the initial docking hit ‘**0541**, such as **Z4929615577** and **Z4929616137**, reached similar potencies of  $10 \mu\text{M}$  and  $8 \mu\text{M}$ , respectively (**Fig. 4.2G**).

### Crystal structures of the non-covalent inhibitors

To investigate how the docked poses of the new inhibitors corresponded to their true binding modes, and to inform further optimization, crystal structures of three of the optimized non-covalent inhibitors were determined with resolutions ranging from 2.12 Å to 2.59 Å. For the ‘0273 analog, **SG-0001** ( $IC_{50} = 55 \mu\text{M}$ , **Fig. 4.2A-C**), the crystal structure revealed only moderate density for the ligand. Still, the predicted binding pose compared well with the experimentally determined pose, with a Hungarian (symmetry corrected) root mean square deviation (RMSD) of 2.2 Å. The isoquinoline group of **SG-0001** is inserted in the S1 subpocket, hydrogen-bonding with His163; this was also predicted for the pyridone carbonyl in the parent molecule ‘0273 (**Fig. 4.2B,C, Fig. 4.S4**). However, the tetrahydrobenzoxazepine ring, predicted to bind in the S2 subpocket in ‘0273, appeared much less buried in the **SG-0001** experimental structure. The crystal structure of  $M^{\text{Pro}}$  in complex with the ‘0541 analog ‘5548 superimposed with high fidelity to the docking-predicted pose, with an RMSD of 1.1 Å (**Fig. 4.2E, Fig. 4.S4**). Here, the compound’s hydantoin core hydrogen bonds with the backbone amine of Glu166 and Gly143. In addition, the crystal structure of  $M^{\text{Pro}}$  in complex with ‘6111 confirms the predicted binding pose (RMSD = 1.4 Å) with the isoquinoline placed in the S1 subpocket and the hydrophobic spirocyclic indane group occupying the S2 pocket (**Fig. 4.2F, Fig. 4.S4**).

## A second docking screen for non-covalent inhibitors of M<sup>Pro</sup>

As our studies progressed, other groups identified potent inhibitors with scaffolds resembling our own<sup>21</sup>. We therefore performed a second docking campaign, seeking to incorporate insights emerging from our own results and those from other studies (Methods) emphasizing the discovery of novel chemotypes.

The new docking screen targeted the SARS-CoV-2 M<sup>Pro</sup> crystal structure in complex with MAT-POS-b3e365b9-1 (MPro-x11612.pdb) (18), a non-covalent ligand reported by the COVID-19 Moonshot consortium. Compared to the previous docking template (PDB 6Y2G), the MAT-POS-b3e365b9-1-bound site is slightly smaller, with the 2-turn alpha helix between Thr45 and Leu50, and the loop between Arg188 and Ala191 shifted inwards by roughly 2 Å, constricting the shape of the P2 sub-pocket. After calibration of docking parameters, ensuring the model prioritizes 15 previously reported M<sup>Pro</sup> inhibitors against different decoy sets (42,48), we used the ZINC library (<https://cartblanche22.docking.org/>) to dock 862 million neutral compounds with 18-29 non-hydrogen atoms from the Enamine *REAL* database (Methods).

High-ranking docked molecules were filtered for novelty by removing those with ECFP4-based Tanimoto coefficients (Tc) greater than 0.35 to 1,716 SARS-CoV-2 M<sup>Pro</sup> inhibitors (Methods). Roughly 9,500 of these were graphically evaluated for favorable contacts, and 146 compounds were *de novo* synthesized by Enamine Ltd. Of these, 17 inhibited M<sup>Pro</sup> with IC<sub>50</sub> values ≤ 200 μM (Table 4.2, Fig. 4.S2.3, Fig. 4.S2.4) for a hit rate of 12% (17 hits/146 tested). To our knowledge, none of the new actives fell into scaffolds that have been previously reported for M<sup>Pro</sup>. Compared to the first docking screen, several initial hits from the second screen showed slightly higher

activity, such as **Z3535317212**, with an  $IC_{50}$  value of 29  $\mu$ M. For **'7212**, the docked pose suggests hydrogen bonds between the compound's dihydrouracil core and Glu166 as well as Gly143, in addition to hydrogen bonds between the compound's pyridinol group (**Fig. 4.S2.3**). Five docking hits (**Z5420225795**:  $IC_{50}$  = 40  $\mu$ M, **Z1669286714**:  $IC_{50}$  = 110  $\mu$ M, **Z1355254448**:  $IC_{50}$  = 110  $\mu$ M, **ZINC5420738300**:  $IC_{50}$  = 160  $\mu$ M, **Z2195811405**:  $IC_{50}$  ~200  $\mu$ M) share a common ketoamide functional group predicted to form one hydrogen bond to Glu166, however, we note that ketoamide might also inhibit  $M^{Pro}$  through covalent linkage to Cys145. As in the first docking campaign, hits were tested for colloidal aggregation. A few compounds (**'7900**, **'8488**, **'1405**, **'8300**) had higher DLS scattering or caused >50% inhibition of MDH in the absence of detergent, which was reversed by 0.01% Triton X-100 (**Fig. 4.S3**). We therefore conclude that the measured activities of those compounds at  $M^{Pro}$ , in presence of 0.05% Tween-20, originate from specific on-target actions, but care should be taken when using related scaffolds in detergent-free experiments. Taken together, the actives from this campaign explored ten different scaffold classes with  $IC_{50}$  values better than 150  $\mu$ M. These scaffolds represent new points of departure for  $M^{Pro}$  inhibitor discovery.

#### **A covalent docking screen targeting $M^{Pro}$ Cys145.**

In addition to non-covalent inhibitors, we also sought electrophiles that could covalently modify the catalytic Cys145. We searched the 1.4 billion molecules in the ZINC15/ZINC20 (43,47) databases for three Cys-reactive covalent warheads: aldehydes, nitriles, and  $\alpha$ -ketoamides. Precedence for covalent inhibitor design is seen with a range of targets, including KRAS G12C anticancer therapeutics (sotorasib, GDC-6036) (49,50), FGFR4 inhibitor (roblitinib) with an reversible aldehyde warhead in clinical trials (51), antivirals for HCV with ketoamides (boceprevir, telaprevir) (52,53), a sickle cell hemoglobin stabilizer aldehyde inhibitor (GBT440)

(54), and the SARS-CoV-2 nitrile M<sup>Pro</sup> covalent inhibitor nirmatrelvir (6). While aldehydes, especially, are reactive electrophiles, previous work has demonstrated the viability of electrophile swapping to less reactive groups, or to irreversible groups for development goals (55-57). Dockable 3D molecules were built for covalent docking with DOCKoalent (58,59) (Methods). The molecules and their DOCKoalent files for the final 6.5 million molecules are openly-available at <http://covalent2022.docking.org>.

We then docked 3.6 million nitriles, 1.5 million aldehydes, and 1.4 million  $\alpha$ -ketoamides against M<sup>Pro</sup> (PDB 6Y2G) (39). The top-ranked molecules were filtered for torsional strain (44), for favorable enzyme interactions, and clustered for chemical diversity using an ECFP4-based best first clustering algorithm (Methods). Remaining molecules were visually prioritized for favorable interactions with His41, Cys145, Gly143, Thr26, or Glu166. Ultimately, 35 aldehydes, 41 nitriles, and 21  $\alpha$ -ketoamides were selected for synthesis, of which 27, 31, 16, respectively, were successfully made and tested for activity against M<sup>Pro</sup> (Methods). Those compounds with single-point percent inhibition >50% at 100  $\mu$ M—a more stringent criterion than we had used earlier—were prioritized for full concentration-dose-response assays.

Defining actives as molecules with  $IC_{50} \leq 150 \mu$ M, the hit rate for covalent docking was 15% (11 actives/74 compounds tested); the most potent had an  $IC_{50}$  of 20  $\mu$ M (**Fig. 4.3**, **Fig. 4.S5**). Eight others had  $IC_{50}$  values 25 to 100  $\mu$ M. Initial nitriles and aldehyde docking hits had activities as low as 20  $\mu$ M in compound '5103, and 55  $\mu$ M in compound '3620, respectively. None of the  $\alpha$ -ketoamides were potent enough to be considered active. Initial docking hits were evaluated for potential M<sup>Pro</sup> inhibition through colloidal aggregation as described above (**Fig. 4.S3**). Some DLS

scattering or non-specific inhibition is observed in the AmpC and MDH enzymatic assays. However, adding 0.01% Triton X-100 in the MDH inhibition assay largely recovered enzymatic activity and eliminated any non-specific inhibition, suggesting that the measured activities in the detergent-containing M<sup>Pro</sup> enzymatic assays are not caused by aggregation (also confirmed by subsequent crystallography, see below).

The covalent inhibitors had diverse chemotype and their docked poses explored different enzyme sub-pockets (**Fig. 4.1, Fig. 4.3, Fig. 4.S5**). In the S1' pocket, hydrophobic interactions were made by compounds **'3620, '6345, '6792** in their docked poses. Hydrogen bonding with His163 in the S1 pocket was made by **'5103, '0431, '2961** in their docked poses. Several compounds, such as **'0892** and **'0292**, occupied the S2 and S3 pockets, making non-polar interactions with Met49 and Phe181. Other compounds appeared to span the binding site between the S1 and S2/S3 pockets, *e.g.* **'5156** hydrogen-bonding with Glu166. Many compounds, such as **'3620** and **'6792**, formed hydrogen-bonds with the peptide backbone atoms of Cys145, Ser144 and Gly143.

We sought to optimize several of the new covalent inhibitors, focusing on the aldehyde **'3620** with an IC<sub>50</sub> of 55 μM. These analogs were identified through multiple strategies, including simply seeking readily available “make-on-demand” congeners that fit in the enzyme site, using SmallWorld and Arthor (NextMove Software, Cambridge, UK) (47), or testing perturbations to what seemed to be key interactions. From these studies emerged 39 analogs with IC<sub>50</sub> values better than **'3620**. The most potent analog **'7021** had an IC<sub>50</sub> of 1 μM acted as a reversible inhibitor (**Fig. 4.S6**). Other analogs ranging from 2 to 48 μM had changes to different benzene substituents or bicyclic systems of **'3620** (**Fig. 4.4, Table 4.S1, Fig. 4.S7**).

In its docked pose, the pyridine nitrogen of **'7021** hydrogen bonds to Gly143 (**Fig. 4.4B**). To test the importance of this interaction, the phenyl analog of the pyridine, compound **'4218**, was synthesized and tested. This molecule lost all measurable activity ( $IC_{50} > 200 \mu M$ ), consistent with the importance of the pyridine hydrogen bonds (**Fig. 4.4C**). However, it is also likely that the more electro-deficient pyridine ring makes the aldehyde more reactive towards the catalytic Cys145. Meanwhile, removing non-polar groups from the distal phenyl ring of **'7021**, as in analogs **'9313** and **'9112**, increased  $IC_{50}$  values to  $22 \mu M$  and  $35 \mu M$ , respectively, indicating more hydrophobic bulk was preferred in the shallow subsite in which this substituted phenyl was docked.

#### **Crystal structures of the covalent inhibitors.**

To investigate how the docked poses of the covalent inhibitors corresponded to true binding modes, and to aid further optimization, crystal structures of five aldehyde inhibitors complexed with  $M^{Pro}$  were determined: **'7021** ( $IC_{50} = 1 \mu M$ ), **'9121** ( $IC_{50} = 6 \mu M$ ), **'8252** ( $IC_{50} = 6 \mu M$ ), **'9218** ( $IC_{50} = 12 \mu M$ ), and **'7356** ( $IC_{50} = 26 \mu M$ ), with resolutions ranging from  $1.90 \text{ \AA}$  to  $2.17 \text{ \AA}$  (**Fig. 4.4B, Fig. 4.4D, Fig. 4.S8**). The structures of these compounds recapitulated the docking predictions with high fidelity, with all-atom Hungarian RMSD values ranging from  $0.78 \text{ \AA}$  to  $1.75 \text{ \AA}$  (**Fig. 4.4B**). Consistent with the docking and with the results of the analogs, the pyridine nitrogen in each inhibitor hydrogen bonds with Gly143 and the thioacetal adduct hydrogen bonds with the backbone of Cys145 in the oxyanion hole of the enzyme. The hydrophobic groups on the distal aryl ring interact with residues in the S2/S3 pockets, including Met49 and Phe181 (**Fig. 4.4B, Fig. 4.4D, Fig. 4.S8**).

### **Lead inhibitors are antiviral with pan-coronaviral M<sup>Pro</sup> inhibition.**

With the progression of covalent and non-covalent inhibitor optimization, we tested several compounds in an RT-qPCR viral infectivity assay in HeLa-ACE2 cells. Compounds ‘7021 and ‘7356 had antiviral IC<sub>50</sub> values of 6.2 μM and 19.5 μM, respectively, consistent with their *in vitro* IC<sub>50</sub> values of 1 μM and 26 μM (**Fig. 4.5A, Table 4.S2**). Meanwhile, no measurable antiviral activity was observed for the covalent aldehyde ‘6690, the covalent nitrile ‘5103, and the non-covalent compound ‘6137, with *in vitro* IC<sub>50</sub> values of 2 μM, 20 μM, and 8 μM, respectively. What separates the antiviral actives from the inactives remains unclear. We also tested ‘7021 for its ability to inhibit M<sup>Pro</sup> of other coronaviruses, SARS-CoV-1 and MERS (**Fig. 4.S9**). ‘7021 inhibited the SARS-CoV-1 M<sup>Pro</sup> with an IC<sub>50</sub> of 8 μM, similar to its SARS-CoV-2 M<sup>Pro</sup> IC<sub>50</sub> of 1 μM, however it was a weaker inhibitor for the MERS M<sup>Pro</sup> with an IC<sub>50</sub> of 50 μM (**Fig. 4.5B, Table 4.S3**).



#### 4.4 Discussion

From this study emerged 132 M<sup>Pro</sup> inhibitors with IC<sub>50</sub> values less than 150 μM, covering 37 different scaffold classes (**Fig. 4.3, Table 4.1, Table 4.2**). Of these, 15 inhibitors in 3 scaffolds inhibited the enzyme with IC<sub>50</sub> values less than 10 μM. The best covalent inhibitor, **'7021**, was confirmed to act reversibly (**Fig. 4.S6**), likely reflecting the fast-on/fast-off kinetics characteristic of aldehyde covalent inhibitors. We also present an optimized M<sup>Pro</sup> substrate for future inhibitor characterization (**Fig. 4.1**). To dock the electrophile library, we first had to create it, drawing on aldehydes, nitriles, and α-ketoamides in the expanding library of tangible molecules. This resulted in a library of over 6.5 million new electrophiles, which is openly available to the community at <https://covalent2022.docking.org>. Crystal structures of eight of the new inhibitors closely corresponded to the docking predictions (**Fig. 4.3, Fig. 4.4**). Two of the new aldehyde inhibitors had antiviral activities close to those of their enzymatic IC<sub>50</sub> values, suggesting that further optimization of this class for on-enzyme potency may presage antiviral activity (**Fig. 4.5A**).

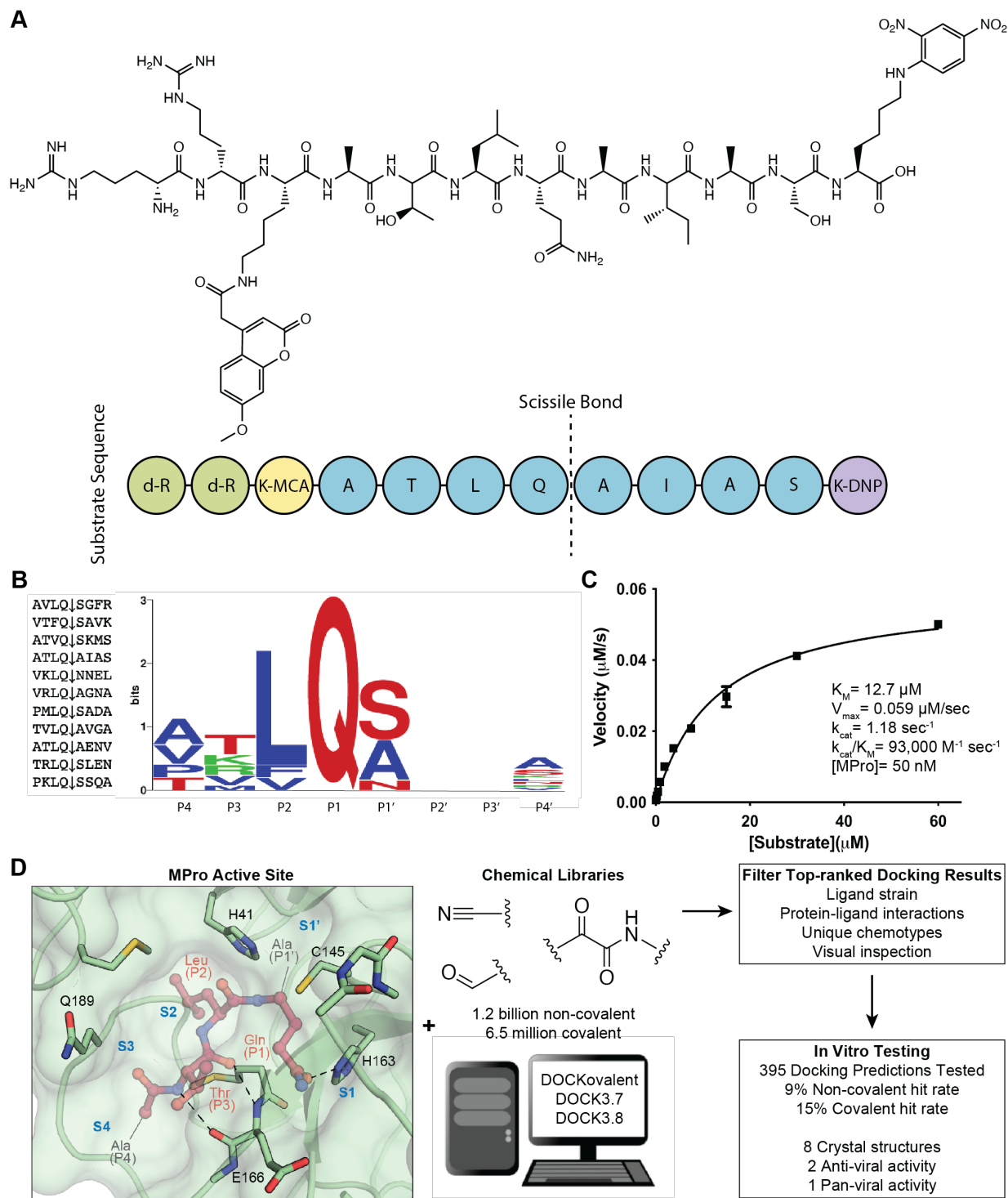
While the strengths of this study were the identification of multiple new M<sup>Pro</sup> inhibitor scaffolds, with subsequent crystal structures supporting the docking predictions, the work also revealed liabilities of the docking strategy we adopted. Docking large libraries of lead-like molecules against G protein-coupled receptors (26,27,29,60) and other integral membrane proteins (25,28) with well-formed orthosteric sites and well-defined ligand-recognition hot-spots can have hit rates in the 25 to 60% range, whereas the hit rates against M<sup>Pro</sup> were in the 6 to 15% range. A similar trend is true for the potency of the docking hits against M<sup>Pro</sup>, which were in the 20 to 100 μM range, not the low- to mid-nM range found against the integral membrane proteins. The low lead-like hit-rates against M<sup>Pro</sup> resemble those against other soluble proteins such as β-lactamase (26,61)

and the macrodomain of SARS-CoV-2 (34,35), or even against allosteric sites or peptide binding sites in GPCRs (62-64). Like M<sup>Pro</sup>, these targets are characterized by more open, solvent-exposed binding sites, lacking the high-complementarity typical of the small molecule GPCRs and other membrane receptors. This deficit can be partly overcome in sites with well-defined ligand-recognition hot spots, which can guide docking campaigns. Thus, in lead-like non-covalent and covalent (58) docking versus  $\beta$ -lactamase it has been possible to find relatively potent hits directly out of docking, notwithstanding sometimes low hit rates, and to optimize these to mid-nM potency fairly rapidly. The same may be true for typically more challenging peptide (24) and lipid receptors (65), where hot-spot recognition can help lead-like docking and optimization. Confronted with a target like M<sup>Pro</sup>, with its solvent-exposed, relatively flat interface and, certainly when this study began, lack of ligands to define receptor hot spots, a more fragment-based approach may have advantages to the larger and more complicated lead-like molecules prioritized here. As was observed in studies by the Covid Moonshot (18), and in our own experience versus the orphan SARS-CoV-2 enzyme Mac1 (34), fragment screens and fragment docking had much higher hit rates and could be used directly for optimization, but they also could be used to define and identify hot-spots enabling subsequent docking of lead-like libraries with better hit-rates and affinities (35). Such fragment based approaches, may find broad use against structurally-enabled viral enzymes that are under-characterized for ligand recognition.

These caveats should not distract from the key observations of this study. Large library docking of both lead-like molecules and covalent electrophiles has revealed 11 scaffold families of M<sup>Pro</sup> inhibitors (**Fig. 4.3, Table 4.1, Table 4.2**), the best of which act in the low  $\mu$ M range (**Fig. 4.2, Fig. 4.4**). Whereas neither hit rates nor affinities rose to levels seen against targets with well-

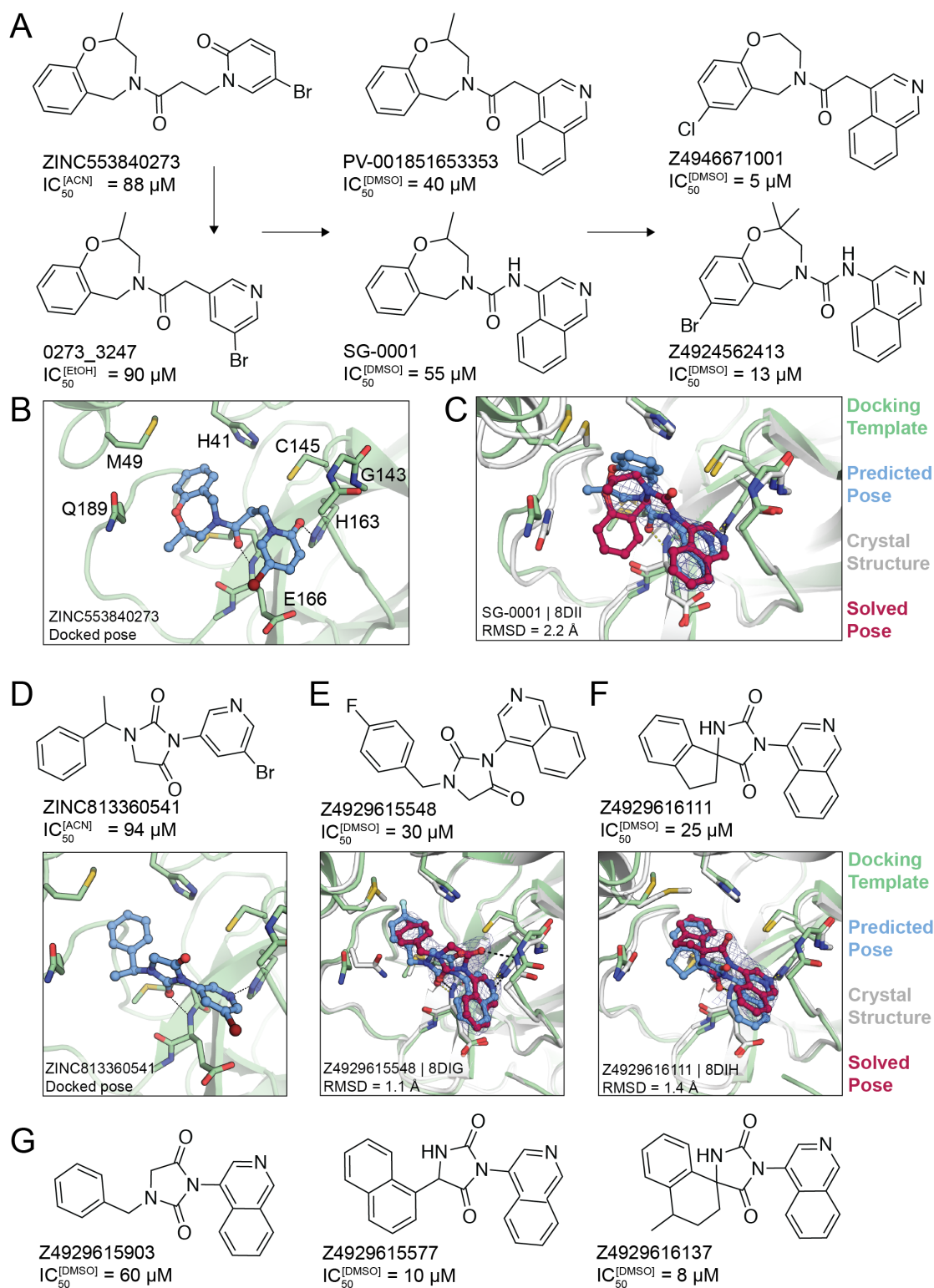
defined binding sites, eight crystal structure of characteristic lead molecules confirmed the docking poses (**Fig. 4.2, Fig. 4.4**), suggesting that, notwithstanding the lower hit rates, when the docking was right it was right for the right reasons. These structures may template the further optimization of these new M<sup>Pro</sup> inhibitors, several of which show initial antiviral activity against the virus.

## 4.5 Figures



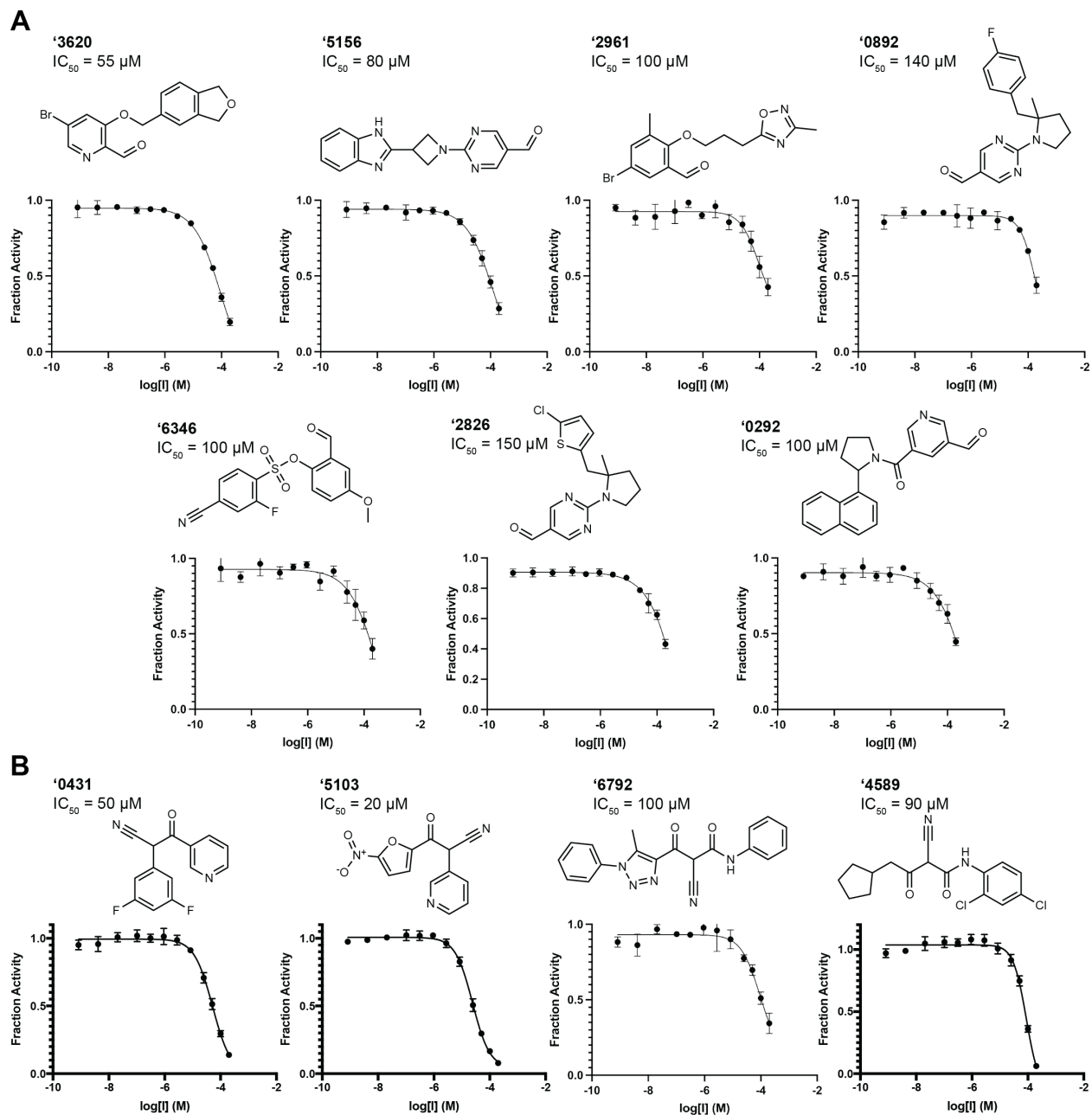
**Figure 4.1. Substrate design and assay development allows structure-based inhibitor discovery.** (A) The chemical structure of the optimized NSP7 substrate shown as a schematic (top) of the substrate sequence highlights the role of each residue (bottom). The substrate contains the P4-P4' NSP7 extended substrate sequence (blue), the fluorophore (yellow), the fluorescent

quencher (purple), and the residues for increasing solubility (green). **(B)** A list of the viral polypeptide NSP sequences (P4-P4') that are cleaved by MPro (left). The sequence logo highlighting the substrate specificity of M<sup>Pro</sup>, yielding a P4-P4' consensus sequence: ATLQ(S/A)XXA (right). **(C)** The Michaelis-Menten kinetics for the NSP7 substrate with M<sup>Pro</sup> yield parameters indicative of an optimized, efficient substrate. **(D)** SARS-CoV-2 M<sup>Pro</sup> active site (PDB 6Y2G) (39) (green; sub-pockets S1', S1, S2, S3, S4), shown here with substrate preferences (pink; P1', P1, P2, P3, P4) (modeled after PDB 3SNE) (40), was used to dock 1.2 billion non-covalent molecules and 6.5 million electrophile molecules. Top-ranked molecules were filtered and 395 were synthesized for *in vitro* testing. Some docking hits were prioritized for compound optimization, crystallography, pan-viral enzymatic activity, and cell-based antiviral activity. For C, experiments were performed in triplicate.



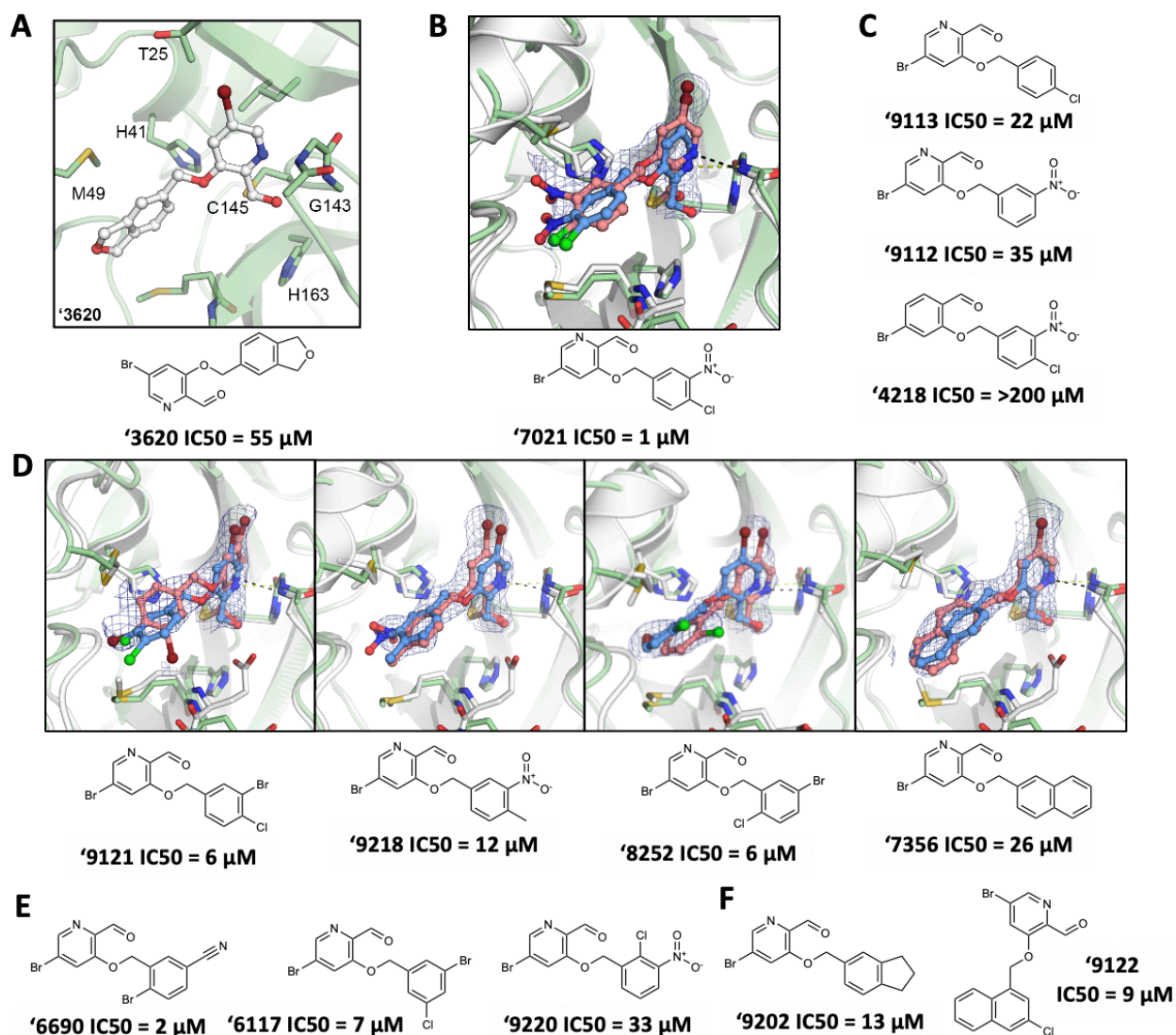
**Figure 4.2. Non-covalent compound optimization to low- $\mu M$  potencies.** (A) Progression of the ‘0273 scaffold. (B) Predicted binding pose of ‘0273. (C) Comparison of crystal structure (grey protein, red compound) and docked complex (green protein, blue compound) of SG-0001 (PDB 8DII). (D) Predicted binding pose of ‘0541. (E), (F) Comparison of crystal structures and docked

complexes of **'5548** (PDB 8DIG) and **'6111** (PDB 8DIH), respectively. **(G)** Additional **'0541** analogs with improved affinities. The 2fo-fc ligand density maps (blue contour) are shown at 1  $\sigma$ . Hungarian root mean square deviations (RMSD) were calculated with DOCK6.

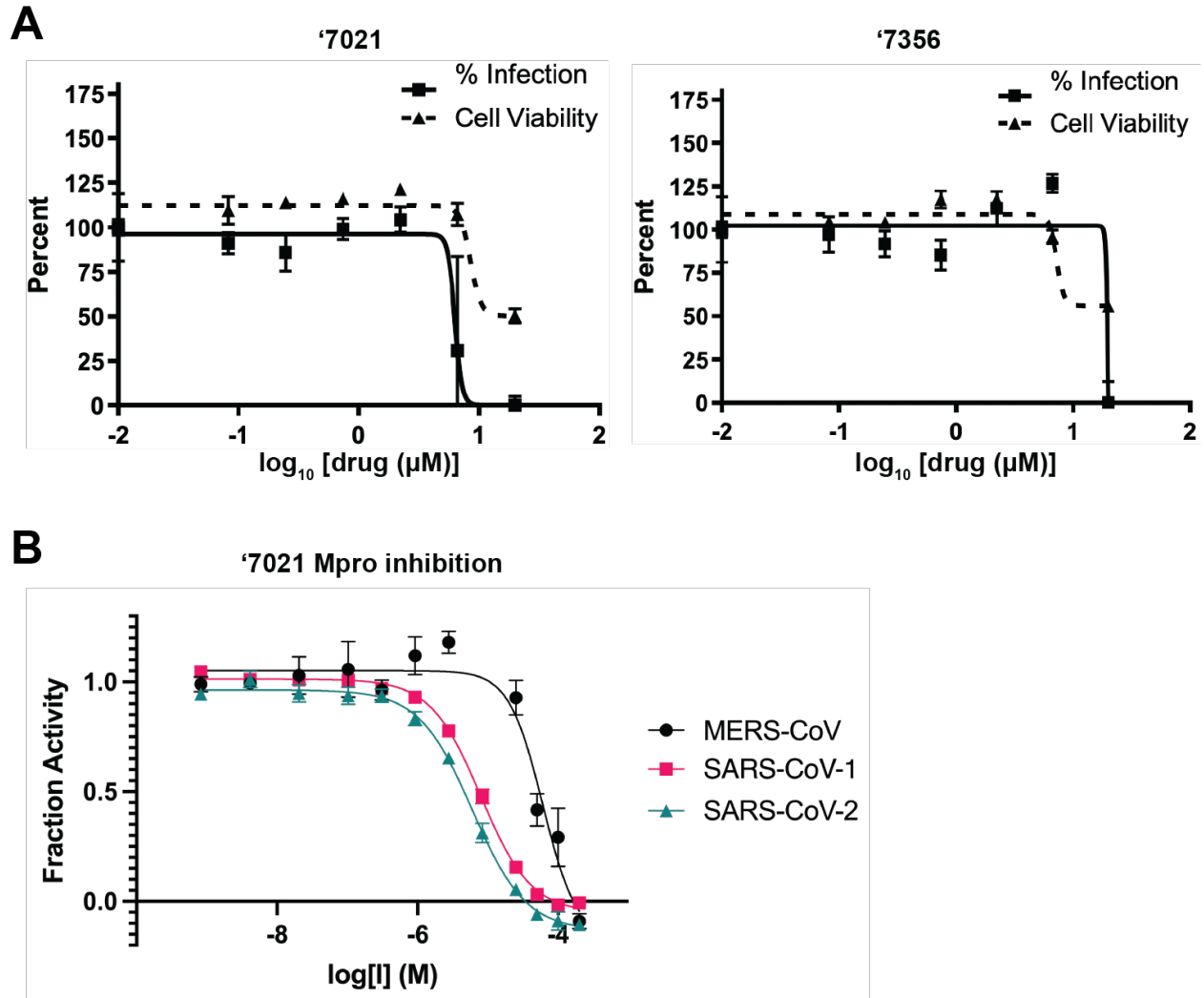


**Figure 4.3. Covalent hits from 6.5 million virtual screen. Dose response curves for (A) aldehyde and (B) nitrile docking hits. IC<sub>50</sub> values shown. All measurements were done in triplicate.**



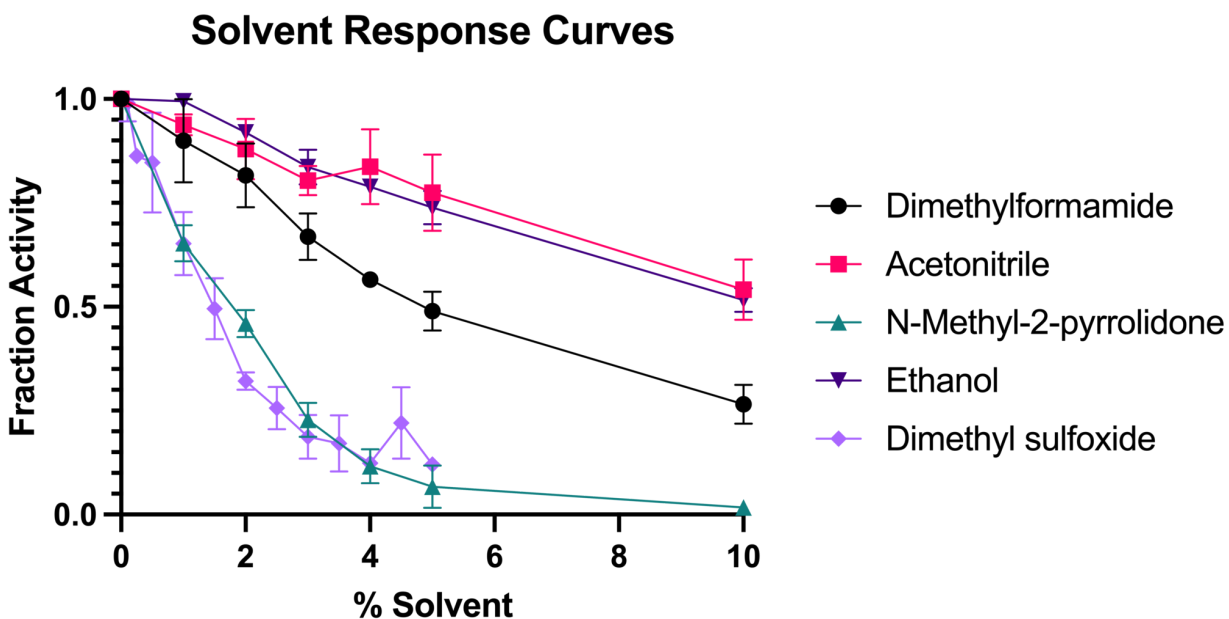


**Figure 4.4. Compound optimization of aldehyde '3620.** (A) Docked pose of docking hit '3620. (B) Crystal structure (pink carbons) and docked pose (blue carbons) comparison for analog '7021 (RMSD 1.29 Å; PDB 8DIB). (C) Hypothesis testing analogs of '7021 included removing the nitro in '9113 and the chlorine in '9112, both with weaker inhibition. Analog '4218 replaced the pyridine with a benzene eliminating inhibition. (D) Crystal structures of additional '3620 analogs comparing experimental (pink carbons) and docked (blue carbons) poses (RMSDs of 1.75 Å, 0.78 Å, 1.18 Å, and 0.84 Å, respectively; PDB 8DIC, 8DIE, 8DID, 8DIF, respectively). (E) Analogs with different benzene substituent orientations ('6690, '6117) inhibit M<sup>Pro</sup> at similar potencies. Substituents oriented like '9220 were weaker inhibitors. (F) Examples of the most potent larger hydrophobic analogs of '3620. For A-F, M<sup>Pro</sup> protein structure is PDB 6Y2G (green carbons) used in docking or from the solved structures (white carbons). Hydrogen bonds shown with dashed lines. The 2fo-*fc* ligand density maps (blue contour) are shown at 1  $\sigma$ . IC<sub>50</sub> values are shown with concentration response curves in Fig. S7. All measurements were done in triplicate.

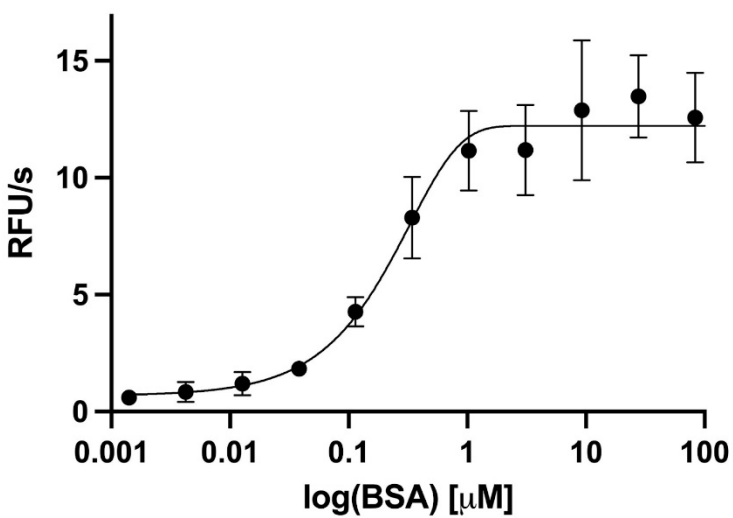


**Figure 4.5. Antiviral activity and pan-coronaviral M<sup>Pro</sup> inhibition by covalent analogs. (A) '7021 and '7356 inhibit SARS-CoV-2 infectivity with minimal impacts on cell viability. (B) '7021 also inhibits SARS-CoV-1 and MERS-CoV M<sup>Pro</sup>. All measurements were done in triplicate**

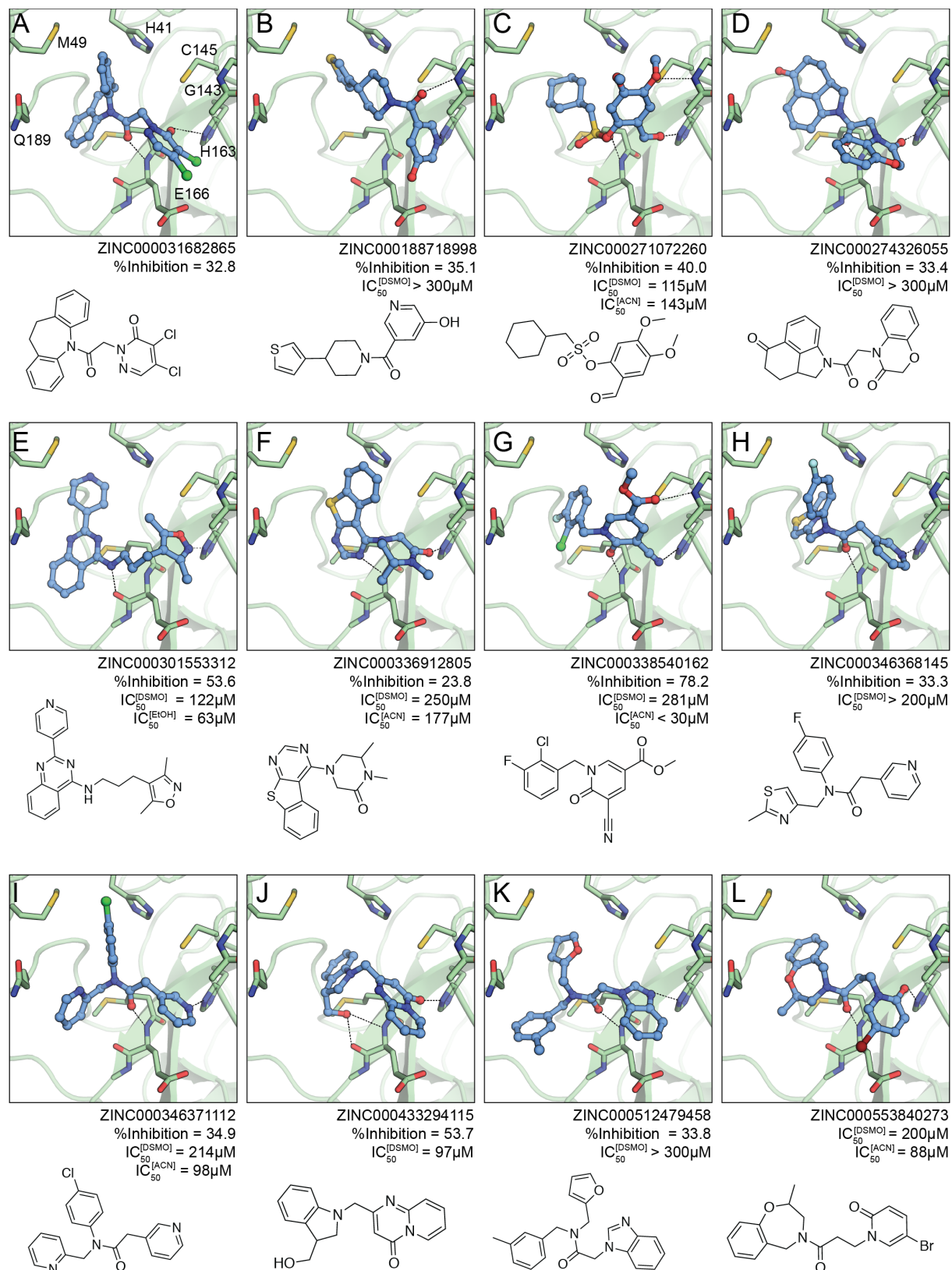
A



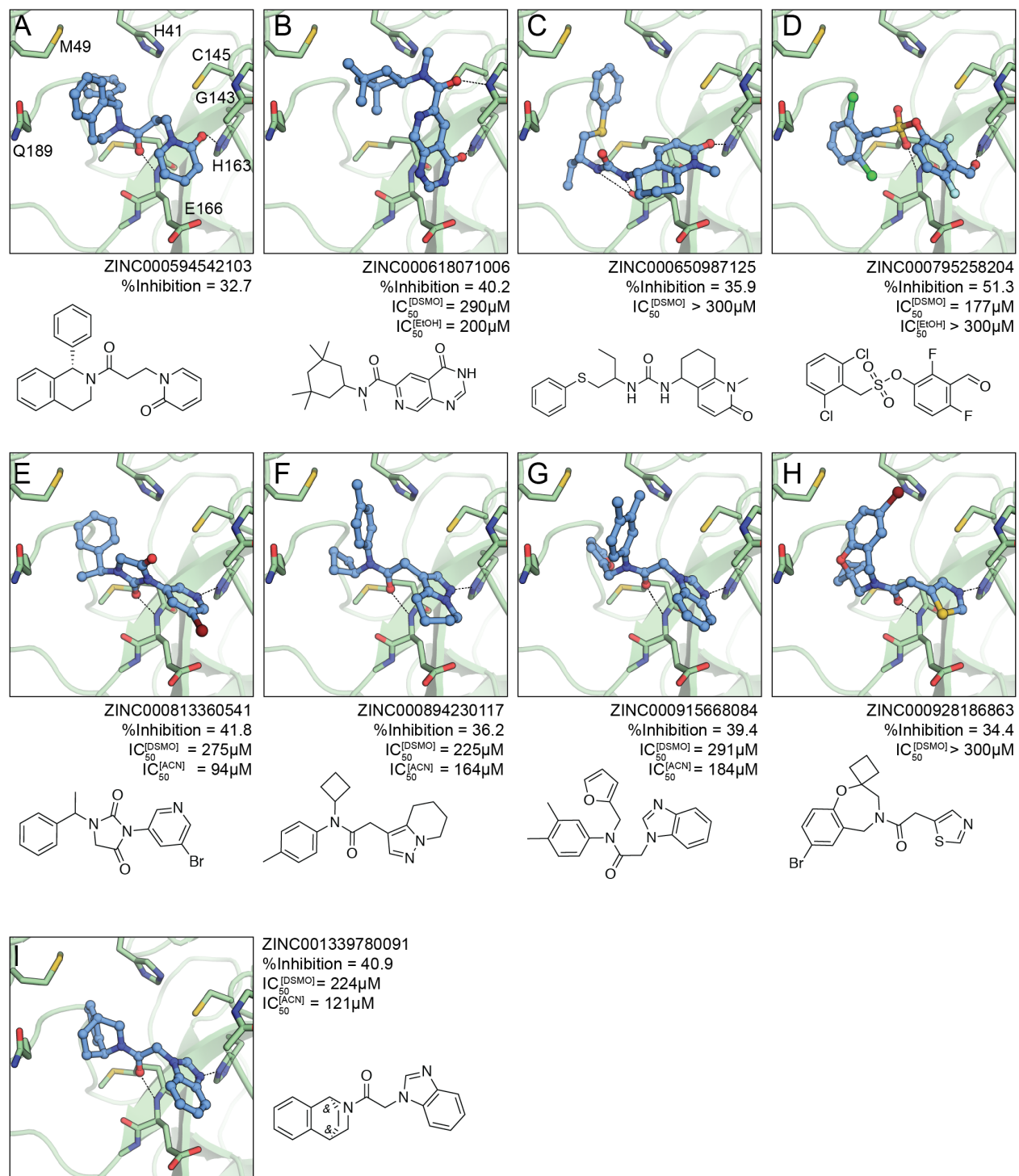
B



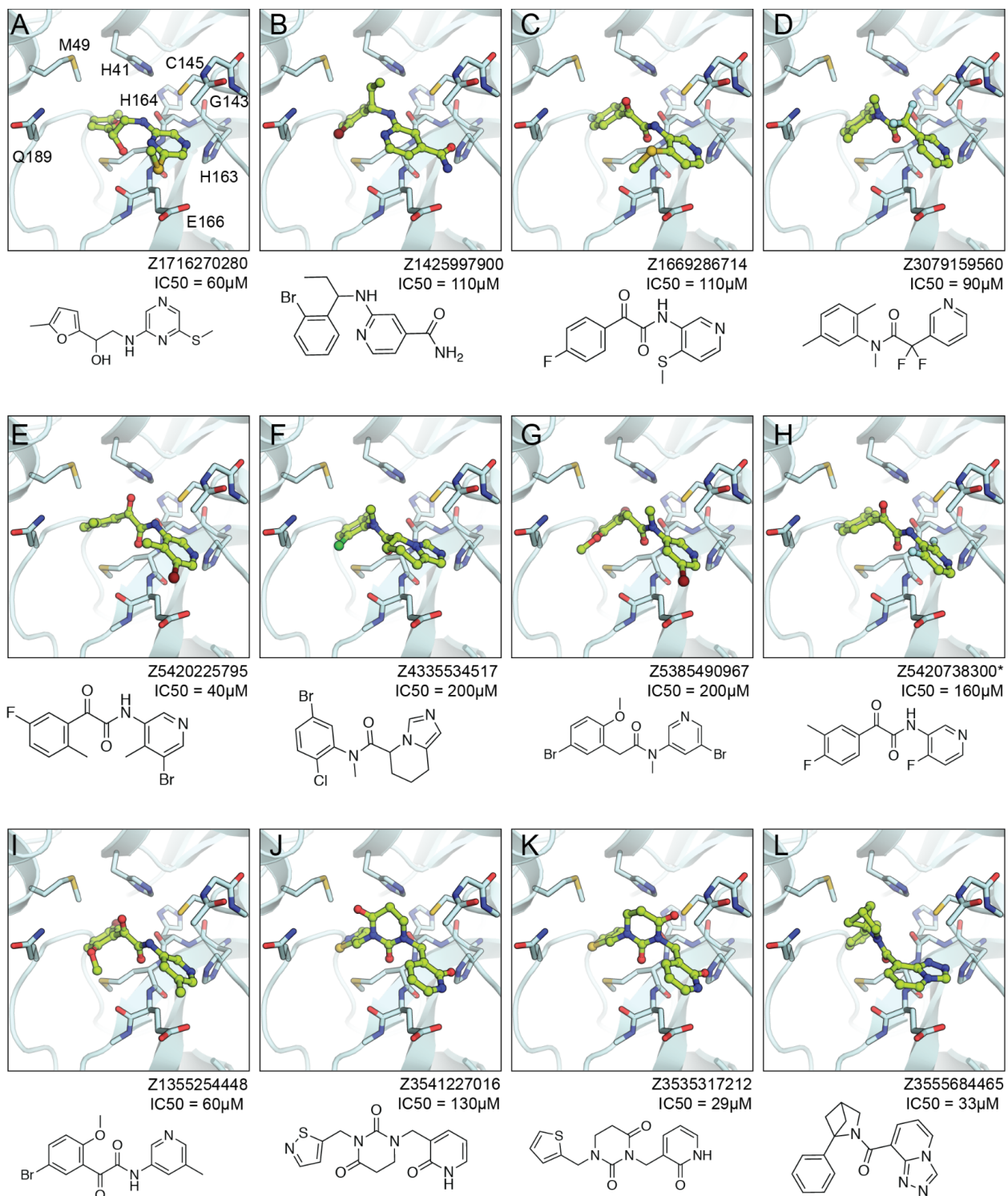
**Figure 4.S1. Assay optimization for solvent and detergent.** (A) The effect of solvents on enzymatic activity of MPro. (B) Activity loss from removal of detergent rescued with addition of BSA.



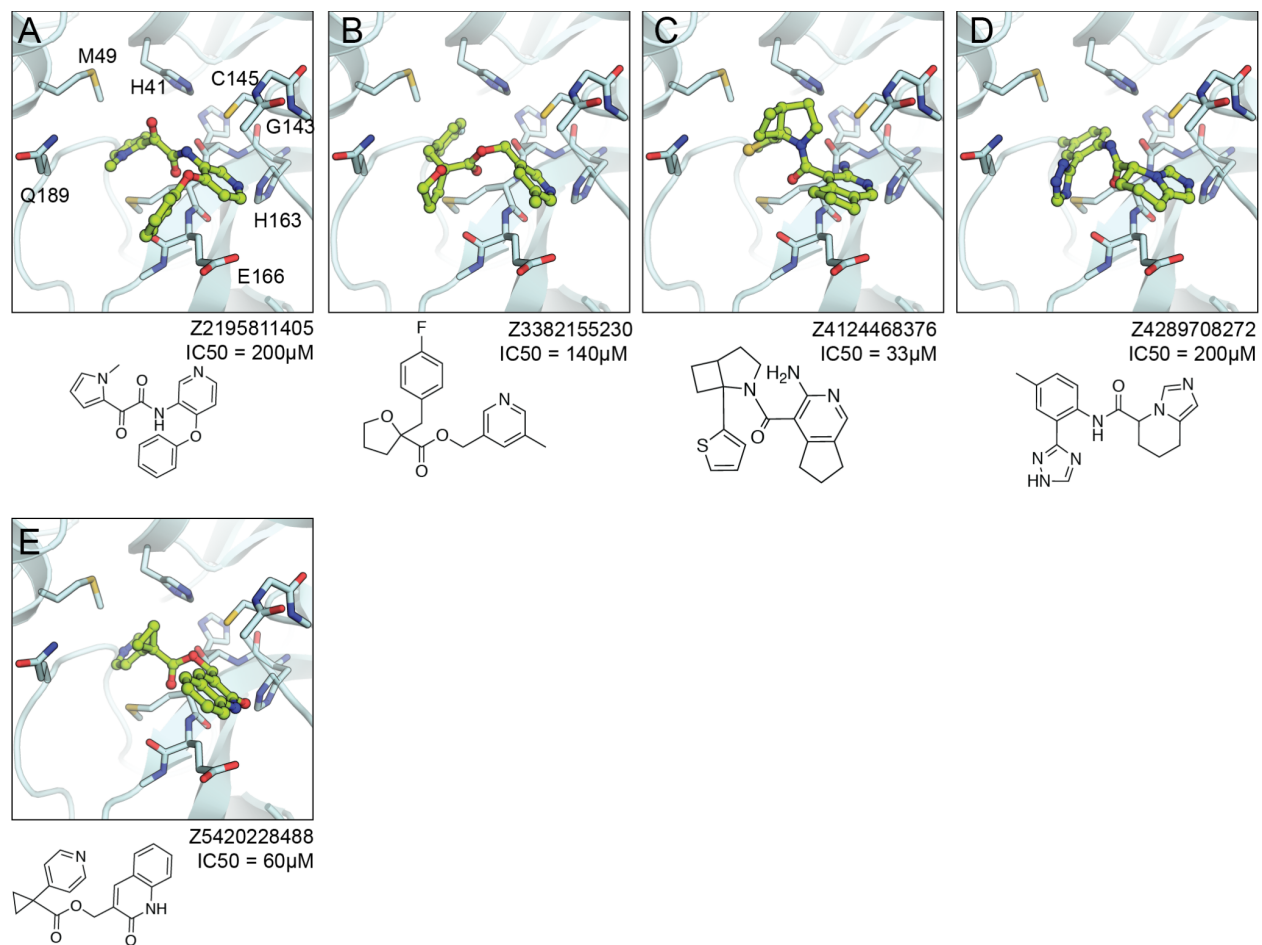
**Figure 4.S2.1. Non-covalent docking hits or compounds with >30% inhibition from first virtual screen.**



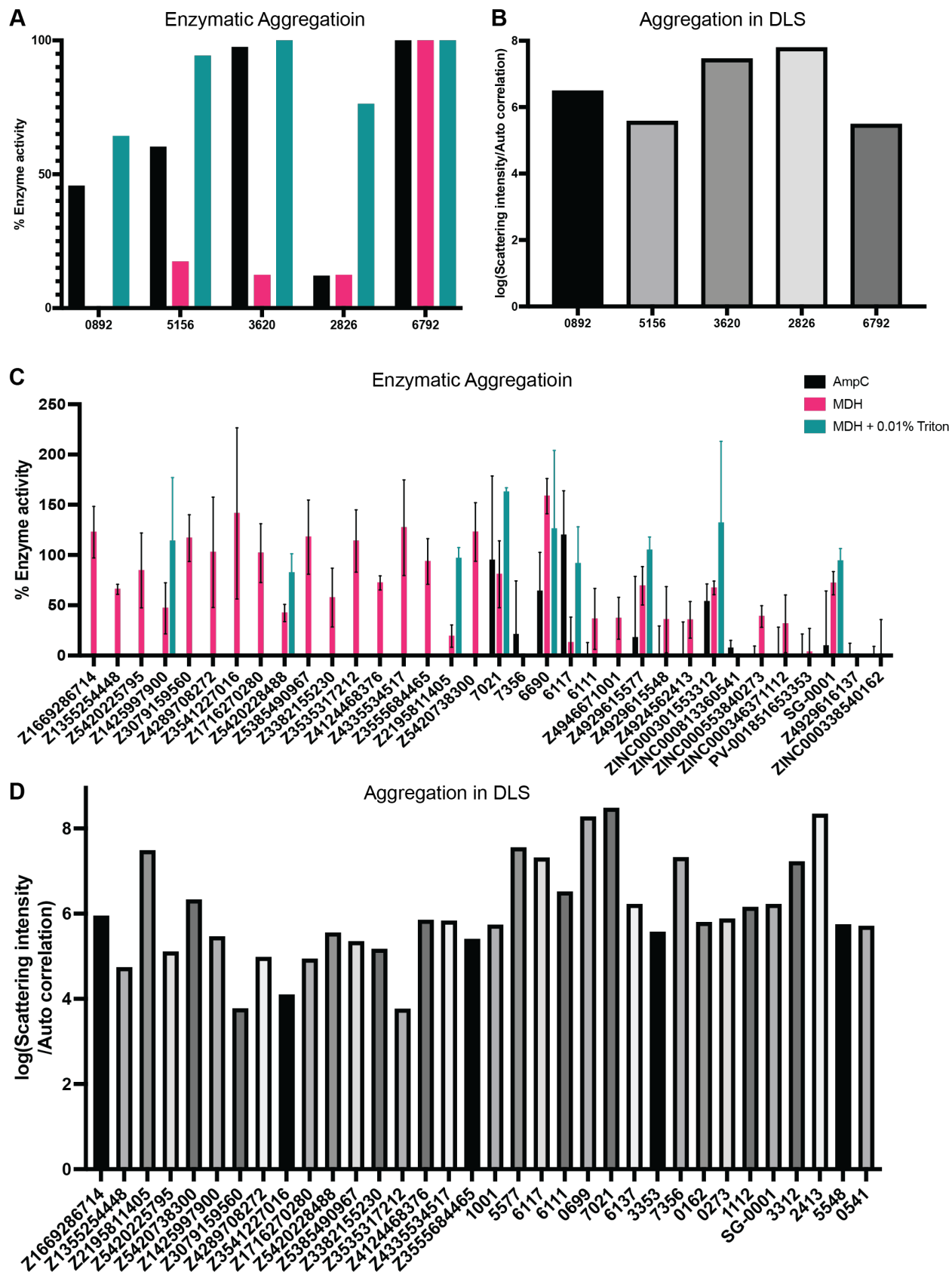
**Figure 4.S2.2. Non-covalent docking hits or compounds with >30% inhibition from first virtual screen.**



**Figure 4.S2.3. Non-covalent docking hits from second virtual screen.**

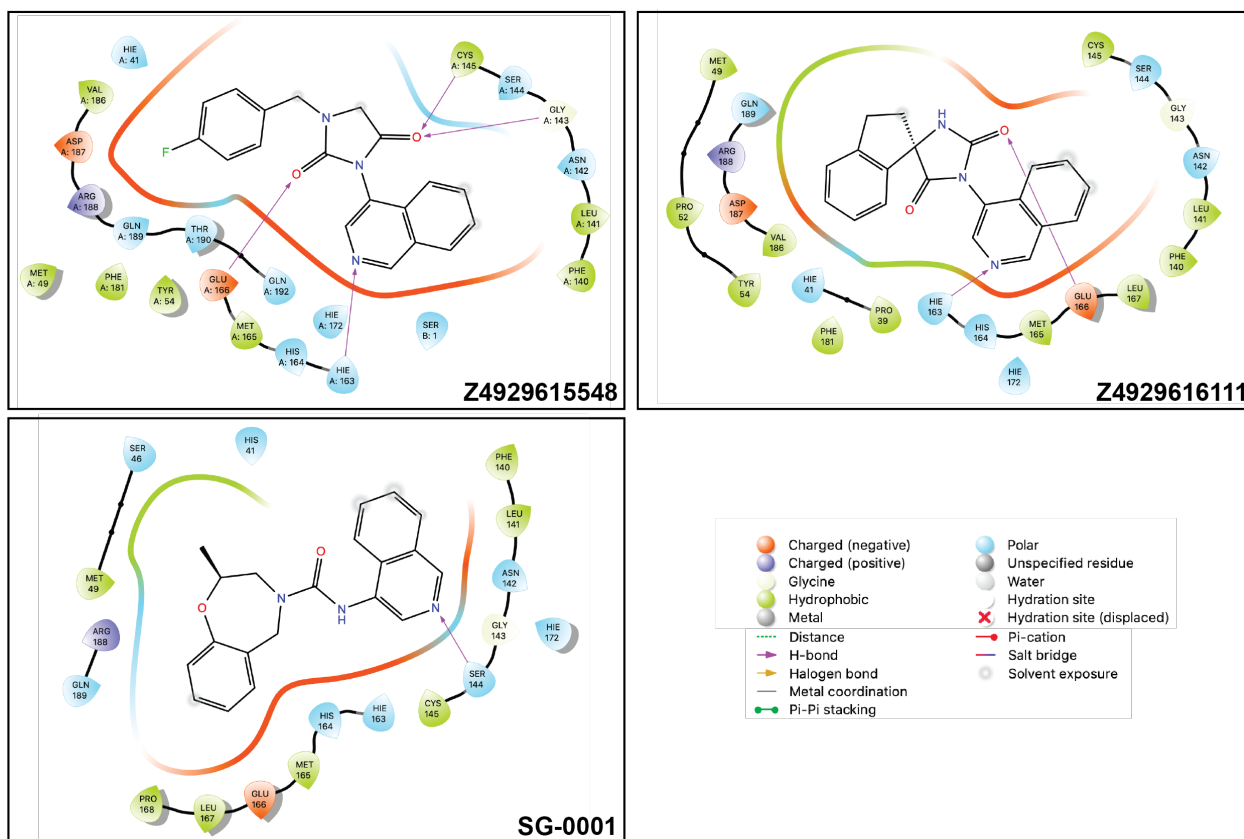


**Figure 4.S2.4. Non-covalent docking hits from second virtual screen.**

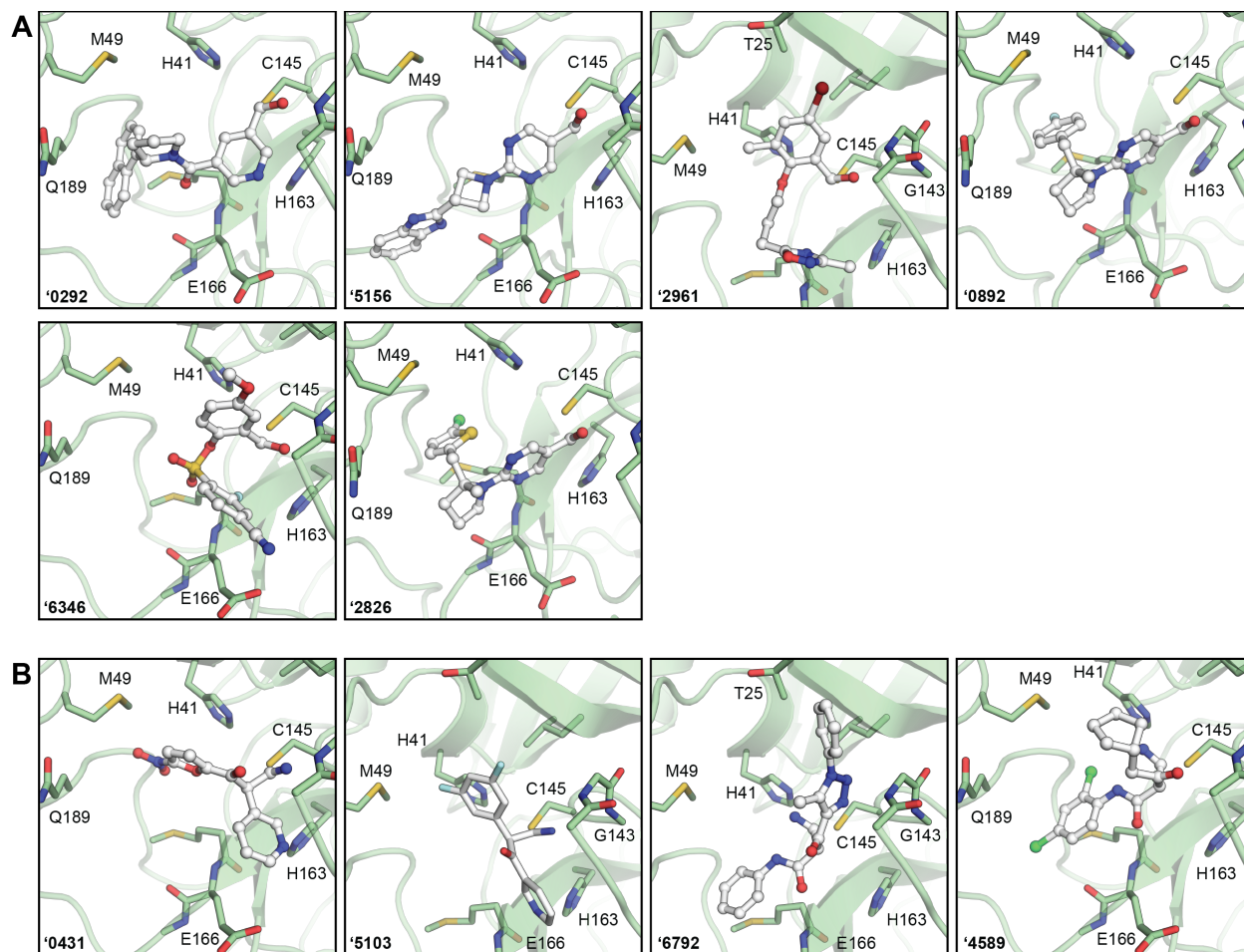




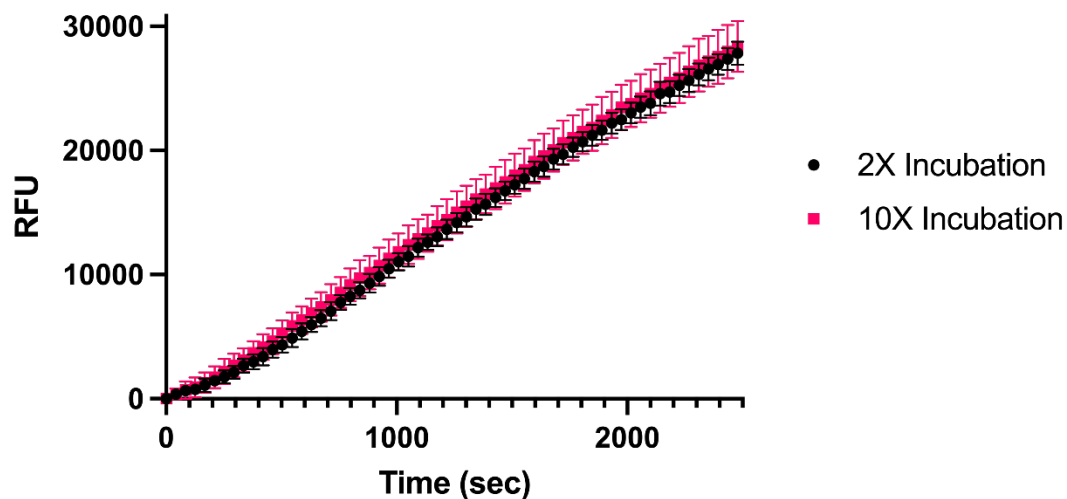
**Figure 4.S3. Evaluating aggregation potential of initial docking hits and potent analogs.** (A,C) Enzymatic inhibition by aggregation tested against AmpC and MDH (with or without 0.01% Triton X-100 detergent). (B,D) Compounds observed in DLS for forming colloidal-like particles where values above  $10^6$  have the potential for aggregating.



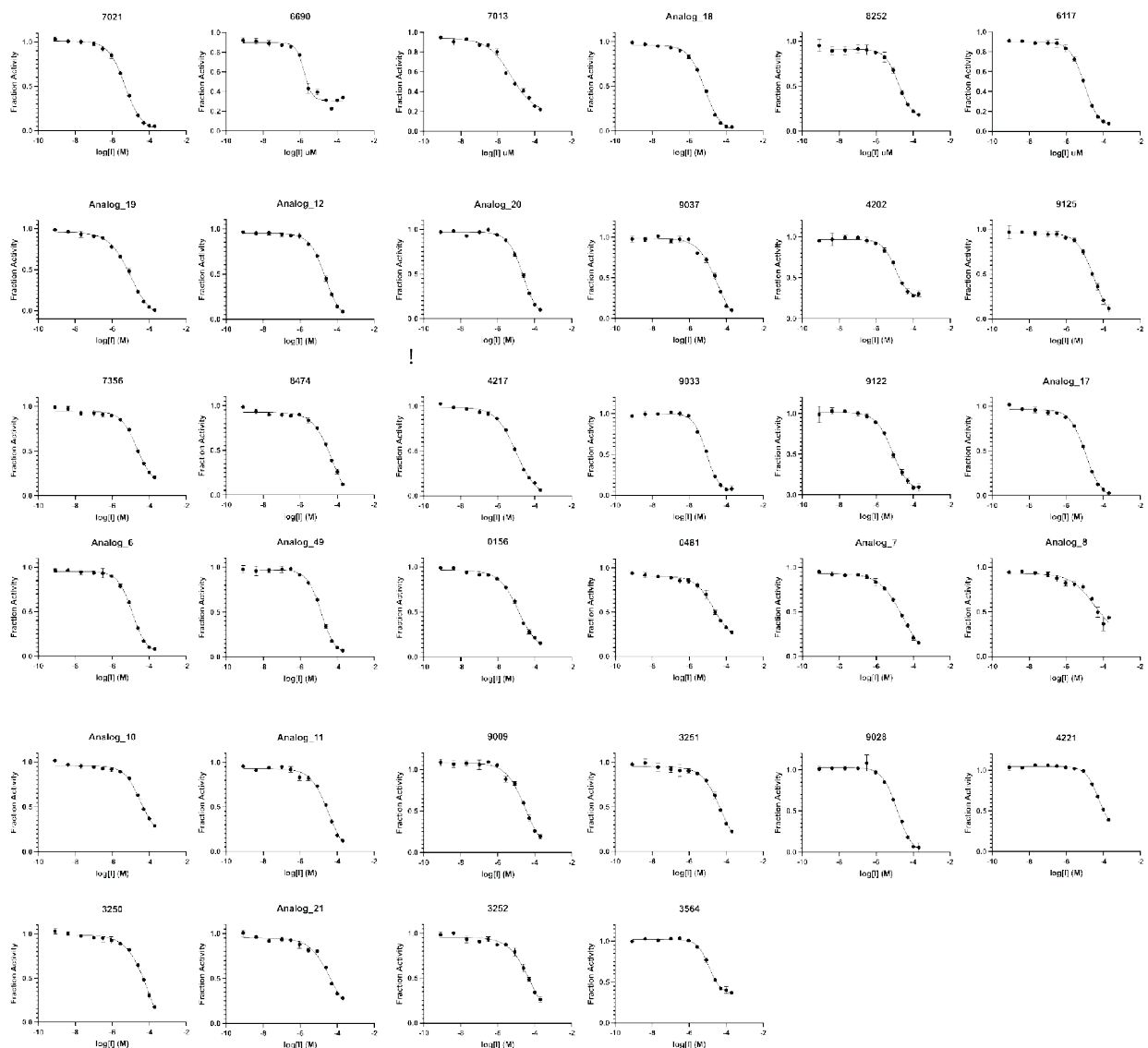
**Figure 4.S4. LigPlot visualization of MPro-noncovalent inhibitor interactions in newly solved structures.**



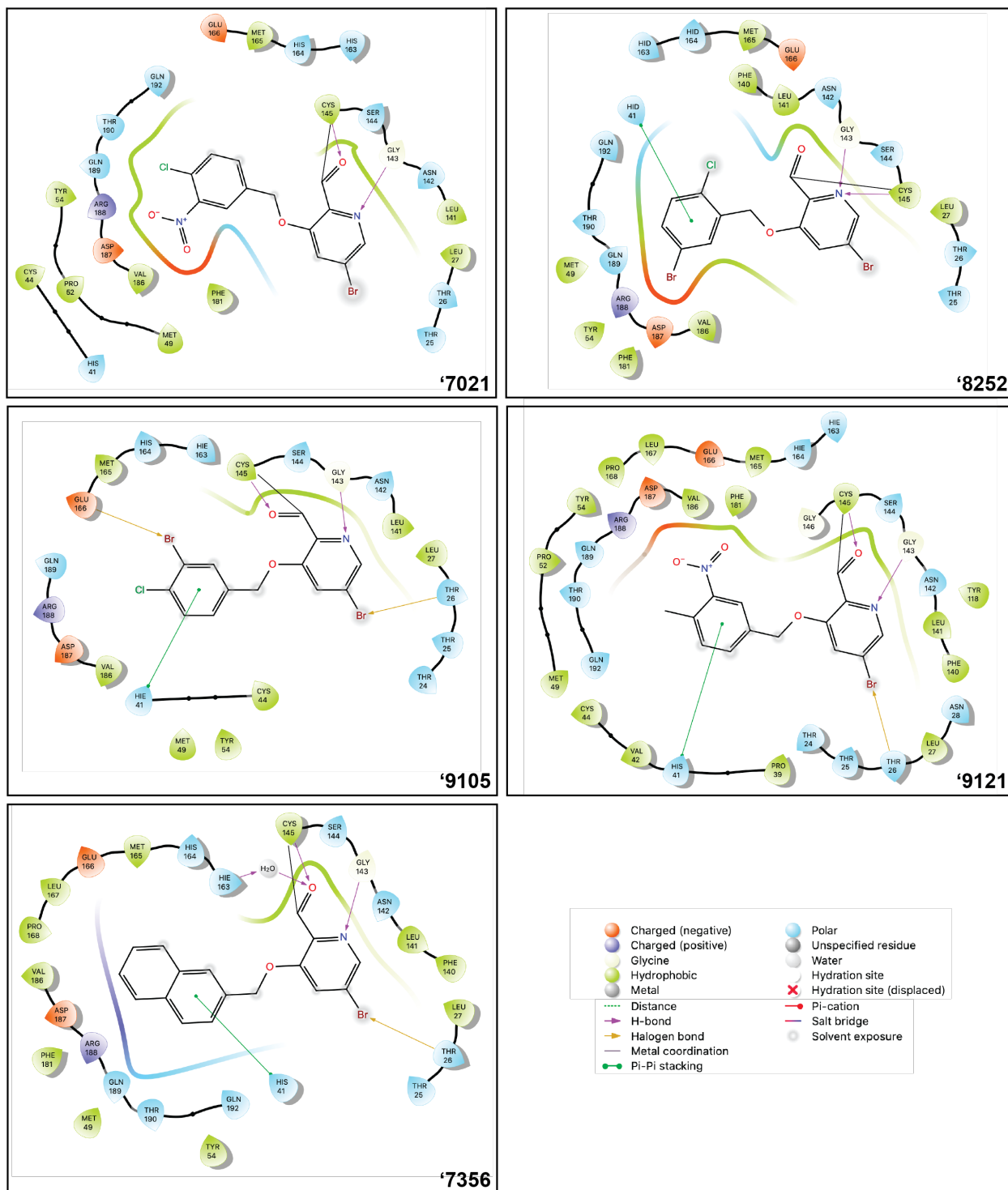
**Figure 4.S5. Docked poses of covalent hits. (A) Aldehydes. (B) Nitriles.** Covalent bonds between the warhead and Cys145 and not explicitly modeled in docking.



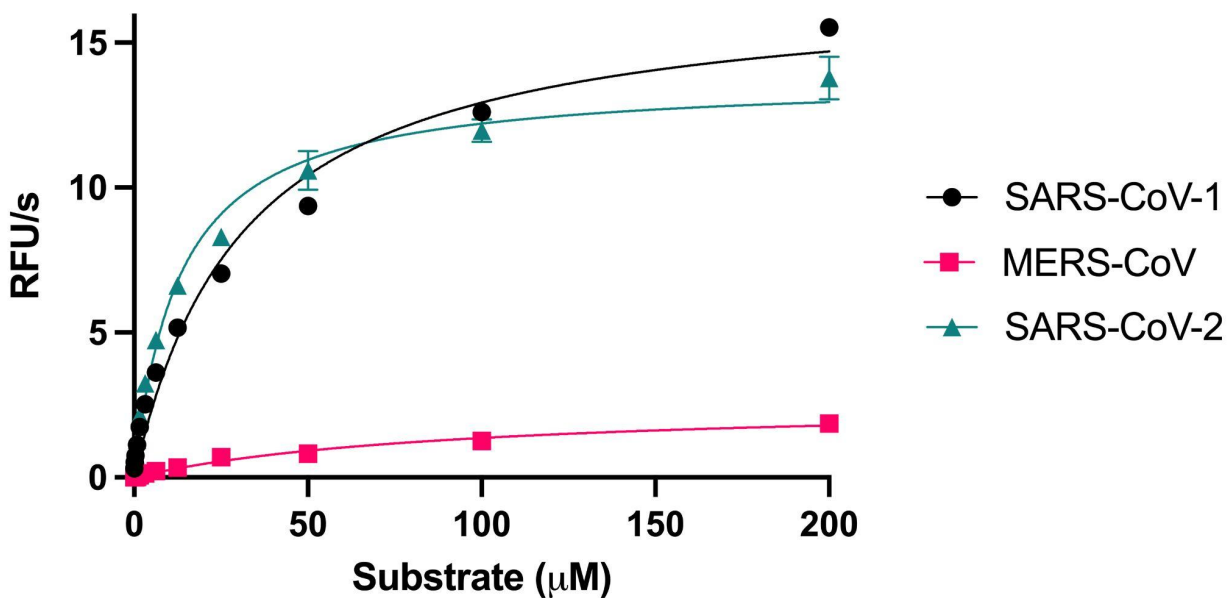
**Figure 4.S6. Reversibility of compound '7021.** Both 2X and 10X concentrations of enzyme and '7021 inhibitor were co-incubated for 1-hour. By Le Chatelier's principle, the 10X incubation should lead to more inhibited enzyme product. Both 2X and 10X incubations were diluted to 1X concentrations with final substrate concentration of 1X NSP7. Both co-incubations yielded the same velocity, evincing concentration dependent diffusion of the enzyme-inhibitor complex, thus evincing the reversibility of the '7021 compound.



**Figure 4.S7. '3620 analogs with improved potencies. IC<sub>50</sub> values in Table 4.S2.**



**Figure 4.S8. LigPlot visualization of MPro-covalent inhibitor interactions in newly solved structures.**



**Figure 4.S9. Pan-viral enzymatic activities.** Michaelis-Menten kinetics were ran side-by-side with the optimized NSP7 substrate for SARS-CoV-2 (green), SARS-CoV-1 (black), and MERS (red) major proteases.  $K_M$  values were found to be  $12\mu\text{M}$ ,  $30\mu\text{M}$ , and  $90\mu\text{M}$  for SARS-CoV-2, SARS-CoV-1, and MERS, respectively. All proceeding inhibitory assays were ran with  $K_M$  concentrations of substrate.

## 4.6 Tables

**Table 4.1. Hits from the first non-covalent docking screen.**

Chemical Structure	Compound ID	IC50 [ $\mu$ M] (solvent)	Chemical Structure	Compound ID	IC50 [ $\mu$ M] (solvent)
	ZINC346371112	214 (DMSO) 98 (ACN)		ZINC813360541	275 (DMSO) 94 (ACN)
	ZINC894230117	225 (DMSO) 164 (ACN)		ZINC553840273	200 (DMSO) 88 (ACN)
	ZINC1339780091	225 (DMSO) 121 (ACN)		ZINC336912805	250 (DMSO) 177 (ACN)
	ZINC433294115	97 (DMSO)		ZINC271072260	115 (DMSO) 143 (ACN)
	ZINC618071006	290 (DMSO) 200 (EtOH)		ZINC338540162	281 (DMSO) <30 (ACN)
	ZINC301553312	122 (DMSO) 63 (EtOH)		ZINC915668084	291 (DMSO) 184 (ACN)



**Table 4.2. Hits from the first non-covalent docking screen.**

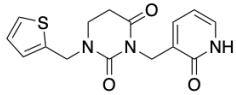
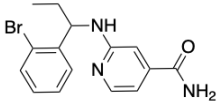
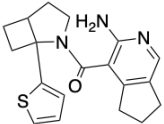
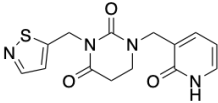
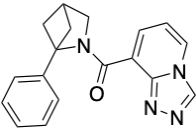
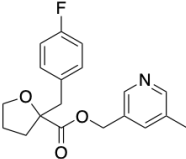
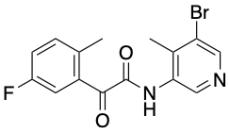
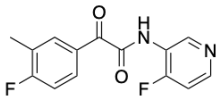
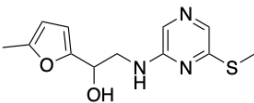
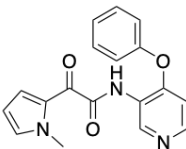
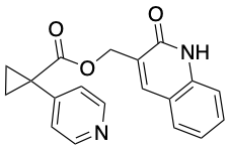
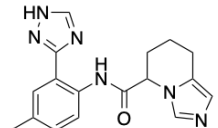
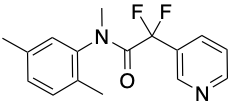
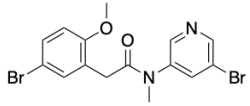
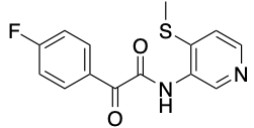
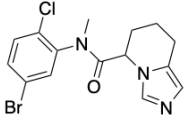
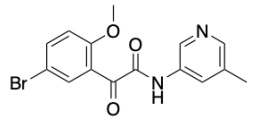
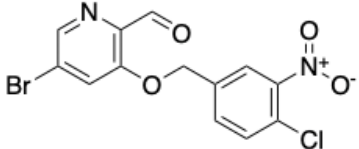
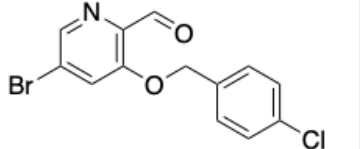
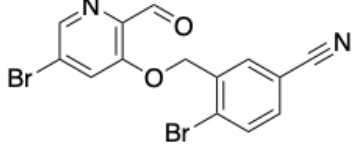
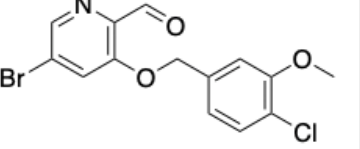
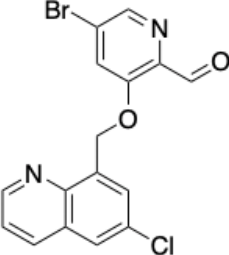
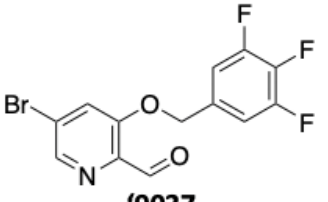
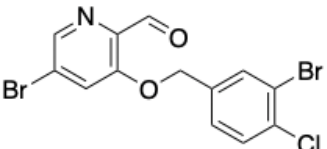
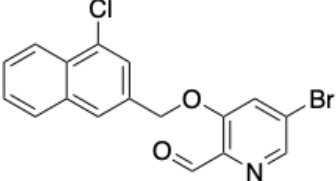
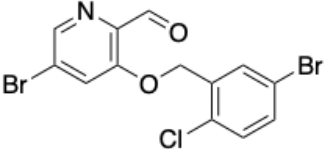
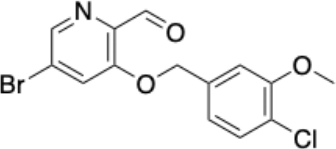
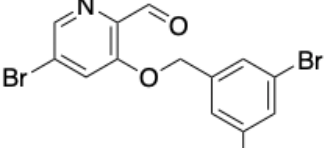
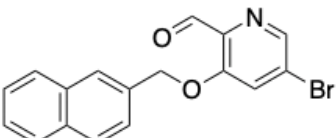
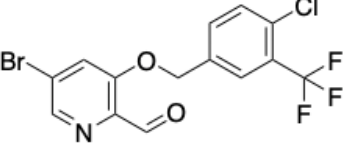
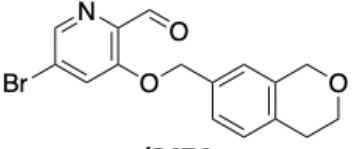
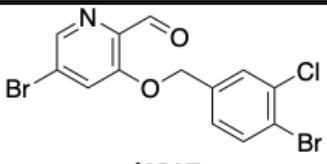
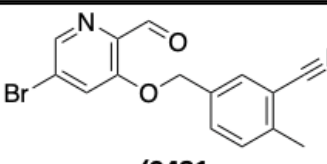
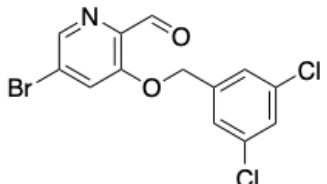
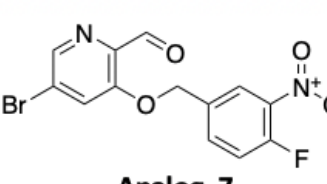
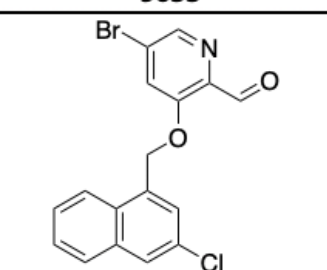
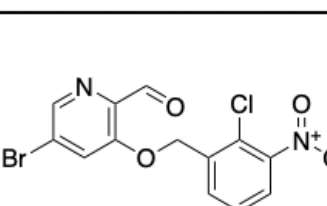
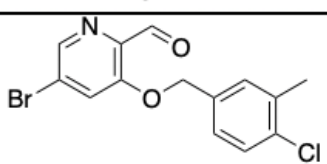
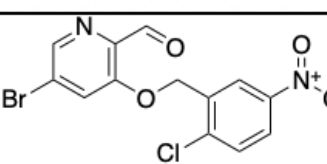
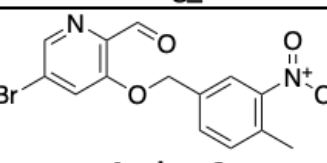
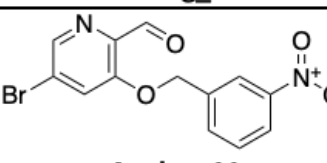
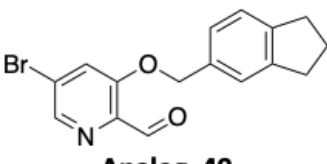
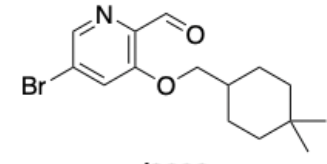
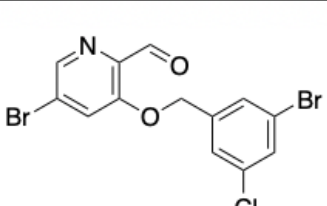
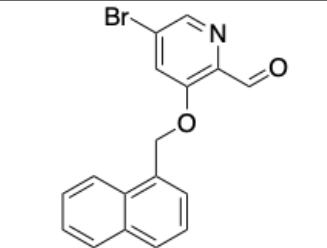
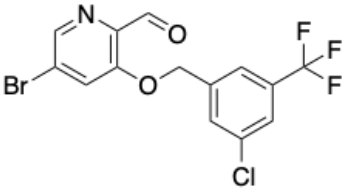
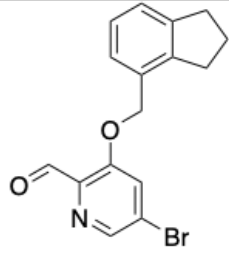
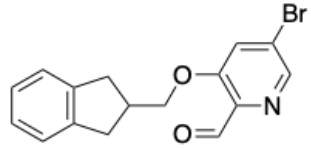
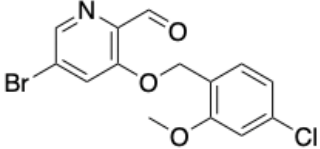
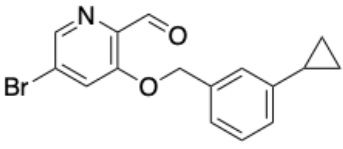
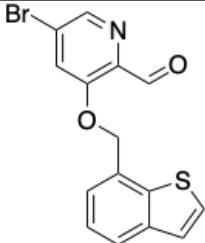
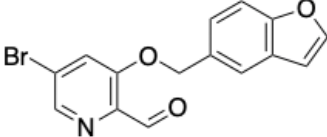
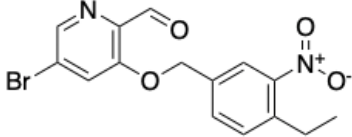
Chemical Structure	Compound ID	IC50 [μM]	Chemical Structure	Compound ID	IC50 [μM]
	Z3535317212	29		Z1425997900	110
	Z4124468376	33		Z3541227016	130
	Z3555684465	33		Z3382155230	140
	Z5420225795	40		Z5420738300	160
	Z1716270280	60		Z2195811405	200
	Z5420228488	60		Z4289708272	200
	Z3079159560	90		Z5385490967	200
	Z1669286714	110		Z4335534517	200
	Z1355254448	110			

Table 4.S1 Analogs of covalent docking hit '3620 with improved potencies.

Compound	IC50 [ $\mu$ M]	Compound	IC50 [ $\mu$ M]
 <p><b>'7021</b></p>	1	 <p><b>Analog_12</b></p>	22
 <p><b>'6690</b></p>	2	 <p><b>Analog_20</b></p>	22
 <p><b>'7013</b></p>	5	 <p><b>'9037</b></p>	23
 <p><b>Analog_18</b></p>	6	 <p><b>'4202</b></p>	24
 <p><b>'8252</b></p>	6	 <p><b>'9125</b></p>	25
 <p><b>'6117</b></p>	7	 <p><b>'7356</b></p>	26
 <p><b>Analog_19</b></p>	9	 <p><b>'8474</b></p>	27

Compound	IC50 [ $\mu$ M]	Compound	IC50 [ $\mu$ M]
 <p><b>'4217</b></p>	9	 <p><b>'0481</b></p>	27
 <p><b>'9033</b></p>	9	 <p><b>Analog_7</b></p>	28
 <p><b>'9122</b></p>	9	 <p><b>Analog_8</b></p>	33
 <p><b>Analog_17</b></p>	11	 <p><b>Analog_10</b></p>	34
 <p><b>Analog_6</b></p>	12	 <p><b>Analog_11</b></p>	35
 <p><b>Analog_49</b></p>	13	 <p><b>'9009</b></p>	38
 <p><b>'0156</b></p>	14	 <p><b>'3251</b></p>	40

Compound	IC50 [ $\mu$ M]	Compound	IC50 [ $\mu$ M]
 <p><b>'9028</b></p>	14	 <p><b>'3250</b></p>	40
 <p><b>'4221</b></p>	19	 <p><b>Analog_21</b></p>	43
 <p><b>'3264</b></p>	20	 <p><b>'3252</b></p>	45
 <p><b>'4212</b></p>	20	 <p><b>'3564</b></p>	48

**Table 4.S2. Antiviral activities.**

Compound	Antiviral IC <sub>50</sub> <sup>a</sup> [μM]	Antiviral IC <sub>90</sub> <sup>a</sup> [μM]	CC <sub>10</sub> <sup>b</sup> [μM]
<b>'7021</b>	6.2	7.3	8.03
<b>'7356</b>	19.53	19.9	6.93

<sup>a</sup> Efficacy in RT-qPCR viral infectivity assay in X cells.

<sup>b</sup> Reduction in cell viability (cytotoxicity).

**Table 4.S3. Pan-viral activities of '7021.**

Viral MPro	IC50 [ $\mu$ M]
MERS	50
SARS-CoV-1	8
SARS-CoV-2	1

**Table 4.S4. Crystallographic statistics.**

Ligand	'5548	'7356	'6111	SG-0001	'9121	'7021	'8252	'9218
PDB ID	8DIG	8DIF	8DIH	8DII	8DIC	8DIB	8DID	8DIE
Resolution range (Å)	72.62 - 2.45 (2.538 - 2.45)	48.44 - 1.98 (2.051 - 1.98)	44.39 - 2.12 (2.196 - 2.12)	48.6 - 2.59 (2.683 - 2.59)	38.13 - 2.09 (2.165 - 2.09)	38.73 - 2.17 (2.248 - 2.17)	38.22 - 1.95 (2.02 - 1.95)	38.64 - 1.9 (1.968 - 1.9)
Space group	P 21 21 21	P 1 21 1	C 1 2 1	I 1 2 1	P 1 21 1	P 1 21 1	P 1 21 1	P 1 21 1
a, b, c (Å)	67.779, 102.704, 102.704	44.897, 53.624, 115.04	114.102, 53.917, 45.258	45.041, 53.835, 115.11	44.346, 53.814, 115.824	44.911, 53.72, 115.527	44.238, 53.666, 115.632	44.923, 53.751, 115.119
a, b, c (°)	90, 90, 90	90, 101.151, 90	90, 101.253, 90	90, 101.13, 90	90, 100.459, 90	90, 100.833, 90	90, 100.938, 90	90, 101.062, 90
Total reflections	52786 (4222)	74588 (6993)	29895 (2343)	17080 (1724)	63019 (6332)	56799 (5689)	76834 (7688)	83998 (8382)
Unique reflections	26433 (2151)	37368 (3516)	14998 (1188)	8552 (864)	31556 (3170)	28435 (2848)	38463 (3848)	42036 (4194)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	97.69 (80.05)	99.03 (92.87)	96.99 (77.24)	99.74 (99.77)	98.17 (98.84)	98.18 (98.41)	98.14 (98.54)	94.67 (79.45)
Mean I/sigma(I)	10.82 (1.33)	13.42 (1.00)	19.51 (1.83)	8.81 (0.84)	8.94 (0.98)	8.68 (0.89)	7.58 (0.98)	4.57 (0.59)
Wilson B-factor	44.04	36.71	48.65	56.27	39.01	40.42	30.04	33.18
R-merge	0.05895 (0.6349)	0.0362 (0.8374)	0.01881 (0.4432)	0.06999 (0.9217)	0.05198 (0.7831)	0.05847 (0.8779)	0.05816 (0.7736)	0.1022 (1.824)
R-meas	0.08337 (0.8979)	0.0512 (1.184)	0.0266 (0.6267)	0.09899 (1.304)	0.07351 (1.107)	0.08269 (1.242)	0.08226 (1.094)	0.1446 (2.579)
R-pim	0.05895 (0.6349)	0.0362 (0.8374)	0.01881 (0.4432)	0.06999 (0.9217)	0.05198 (0.7831)	0.05847 (0.8779)	0.05816 (0.7736)	0.1022 (1.824)
CC1/2	0.997 (0.513)	1 (0.41)	1 (0.625)	0.995 (0.333)	0.999 (0.474)	0.998 (0.397)	0.998 (0.513)	0.993 (0.201)
CC*	0.999 (0.823)	1 (0.763)	1 (0.877)	0.999 (0.707)	1 (0.802)	1 (0.754)	1 (0.823)	0.998 (0.579)
<b>Refinement</b>								
Reflections used in refinement	26397 (2118)	37312 (3502)	14996 (1188)	8548 (862)	31513 (3163)	28394 (2844)	38390 (3844)	40479 (3371)
Reflections used for R-free	1317 (116)	1835 (176)	723 (54)	419 (45)	1615 (173)	1341 (143)	1894 (224)	2128 (174)
R-work	0.2138 (0.3321)	0.2432 (0.3608)	0.2114 (0.3256)	0.2164 (0.3640)	0.2275 (0.3419)	0.2267 (0.3265)	0.2574 (0.3563)	0.2534 (0.3559)
R-free	0.2717 (0.3771)	0.2955 (0.3532)	0.2760 (0.3358)	0.2672 (0.3837)	0.2764 (0.3772)	0.2814 (0.3331)	0.2915 (0.4119)	0.3032 (0.3935)
CC(work)	0.954 (0.748)	0.957 (0.639)	0.964 (0.698)	0.957 (0.515)	0.959 (0.639)	0.962 (0.651)	0.954 (0.698)	0.953 (0.502)
CC(free)	0.922 (0.479)	0.953 (0.613)	0.963 (0.738)	0.947 (0.420)	0.932 (0.465)	0.955 (0.574)	0.948 (0.514)	0.916 (0.400)
Number of non-hydrogen atoms	4589	4365	2317	2347	4502	4579	4531	4604
macromolecules	4540	4269	2267	2310	4433	4506	4413	4509
ligands	25	42	25	25	38	42	38	42
solvent	24	54	25	12	31	31	80	53
Protein residues	601	585	303	303	599	599	591	605
RMS(bonds)	0.008	0.01	0.008	0.002	0.009	0.004	0.015	0.009
RMS(angles)	0.94	0.99	1.01	0.57	0.98	0.7	1.14	0.92
Ramachandran favored (%)	94.76	97.68	96.62	96.66	97.23	98.11	97.33	97.45
Ramachandran allowed (%)	5.24	2.32	3.38	3.34	2.6	1.72	2.49	2.38
Ramachandran outliers (%)	0	0	0	0	0.17	0.17	0.18	0.17
Rotamer outliers (%)	0	0.23	0	0	0.43	0	0.21	0

Ligand	'5548	'7356	'6111	SG-0001	'9121	'7021	'8252	'9218
PDB ID	8DIG	8DIF	8DIH	8DII	8DIC	8DIB	8DID	8DIE
<b>Refinement</b>								
Clashscore	7.5	3.99	4.71	5.48	6.71	3.84	8.22	5.1
Average B-factor	49.4	40.26	56.53	63.66	44.1	48.58	33.91	37.93
macromolecules	49.35	40.11	56.35	63.09	43.98	48.12	33.78	37.84
ligands	60.21	54.15	73.4	122.98	61.63	99.97	55.66	50.55
solvent	48.32	41.14	55.62	49.68	39.64	46.39	30.93	35.8

(One crystal for each structure)

\* Values in parentheses are for the highest-resolution shell.



## 4.7 Materials and Methods

### Expression and purification of M<sup>Pro</sup>

All reducing agents were added to buffers immediately before use. We received nsp5 cloned into pGEX6p-1 with a N-terminal GST tag and M<sup>Pro</sup> cleavage-site SAVLQ↓SGFRK and an HRV-3C-cleavable C-terminal 6xHis tag from Rolf Hilgenfeld. This bacterial expression plasmids were transformed into One Shot™ BL21(DE3)pLysS Chemically Competent *E. coli* (Thermo). The expression for M<sup>Pro</sup> in *E. coli* was modified from the expression previously described (39). In brief, a transformed clone of BL21(DE3)pLysS *E. coli* was added to a 50 mL culture of 2xYT media supplemented with 2% glucose and 100 µg/mL ampicillin grown overnight at 37°C. 30 mL of overnight culture were used to inoculate 1 L of 2xYT media supplemented with 100 µg/mL ampicillin. The inoculated culture was shaken at 225 rpm at 37°C and then induced when culture OD600 reached 0.8 (after ~3 h) by adding 1 mL of 1 M IPTG. After 5 h of expression at 37°C, the culture was centrifuged at 9,000 rpm for 15 min. Supernatant was discarded and cell pellet stored at -80°C. The frozen cell pellet was thawed on ice in 30 mL of 20 mM Tris 150 mM NaCl pH 7.4 buffer. The resuspended sample was sonicated until lysis was complete. Sonicated cell lysate was clarified by centrifugation at 15000 rpm for 30 min. 3 mL of Super Ni-NTA Affinity HP Resin beads (Protein Ark) equilibrated with wash buffer (20 mM TRIS 150 mM NaCl 20 mM imidazole) were incubated with 57 mL of the clarified supernatant for 1 h at 4°C. Beads were centrifuged at 200 rpm for 2 min and the supernatant was decanted. The Ni-NTA beads were washed with ~3 column volumes of wash buffer.. Hexahistidine tagged protein was eluted with 1 mL fractions of elution buffer (20 mM Tris 150 mM NaCl 350 mM Imidazole). Fractions containing protein were pooled and then buffer exchanged into 20% Glycerol 20 mM Tris 150 mM NaCl pH 7.4 using Amicon concentrators. 3C protease was added in a 5:1 ratio of M<sup>Pro</sup> to 3C protease and incubated

overnight at 4°C. A 2 L of culture yielded 2.28 mg of M<sup>Pro</sup> following 3C cleavage. 3C protease and 6xHis-tag were removed by incubation with Ni-NTA beads. The active dimer was isolated with a MonoQ™ 5/50 GL column (GE Healthcare). MonoQ column equilibrated with buffer A (20 mM Tris 1 mM DTT pH 8) and the protein sample was then loaded onto the column and eluted with a linear gradient of buffer B (1 M NaCl 20 mM Tris 1 mM DTT pH 8) 0 mM to 500 mM NaCl over 20 column volumes.

### **M<sup>Pro</sup> Inhibition assay**

A fluorescence-quenched substrate with the sequence H<sub>2</sub>N(d-Arg)(d-Arg)-K(MCA)-ATLQAIAS-K(DNP)-COOH was synthesized via the Fmoc solid-phase peptide synthesis as previously described (66). 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  $K_M = 12.7 \mu\text{M}$ , and M<sup>Pro</sup> 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. Compounds were incubated with protease prior to substrate addition at 37 °C for 1 h. After incubation, the substrate was added, and kinetic activity was monitored for 1 h at 37 °C. Initial velocities were calculated at 1 to 30 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. Kinetics measurements were carried out in triplicate. SARS-CoV-1 and MERS M<sup>Pro</sup> were both purchased from Bio-Techne (catalogue #: E-718-050 and E-719-050, respectively).  $K_M$  was derived with the NSP7 substrate for each protease (**Fig. 4.S9**), which was the substrate concentration used for each protease for comparative dose-response curves. Enzyme concentration was 50 nM for SARS-CoV-

1 and 100 nM for MERS. The same assay buffer described above was used for all kinetic assays with each protease.

### **Non-covalent molecular docking**

The protein template was modeled based on the crystal structure of the M<sup>Pro</sup> dimer in complex with a covalent alpha-ketoamide inhibitor (PDB 6Y2G) (39). All water molecules except for HOH 585 and HOH 602, which are located at the dimeric interface, were deleted. The binding pocket of the crystal structure's chain A was selected for docking. The alpha-ketoamide inhibitor was replaced by the non-covalent SARS-CoV inhibitor SID87915542 (41). Here, the SID87915542-bound M<sup>Pro</sup> crystal structure (PDB 3V3M) was aligned onto the SARS-CoV-2 M<sup>Pro</sup> crystal structure in order to project SID87915542 into the SARS-CoV-2 M<sup>Pro</sup> binding site. Next, the modeled protein-ligand complex and selected water molecules were prepared for docking with the protein prepwizard protocol of Maestro (Schrödinger, Inc. v. 2019-3) (67). Protons were added with Epik and protonation states were optimized with PropKa at pH 7. The C-terminus (Ser301) of each protein monomer structure was capped with N-methyl groups while the N-termini (Ser1) were positively charged. Subsequently, the modeled complex was energetically minimized using the OPLS3e force field. To better accommodate the modeled non-covalent ligand SID87915542, the CE atom of Met49 was displaced by 1.7Å from its initial position in the covalently ligated crystal structure (PDB 6Y2G).

Computational docking was performed using DOCK3.7 (68). Precomputed scoring grids for efficient quantification of van der Waals interaction between M<sup>Pro</sup> and docked molecules were generated with CHEMGRID (69). Using the AMBER united-atom partial charges (70),

electrostatic potentials within the binding pocket were computed following the numerical solution of the Poisson-Boltzmann equation with QNIFFT (71). The partial charges of the hydrogen at the epsilon nitrogen of His163, as well as the hydrogen atoms of the backbone amines of Gly143 and Glu166 were increased by 0.4 elementary charge units (e). In turn, the partial charges of oxygen atoms of the corresponding backbone carbonyl groups were decreased by 0.4e to maintain the initial net charge of each residue (42). The low dielectric protein environment was extended by 1.2 Å from the protein surface, as previously described (64). Similarly, the low dielectric boundary was extended by 0.7 Å from the protein surface for the calculation of ligand desolvation scoring grids with SOLVMAP (72). The atomic coordinates of SID87915542 (PDB 3V3M) (41), the alpha-ketamide inhibitor of the initial crystal structure (PDB 6Y2G) (39), BDBM512845 (PDB 4MDS) (73), as well as fragment hits MAT-POS-7dfc56d9-1 (M<sup>Pro</sup>-x0161) (18) and AAR-POS-d2a4d1df-5 (M<sup>Pro</sup>-x0305) (18) obtained from the Covid-19 Moonshot screening efforts, were used to generate 80 matching spheres (68) for ligand placement in the docking calculations.

The obtained docking parameters were evaluated based on their ability to prioritize 34 previously reported ligands of SARS-CoV M<sup>Pro</sup> obtained from the ChEMBL database (74), against a background of 1,805 property matched decoys generated with the DUDE-Z approach (48). In addition, an ‘Extrema’ set (27,42) of 194,921 molecules, including compounds with net-charges ranging from -2 to +2, was screened against the docking model in order to assess the parameters’ ability to prioritize neutral molecules.

Using the ZINC15 database (43), 225,327,212 neutral molecules mainly from the lead-like chemical space, i.e. molecular weight (MWT) between 250 and 350 amu and calculated (c)logP ≤

4.5, from the make-on-demand compound libraries from Enamine Ltd. and WuXi Appetec. (Shanghai, China), were screened. Thereby, 219,305,079 molecules were successfully scored with each molecule sampling on average 3,588 orientations and 425 conformations which resulted in the evaluation of approximately 148 trillion complexes in roughly 70 h on a 1,000-core computer cluster. In addition, 110,898,461 molecules with  $350 < \text{MWT} \leq 500$  and  $\text{clogP} \leq 4.5$  from ZINC15 were screened in a separate docking campaign. 107,486,710 compounds were successfully scored, each exploring on average 4,175 orientations and 540 conformations within the binding pocket. Nearly 90 trillion complexes were scored in roughly 45 h using a 1,000-core cluster.

From each docking screen, the predicted binding poses of the 500,000 top-ranked molecules were analyzed for internal molecular strain (44). Molecules that passed the strain criteria (total strain  $< 6.5$  TEU; maximum single torsion  $< 1.8$  TEU), were judged by their ability to form hydrogen bonds with Gly143, His163 (S1 subpocket) or Glu166 and proximity to residues forming the S2 subpocket such as Met49 or Asp187. Finally, 120 compounds, selected from the lead-like docking screen, were ordered from Enamine Ltd., of which 105 were successfully synthesized (87.5%) in addition to 100 molecules of larger MWT that were ordered from the second docking screen, 89 of which were successfully synthesized by Enamine Ltd.

A second docking campaign for non-covalent inhibitors was performed against the crystal structure of M<sup>Pro</sup> in complex with MAT-POS-b3e365b9-1 (M<sup>Pro</sup>-x11612) (18) from the Covid-19 Moonshot consortium. All water molecules except HOH6 and HOH300 were removed and the protein-ligand complex structure was prepared for docking following the protein prepwizard protocol of Maestro (Schrödinger v. 2019-3) as described above.

As described above in the previous docking campaign, the partial charges of the hydrogen atoms at the epsilon nitrogen of His163 and the backbone amine of Glu166 were increased by 0.4e, whereas the partial charges of corresponding backbone carbonyl oxygen atoms were decreased by 0.4e to maintain the net charge of each residue. For calculating electrostatic scoring grids, the low-dielectric volume of the protein was extended by 1.9 Å from the protein surface (based on surface mapping spheres generated by Sphgen). In addition, the low dielectric boundary was extended by 1.0 Å from the protein surface for calculating ligand desolvation scoring grids with SOLVMAP. The atomic coordinates of MAT-POS-b3e365b9-1 were used to generate 45 matching spheres for ligand placement with DOCK3.8. The performance of the obtained docking grids was evaluated by their ability to enrich 15 previously reported SARS-CoV-2 M<sup>Pro</sup> inhibitors over 650 property-matched decoys or an Extrema set containing 153,256 molecules with net charges ranging from -2 to +2, molecular weight between 300 and 500 amu. Finally, 862,382,088 neutral compounds with 18-29 heavy atoms from the Enamine REAL chemical library were screened using the ZINC22 database (<http://files.docking.org/zinc22/>). Molecules with strained conformations (total strain > 8 TEU, maximum single strain > 3 TEU), were excluded by the docking program. 778,517,250 molecules were successfully scored, each sampled in approximately 836 conformations and 3,439 orientations, leading to the evaluation of roughly 905.8 trillion complexes within 481h on a 1000-core computer cluster.

21,284,498 compounds scored lower than -35 kcal/mol and the poses of top scoring 5,004,192 compounds were extracted. 214,580 compounds formed favorable interactions with key residues such as His163, Glu166 and the P2 subpocket, 181,866 of which obtained ECFP4-based TC

coefficients of less than 0.35 to the 1,716 known SARS-CoV and SARS-CoV-2 M<sup>Pro</sup> inhibitors reported in the literature (2-4,7,9,10,12,13,15,18,21,39,41,75-90). Finally, roughly 9,000 top-ranking compounds were visually inspected, and 167 molecules were ordered from Enamine Ltd., 146 of which (87.4%) were successfully synthesized.

### Covalent molecular docking

Cysteine-reactive warheads of aldehydes, nitriles, and alpha-ketoamides were searched in the ZINC20/Enamine *REAL* databases of 1.4 billion molecules using their respective SMARTS patterns (ketoamides O=[CR0]([#6])[CR0](=O)N[#6]; aldehydes [CX3H1](=O)[#6]; nitriles [CX4]-C#N). This returned 25.7 million nitriles, 2.5 million aldehydes, and 1.5 million ketoamides. Molecules were filtered to have at least one ring, and to be fragment to lead-like molecular weights (<350). Three-dimensional “dockable” conformations were generated with molecules in their transition-state form and a dummy atom in place for the covalent docking algorithm to indicate which atom should be modeled covalently bound to the Cysteine sulfur (58,59). Overall, 6.5 million molecules were docked – 3.6 million nitriles, 1.4 million ketoamides, and 1.5 million aldehydes.

The protein was prepared in DOCK3.7 (68). Pose reproduction of the truncated covalent molecule of PDB 6Y2G (39) (smiles of dockable ligand: O=C1NCC[C@H]1CC[C@]([SiH3])(C(NCC2=CC=CC=C2)=O)O) was checked for the docking setup. Default generated grids were used for electrostatic (radius size 1.9Å) and VDW scoring, and no matching spheres were used in docking calculations as they are not used by the covalent docking DOCKoalent (58,59) algorithm. For covalent docking, the Cys145 SH group was

indicated as the anchor for molecules screened. The distance was slightly relaxed from the C-C bond distance to 1.85Å. For His41 protonation, aldehydes, nitriles, and neutral ketoamides used HID, while negative ketoamides used HIP. Each warhead was docked separately with a total 6.5 million molecules screened. Accordingly, each warhead was also processed separately.

For the aldehydes, the top 300,000 ranked molecules were evaluated for torsional strain (44), and those with a total torsional strain greater than 9.8 (around 3.7 incurred due to strain on atom types on the warhead and this was disregarded, therefore total energy was 6) and single torsional strain greater than 2.5 were excluded (155,386 left). Molecules making more than 1 hydrogen bond to the protein, having no hydrogen bond clashes, no unpaired hydrogen bond donors (56,969 left) were prioritized. Remaining molecules were clustered for chemical similarity based on ECFP4-based Tanimoto coefficient (Tc) of 0.5. Viable poses filling the S1', S1 or S2 sites were selected during visual inspection. A total of 35 aldehydes were selected for make-on-demand synthesis of which 27 were successfully synthesized. For the nitriles, the top 100,000 ranked molecules were evaluated for torsional strain (17,424 left), then filtered for favorable interactions (6,201 left). Lastly, we visually inspected remaining molecules for favorable hydrogen bonds formed with His41, Gly143, Thr26, Glu166, or Cys145. Finally, 41 compounds were ordered for synthesis (31 were successfully obtained). For the ketoamides the top 393,000 ranked molecules with scores less than 0.0 were evaluated for torsional strain (121,234 left), and favorable interactions with the enzyme (37,267 remained). Visual inspection focused on those making hydrogen bonds with His41, Cys145, Gly143, Thr26. In total 21 molecules were prioritized and 16 were successfully synthesized.



### **Make-on-demand synthesis**

Non-covalent and covalent compounds purchased from docking screens, as well as analogs, were synthesized by Enamine Ltd. ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Purities of molecules were at least 90% and most active compounds were at least 95% (based on LC/MS data) ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)).

### **Compound optimization**

Optimization of docking hits **ZINC346371112**, **ZINC301553312**, **ZINC813360541**, **ZINC553840273**, **'3620**, **'0431**, **'4589**, **'5103**, **'5156**, **'6246**, **'6792**, **'0292**, **'2826/'0892** were attempted ([https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Analogs were designed for desired chemical perturbations or searched in SmallWorld and Arthor catalogs and synthesized by Enamine Ltd. For **'3620**, compounds were also designed from the **'7356** and **'7021** crystal structures and were modeled with covalent docking or with Maestro (v. 2021-2, Schrödinger, LLC) ligand alignment.

### **Protein crystallization**

Both covalent and non-covalent compounds including **7021**, **'9121**, **8252**, **'9218**, **7356**, **5548**, **6111** and **SG-0001** were co-crystallized with SARS-CoV2 M<sup>Pro</sup>. Before setting up crystals, 10 mg/ml of protein was incubated with either 0.3 mM of covalent compounds or 1.5 mM of non-covalent compounds on ice for 1 h. Crystals were set using vapor diffusion hanging drop method at 20 °C in conditions including 0.1 M Tris pH 7.4 and 20% PEG 8000; and 0.1 M MES pH 6.5, 20% PEG 6000. Crystals took 3-4 days to grow for all compounds. Before data collection, crystals were cryo-cooled in a solution containing reservoir solution and 25% glycerol.

## Structure determination and refinement

The M<sup>Pro</sup> inhibitor compound datasets were either collected at the Advanced Light Source beamline 8.3.1 (Lawrence Berkeley laboratory) or SSRL beamline 12-2 beamline (Stanford, United States) at a temperature of 100K. The diffraction datasets were processed using XDS (91) and CCP4 software's suite (92). AIMLESS (93) was used for scaling and merging. Molecular replacement was performed either using PHASER (94) using the protein model from PDB entry 7NG3 (95) as the search model. The bound ligand in the PDB 7NG3 was removed from the search model during molecular replacement, giving unbiased electron density for ligands in the initial electron density maps. The initial model fitting and addition of waters was done in COOT (96) followed by refinement in REFMAC (97). Geometry restraints for the ligands were created in eLBOW-PHENIX (98) and following rounds of refinement were carried out in PHENIX. Geometry for each structure was assessed using Molprobdity and PHENIX polygon. Datasets have been deposited to the PDB with PDB IDs 8DIB, 8DIC, 8DID, 8DIE, 8DIF, 8DIG, 8DIH and 8DII. Statistics for data collection and refinement are in **Table 4.S5**. The ligand symmetry accounted RMSDs between the docked pose and experimental pose were calculated by the Hungarian algorithm in DOCK6 (99).

## Antiviral and cytotoxicity assays

Two thousand (2,000) HeLa-ACE2 cells were seeded into 96-well plates and incubated for 24 h. 2 h before infection, the medium was replaced with a new media containing the compound of interest, including a DMSO control. Plates were then transferred into the biosafety level 3 (BSL-3) facility and 1,000 PFU (MOI = 0.25) of SARS-CoV-2 was added, bringing the final compound concentration to those indicated. SARS-CoV-2/WA1 variant was used as indicated. Plates were

then incubated for 48 h. Infectivity was measured by the accumulation of viral NP protein in the nucleus of the HeLa-ACE2 cells (fluorescence accumulation). Percent infection was quantified as  $((\text{Infected cells}/\text{Total cells}) - \text{Background}) \times 100$ , and the DMSO control was then set to 100% infection for analysis. Cytotoxicity was also performed at matched concentrations using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected HeLa-ACE2 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

### **Dynamic Light Scattering (DLS)**

Samples were prepared in filtered 50 mM KPi buffer pH 7.0 with final DMSO concentration at 1% (v/v). Colloidal particle formation was detected using DynaPro Plate Reader II (Wyatt Technologies). All compounds were screened in triplicate at roughly 2-fold higher concentration than reported  $IC_{50}$  (concentrations can be found at [https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Analysis was performed with GraphPad Prism software version 9.1.1 (San Diego, CA).

### **Enzyme Inhibition Assays for Aggregation**

Enzyme inhibition assays 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 4 nM AmpC  $\beta$ -lactamase (AmpC) or Malate dehydrogenase (MDH) for 5 min. AmpC reactions were initiated by the addition of 50  $\mu$ M CENTA chromogenic substrate. The change in absorbance was monitored at 405 nm for 1 min 45 sec. MDH reactions were initiated by the addition of 200  $\mu$ M nicotinamide adenine dinucleotide

(NADH) (54839, Sigma Aldrich) and 200  $\mu$ M oxaloacetic acid (324427, Sigma Aldrich). The change in absorbance was monitored at 340 nm for 1 min 45 sec. Initial rates were divided by the DMSO control rate to determine % enzyme activity. Each compound was screened at roughly 2-fold higher concentration than reported the IC<sub>50</sub> value in triplicate (concentrations can be found at [https://github.com/efink14/Fink\\_2022\\_Dissertation](https://github.com/efink14/Fink_2022_Dissertation)). Data was analyzed using GraphPad Prism software version 9.1.1 (San Diego, CA). For detergent reversibility experiments, inhibition was screened near IC<sub>75</sub> with or without 0.01% (v/v) Triton X-100 in triplicate. Enzymatic progress was performed/monitored as previously described (45,100).

### **Statistical analyses**

All statistical analyses were performed on the GraphPad Prism version 8.0 or 9.1.1 software. Changes only at the 95% confidence level ( $P < 0.05$ ) were considered as statistically significant.

### **Data availability.**

All crystallographic structures have been deposited in the PDB as 8DIB ('7021), 8DIC ('9121), 8DID (8252), 8DIE ('9218), 8DIF ('7356), 8DIG ('5548), 8DIH ('6111), 8DII (SG-0001). The identities of compounds docked in non-covalent screens can be found at ZINC15/ZINC20 (<http://zinc15.docking.org> and <http://zinc20.docking.org>) and ZINC22 (<http://files.docking.org/zinc22/>). The covalent compounds have been deposited in <http://covalent2022.docking.org> along with their DOCKoValent files. Active compounds may be purchased from Enamine Ltd. All other data are available from the corresponding authors on request.

**Code availability**

DOCK3.7 and DOCK3.8 are freely available for non-commercial research from the authors; commercial licenses are available via the UC Regents. An open-source web-based version of the program is available without restriction to all (<https://blaster.docking.org>), as are the Arthor and Small World analoging tools used in this study.

## 4.6 Acknowledgements

This work was supported by DARPA grant HR0011-19-2-0020 (B.K.S., J.J.I., A.G.-S.) and by NIH grant R35GM122481 (B.K.S., J.J.I.). This was also funded by NIAID grant U19AI171110 (C.S.C., B.K.S., A.G.-S., A.R.R.) run by Principal Investigator Nevan Krogan. This work was also partly funded by CRIPT (Center for Research on Influenza Pathogenesis and Transmission), an NIAID funded Center of Excellence for Influenza Research and Response (CEIRR, contract #75N93021C00014) and by supplements to DoD grant W81XWH-20-1-0270 and to NIAID grant U19AI135972 to A.G.-S. B.A. received the Covid Catalyst Award from the Center for Emerging and Neglected Diseases (CEND). We gratefully acknowledge OpenEye Software for Omega and related tools, and Schrodinger LLC for the Maestro package. We also thank Dr. Randy Albrecht for support with the BSL3 facility and procedures at the ISMMS as well as Richard Cadagan and Daniel Flores for excellent technical assistance. In addition, we thank Dr. Rolf Hilgenfeld for providing the SARS-CoV-2 M<sup>Pro</sup> plasmid. We acknowledge the contributions of the UCSF Chemical Underpinnings of Biological Systems (CUBS) 2021 cohort which included Siyi Wang, Isabel Lee, Vineet Mathur, Sham Rampersaud, Luis Santiago, Sara Warrington, and Rose Yang.

#### **4.7 Author contributions.**

S.G. conducted non-covalent docking screens and compound optimization, with input from B.K.S., J.L., S.V., and the 2021 CUBS Cohort. E.A.F., S.Gu, and X.W. performed covalent database building with input from J.T. and B.K.S., covalent docking with input from B.K.S., and compound optimization with input from B.K.S, A.R.R., S.G., and I.S. Enzymatic testing was conducted by C.B., assisted by N.J.Y., and supervised by C.S.C. Antiviral and cytotoxicity assays were performed by K.W., with supervision by A.G.-S. Protein purification was done by I.S., B.A., P.F., with supervision by B.K.S., C.S.C., and A.O. Crystallography was done by I.S. assisted by J.O. and with input from B.K.S. Aggregation testing was performed by I.G. and H.O. with input from B.K.S. ZINC15 and ZINC22 databases were built by J.J.I. Y.S.M. supervised compound synthesis of Enamine compounds, assisted by I.S.K. C.S.C. and B.K.S. supervised the project. E.A.F., S.G., and C.B. wrote the paper with input from all other authors, and primary editing from C.S.C. and B.K.S. C.S.C. and B.K.S. conceived the project.

## 4.8 References

1. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59* (14), 6595–6628. <https://doi.org/10.1021/acs.jmedchem.5b01461>.
2. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>.
3. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Proteinase (3CL<sup>pro</sup>) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003**, *300* (5626), 1763–1767. <https://doi.org/10.1126/science.1085658>.
4. Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R.  $\alpha$ -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63* (9), 4562–4578. <https://doi.org/10.1021/acs.jmedchem.9b01828>.
5. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Wang, X.; Jochmans, D.; Neyts, J.; Młynarski, W.; Hilgenfeld, R.; Drag, M. SARS-CoV-2 Mpro Inhibitors and Activity-Based Probes for Patient-Sample Imaging. *Nat. Chem. Biol.* **2021**, *17* (2), 222–228. <https://doi.org/10.1038/s41589-020-00689-z>.



6. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.; Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Steppan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. An Oral SARS-CoV-2 Mpro Inhibitor Clinical Candidate for the Treatment of COVID-19. *Science* **2021**, *374* (6575), 1586–1593. <https://doi.org/10.1126/science.abl4784>.
7. Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y.; Lockner, J. W.; Lou, J.; Marra, M. T.; Mitchell, L. J.; Murray, B. W.; Nieman, J. A.; Noell, S.; Planken, S. P.; Rowe, T.; Ryan, K.; Smith, G. J.; Solowiej, J. E.; Steppan, C. M.; Taggart, B. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63* (21), 12725–12747. <https://doi.org/10.1021/acs.jmedchem.0c01063>.
8. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T.; Kouki, K.; Nakahashi, A.; Kawashima, S.; Sanaki, T.; Toba, S.; Uemura, K.; Mizutare, T.; Ando, S.; Sasaki, M.; Orba, Y.; Sawa, H.; Sato, A.; Sato, T.; Kato, T.; Tachibana, Y. Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022**, *65* (9), 6499–6512. <https://doi.org/10.1021/acs.jmedchem.2c00117>.
9. Zhang, C.-H.; Spasov, K. A.; Reilly, R. A.; Hollander, K.; Stone, E. A.; Ippolito, J. A.; Liosi, M.-E.; Deshmukh, M. G.; Tirado-Rives, J.; Zhang, S.; Liang, Z.; Miller, S. J.; Isaacs, F.;

- Lindenbach, B. D.; Anderson, K. S.; Jorgensen, W. L. Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med. Chem. Lett.* **2021**, *12* (8), 1325–1332. <https://doi.org/10.1021/acsmchemlett.1c00326>.
10. Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.; Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N.; Liang, Z.; Isaacs, F.; Lindenbach, B.; Miller, S. J.; Anderson, K. S.; Jorgensen, W. L. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7* (3), 467–475. <https://doi.org/10.1021/acscentsci.1c00039>.
11. Ge, R.; Shen, Z.; Yin, J.; Chen, W.; Zhang, Q.; An, Y.; Tang, D.; Satz, A. L.; Su, W.; Kuai, L. Discovery of SARS-CoV-2 Main Protease Covalent Inhibitors from a DNA-Encoded Library Selection. *SLAS Discov.* **2022**, *27* (2), 79–85. <https://doi.org/10.1016/j.slasd.2022.01.001>.
12. Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368* (6497), 1331–1335. <https://doi.org/10.1126/science.abb4489>.
13. Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Boceprevir, GC-376, and Calpain Inhibitors II, XII Inhibit SARS-CoV-2 Viral Replication by Targeting the Viral Main Protease. *Cell Res.* **2020**, *30* (8), 678–692. <https://doi.org/10.1038/s41422-020-0356-z>.
14. Moon, P.; Boike, L.; Dovala, D.; Henning, N. J.; Knapp, M.; Spradlin, J. N.; Ward, C. C.; Wolleb, H.; Zammit, C. M.; Fuller, D.; Blake, G.; Murphy, J. P.; Wang, F.; Lu, Y.; Moquin,

- S. A.; Tandeske, L.; Hesse, M. J.; McKenna, J. M.; Tallarico, J. A.; Schirle, M.; Toste, F. D.; Nomura, D. K. *Discovery of Potent Pyrazoline-Based Covalent SARS-CoV-2 Main Protease Inhibitors*; preprint; *Biochemistry*, 2022. <https://doi.org/10.1101/2022.03.05.483025>.
15. Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H.; Huang, C.; Wang, Y.-F.; Zhang, J.; Quan, B.; Shen, C.; Mao, X.; Liu, X.; Sun, W.; Yang, W.; Ni, X.; Wang, K.; Xu, L.; Duan, Z.-L.; Zou, Q.-C.; Zhang, H.-L.; Qu, W.; Long, Y.-H.-P.; Li, M.-H.; Yang, R.-C.; Liu, X.; You, J.; Zhou, Y.; Yao, R.; Li, W.-P.; Liu, J.-M.; Chen, P.; Liu, Y.; Lin, G.-F.; Yang, X.; Zou, J.; Li, L.; Hu, Y.; Lu, G.-W.; Li, W.-M.; Wei, Y.-Q.; Zheng, Y.-T.; Lei, J.; Yang, S. SARS-CoV-2 Mpro Inhibitors with Antiviral Activity in a Transgenic Mouse Model. *Science* **2021**, *371* (6536), 1374–1378. <https://doi.org/10.1126/science.abf1611>.
16. Lu, J.; Chen, S. A.; Khan, M. B.; Brassard, R.; Arutyunova, E.; Lamer, T.; Vuong, W.; Fischer, C.; Young, H. S.; Vederas, J. C.; Lemieux, M. J. Crystallization of Feline Coronavirus Mpro With GC376 Reveals Mechanism of Inhibition. *Front. Chem.* **2022**, *10*, 852210. <https://doi.org/10.3389/fchem.2022.852210>.
17. Vuong, W.; Fischer, C.; Khan, M. B.; van Belkum, M. J.; Lamer, T.; Willoughby, K. D.; Lu, J.; Arutyunova, E.; Joyce, M. A.; Saffran, H. A.; Shields, J. A.; Young, H. S.; Nieman, J. A.; Tyrrell, D. L.; Lemieux, M. J.; Vederas, J. C. Improved SARS-CoV-2 Mpro Inhibitors Based on Feline Antiviral Drug GC376: Structural Enhancements, Increased Solubility, and Micellar Studies. *Eur. J. Med. Chem.* **2021**, *222*, 113584. <https://doi.org/10.1016/j.ejmech.2021.113584>.
18. The COVID Moonshot Consortium; Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.; Bennett, J.; Bilenko, V. A.; Bilenko, V. A.; Boby, M. L.; Borden, B.; Bowman, G. R.; Brun, J.; Bvnbs, S.; Calmiano, M.; Carbery, A.; Carney, D.; Cattermole, E.; Chang, E.;

Chernyshenko, E.; Chodera, J. D.; Clyde, A.; Coffland, J. E.; Cohen, G.; Cole, J.; Contini, A.; Cox, L.; Cvitkovic, M.; Dias, A.; Donckers, K.; Dotson, D. L.; Douangamath, A.; Duberstein, S.; Dudgeon, T.; Dunnett, L.; Eastman, P. K.; Erez, N.; Eyermann, C. J.; Fairhead, M.; Fate, G.; Fearon, D.; Fedorov, O.; Ferla, M.; Fernandes, R. S.; Ferrins, L.; Foster, R.; Foster, H.; Gabizon, R.; Garcia-Sastre, A.; Gawriljuk, V. O.; Gehrtz, P.; Gileadi, C.; Giroud, C.; Glass, W. G.; Glen, R.; Glinert, I.; Godoy, A. S.; Gorichko, M.; Gorrie-Stone, T.; Griffen, E. J.; Hart, S. H.; Heer, J.; Henry, M.; Hill, M.; Horrell, S.; Huliak, V. D.; Hurley, M. F. D.; Israely, T.; Jajack, A.; Jansen, J.; Jnoff, E.; Jochmans, D.; John, T.; Jonghe, S. D.; Kantsadi, A. L.; Kenny, P. W.; Kiappes, J. L.; Kinakh, S. O.; Koekemoer, L.; Kovar, B.; Krojer, T.; Lee, A.; Lefker, B. A.; Levy, H.; Logvinenko, I. G.; London, N.; Lukacik, P.; Macdonald, H. B.; MacLean, B.; Malla, T. R.; Matviuk, T.; McCorkindale, W.; McGovern, B. L.; Melamed, S.; Melnykov, K. P.; Michurin, O.; Mikolajek, H.; Milne, B. F.; Morris, A.; Morris, G. M.; Morwitzer, M. J.; Moustakas, D.; Nakamura, A. M.; Neto, J. B.; Neyts, J.; Nguyen, L.; Noske, G. D.; Oleinikovas, V.; Oliva, G.; Overheul, G. J.; Owen, D.; Pai, R.; Pan, J.; Paran, N.; Perry, B.; Pingle, M.; Pinjari, J.; Politi, B.; Powell, A.; Psenak, V.; Puni, R.; Rangel, V. L.; Reddi, R. N.; Reid, S. P.; Resnick, E.; Ripka, E. G.; Robinson, M. C.; Robinson, R. P.; Rodriguez-Guerra, J.; Rosales, R.; Rufa, D.; Saar, K.; Saikatendu, K. S.; Schofield, C.; Shafeev, M.; Shaikh, A.; Shi, J.; Shurrush, K.; Singh, S.; Sittner, A.; Skyner, R.; Smalley, A.; Smeets, B.; Smilova, M. D.; Solmesky, L. J.; Spencer, J.; Strain-Damerell, C.; Swamy, V.; Tamir, H.; Tennant, R.; Thompson, W.; Thompson, A.; Tomasio, S.; Tsurupa, I. S.; Tumber, A.; Vakonakis, I.; van Rij, R. P.; Vangeel, L.; Varghese, F. S.; Vaschetto, M.; Vitner, E. B.; Voelz, V.; Volkamer, A.; von Delft, F.; von Delft, A.; Walsh, M.; Ward, W.; Weatherall, C.; Weiss, S.; White, K. M.; Wild, C. F.; Wittmann, M.; Wright, N.; Yahalom-Ronen, Y.; Zaidmann, D.; Zidane, H.;

- Zitzmann, N. *Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitor Therapeutics*; preprint; Biochemistry, 2020. <https://doi.org/10.1101/2020.10.29.339317>.
19. Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing Drug Discovery for Pandemics. *Nat. Chem.* **2020**, *12* (7), 581–581. <https://doi.org/10.1038/s41557-020-0496-2>.
20. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* **2020**, *11* (1), 5047. <https://doi.org/10.1038/s41467-020-18709-w>.
21. Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Nekhotiaeva, N.; Vangeel, L.; De Jonghe, S.; Jochmans, D.; Krambrich, J.; Tas, A.; Lundgren, B.; Gravenfors, Y.; Craig, A. J.; Atilaw, Y.; Sandström, A.; Moodie, L. W. K.; Lundkvist, Å.; van Hemert, M. J.; Neyts, J.; Lennerstrand, J.; Kihlberg, J.; Sandberg, K.; Danielson, U. H.; Carlsson, J. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J. Am. Chem. Soc.* **2022**, *144* (7), 2905–2920. <https://doi.org/10.1021/jacs.1c08402>.
22. Flynn, J. M.; Samant, N.; Schneider-Nachum, G.; Bakan, D. T.; Yilmaz, N. K.; Schiffer, C. A.; Moquin, S. A.; Dovala, D.; Bolon, D. N. A. Comprehensive Fitness Landscape of SARS-CoV-2 Mpro Reveals Insights into Viral Resistance Mechanisms. *eLife* **2022**, *11*, e77433. <https://doi.org/10.7554/eLife.77433>.

23. Shaqra, A. M.; Zvornicanin, S. N.; Huang, Q. Y. J.; Lockbaum, G. J.; Knapp, M.; Tandeske, L.; Bakan, D. T.; Flynn, J.; Bolon, D. N. A.; Moquin, S.; Dovala, D.; Kurt Yilmaz, N.; Schiffer, C. A. Defining the Substrate Envelope of SARS-CoV-2 Main Protease to Predict and Avoid Drug Resistance. *Nat. Commun.* **2022**, *13* (1), 3556. <https://doi.org/10.1038/s41467-022-31210-w>.
24. Manglik, A.; Lin, H.; Aryal, D. K.; McCorvy, J. D.; Dengler, D.; Corder, G.; Levit, A.; Kling, R. C.; Bernat, V.; Hübner, H.; Huang, X.-P.; Sassano, M. F.; Giguère, P. M.; Löber, S.; Da Duan; Scherrer, G.; Kobilka, B. K.; Gmeiner, P.; Roth, B. L.; Shoichet, B. K. Structure-Based Discovery of Opioid Analgesics with Reduced Side Effects. *Nature* **2016**, *537* (7619), 185–190. <https://doi.org/10.1038/nature19112>.
35. Singh, I.; Seth, A.; Billesbølle, C. B.; Braz, J.; Rodriguiz, R. M.; Roy, K.; Bekele, B.; Craik, V.; Huang, X.-P.; Boytsov, D.; Lak, P.; O'Donnell, H.; Sandtner, W.; Roth, B. L.; Basbaum, A. I.; Wetsel, W. C.; Manglik, A.; Shoichet, B. K.; Rudnick, G. *Structure-Based Discovery of Conformationally Selective Inhibitors of the Serotonin Transporter*; preprint; *Biochemistry*, 2022. <https://doi.org/10.1101/2022.06.13.495991>.
26. Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Alga, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224–229. <https://doi.org/10.1038/s41586-019-0917-9>.
27. Stein, R. M.; Kang, H. J.; McCorvy, J. D.; Glatfelter, G. C.; Jones, A. J.; Che, T.; Slocum, S.; Huang, X.-P.; Savych, O.; Moroz, Y. S.; Stauch, B.; Johansson, L. C.; Cherezov, V.; Kenakin, T.; Irwin, J. J.; Shoichet, B. K.; Roth, B. L.; Dubocovich, M. L. Virtual Discovery of Melatonin

- Receptor Ligands to Modulate Circadian Rhythms. *Nature* **2020**, *579* (7800), 609–614. <https://doi.org/10.1038/s41586-020-2027-0>.
28. Alon, A.; Lyu, J.; Braz, J. M.; Tummino, T. A.; Craik, V.; O’Meara, M. J.; Webb, C. M.; Radchenko, D. S.; Moroz, Y. S.; Huang, X.-P.; Liu, Y.; Roth, B. L.; Irwin, J. J.; Basbaum, A. I.; Shoichet, B. K.; Kruse, A. C. *Crystal Structures of the  $\sigma_2$  Receptor Template Large-Library Docking for Selective Chemotypes Active in Vivo*; preprint; Pharmacology and Toxicology, 2021. <https://doi.org/10.1101/2021.04.29.441652>.
29. Levit Kaplan, A.; Strachan, R. T.; Braz, J. M.; Craik, V.; Slocum, S.; Mangano, T.; Amabo, V.; O’Donnell, H.; Lak, P.; Basbaum, A. I.; Roth, B. L.; Shoichet, B. K. Structure-Based Design of a Chemical Probe Set for the 5-HT<sub>5A</sub> Serotonin Receptor. *J. Med. Chem.* **2022**, *65* (5), 4201–4217. <https://doi.org/10.1021/acs.jmedchem.1c02031>.
30. Carlsson, J.; Coleman, R. G.; Setola, V.; Irwin, J. J.; Fan, H.; Schlessinger, A.; Sali, A.; Roth, B. L.; Shoichet, B. K. Ligand Discovery from a Dopamine D3 Receptor Homology Model and Crystal Structure. *Nat. Chem. Biol.* **2011**, *7* (11), 769–778. <https://doi.org/10.1038/nchembio.662>.
31. Wang, S.; Che, T.; Levit, A.; Shoichet, B. K.; Wacker, D.; Roth, B. L. Structure of the D2 Dopamine Receptor Bound to the Atypical Antipsychotic Drug Risperidone. *Nature* **2018**, *555* (7695), 269–273. <https://doi.org/10.1038/nature25758>.
32. Fink, E. A.; Xu, J.; Hübner, H.; Braz, J. M.; Seemann, P.; Avet, C.; Craik, V.; Weikert, D.; Schmidt, M. F.; Webb, C. M.; Tolmachova, N. A.; Moroz, Y. S.; Huang, X.-P.; Kalyanaraman, C.; Gahbauer, S.; Chen, G.; Liu, Z.; Jacobson, M. P.; Irwin, J. J.; Bouvier, M.; Du, Y.; Shoichet, B. K.; Basbaum, A. I.; Gmeiner, P. Structure-Based Discovery of Nonopioid

- Analgesics Acting through the A2A-Adrenergic Receptor. *Science* **2022**, 377 (6614), eabn7065. <https://doi.org/10.1126/science.abn7065>.
33. Kaplan, A. L.; Confair, D. N.; Kim, K.; Barros-Álvarez, X.; Rodriguiz, R. M.; Yang, Y.; Kweon, O. S.; Che, T.; McCorvy, J. D.; Kamber, D. N.; Phelan, J. P.; Martins, L. C.; Pogorelov, V. M.; DiBerto, J. F.; Slocum, S. T.; Huang, X.-P.; Kumar, J. M.; Robertson, M. J.; Panova, O.; Seven, A. B.; Wetsel, A. Q.; Wetsel, W. C.; Irwin, J. J.; Skiniotis, G.; Shoichet, B. K.; Roth, B. L.; Ellman, J. A. Bespoke Library Docking for 5-HT<sub>2A</sub> Receptor Agonists with Antidepressant Activity. *Nature* **2022**. <https://doi.org/10.1038/s41586-022-05258-z>.
34. Schuller, M.; Correy, G. J.; Gahbauer, S.; Fearon, D.; Wu, T.; Díaz, R. E.; Young, I. D.; Carvalho Martins, L.; Smith, D. H.; Schulze-Gahmen, U.; Owens, T. W.; Deshpande, I.; Merz, G. E.; Thwin, A. C.; Biel, J. T.; Peters, J. K.; Moritz, M.; Herrera, N.; Kratochvil, H. T.; QCRG Structural Biology Consortium; Aimon, A.; Bennett, J. M.; Brandao Neto, J.; Cohen, A. E.; Dias, A.; Douangamath, A.; Dunnett, L.; Fedorov, O.; Ferla, M. P.; Fuchs, M. R.; Gorrie-Stone, T. J.; Holton, J. M.; Johnson, M. G.; Krojer, T.; Meigs, G.; Powell, A. J.; Rack, J. G. M.; Rangel, V. L.; Russi, S.; Skyner, R. E.; Smith, C. A.; Soares, A. S.; Wierman, J. L.; Zhu, K.; O'Brien, P.; Jura, N.; Ashworth, A.; Irwin, J. J.; Thompson, M. C.; Gestwicki, J. E.; von Delft, F.; Shoichet, B. K.; Fraser, J. S.; Ahel, I. Fragment Binding to the Nsp3 Macrodomein of SARS-CoV-2 Identified through Crystallographic Screening and Computational Docking. *Sci. Adv.* **2021**, 7 (16), eabf8711. <https://doi.org/10.1126/sciadv.abf8711>.
35. Gahbauer, S.; Correy, G. J.; Schuller, M.; Ferla, M. P.; Doruk, Y. U.; Rachman, M.; Wu, T.; Diolaiti, M.; Wang, S.; Neitz, R. J.; Fearon, D.; Radchenko, D.; Moroz, Y.; Irwin, J. J.; Renslo, A. R.; Taylor, J. C.; Gestwicki, J. E.; Delft, F. von; Ashworth, A.; Ahel, I.; Shoichet, B. K.;



- Fraser, J. S. *Structure-Based Inhibitor Optimization for the Nsp3 Macrodomein of SARS-CoV-2*; preprint; Biophysics, 2022. <https://doi.org/10.1101/2022.06.27.497816>.
36. Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X).
37. Wang, H.; He, S.; Deng, W.; Zhang, Y.; Li, G.; Sun, J.; Zhao, W.; Guo, Y.; Yin, Z.; Li, D.; Shang, L. Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease. *ACS Catal.* **2020**, *10* (10), 5871–5890. <https://doi.org/10.1021/acscatal.0c00110>.
38. Goetz, D. H.; Choe, Y.; Hansell, E.; Chen, Y. T.; McDowell, M.; Jonsson, C. B.; Roush, W. R.; McKerrow, J.; Craik, C. S. Substrate Specificity Profiling and Identification of a New Class of Inhibitor for the Major Protease of the SARS Coronavirus . *Biochemistry* **2007**, *46* (30), 8744–8752. <https://doi.org/10.1021/bi0621415>.
39. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved  $\alpha$ -Ketoamide Inhibitors. *Science* **2020**, *368* (6489), 409–412. <https://doi.org/10.1126/science.abb3405>.
40. Zhu, L.; George, S.; Schmidt, M. F.; Al-Gharabli, S. I.; Rademann, J.; Hilgenfeld, R. Peptide Aldehyde Inhibitors Challenge the Substrate Specificity of the SARS-Coronavirus Main Protease. *Antiviral Res.* **2011**, *92* (2), 204–212. <https://doi.org/10.1016/j.antiviral.2011.08.001>.
41. Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S. A.; Chase, P.; Eggler, A.; Dawson, E. S.; Baez-Santos, Y. M.; Tomar, S.; Mielech, A. M.; Baker, S. C.; Lindsley, C. W.;

- Hodder, P.; Mesecar, A.; Stauffer, S. R. Discovery, Synthesis, and Structure-Based Optimization of a Series of N-(Tert-Butyl)-2-(N-Arylamido)-2-(Pyridin-3-Yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease. *J. Med. Chem.* **2013**, *56* (2), 534–546. <https://doi.org/10.1021/jm301580n>.
42. Bender, B. J.; Gahbauer, S.; Lutgens, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balias, T. E.; Carlsson, J.; Irwin, J.; Shoichet, B. K. A Practical Guide to Large-Scale Docking. *Nat Protoc* **2021**. <https://doi.org/10.1038/s41596-021-00597-z>.
43. Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>.
44. Gu, S.; Smith, M. S.; Yang, Y.; Irwin, J. J.; Shoichet, B. K. *Ligand Strain Energy in Large Library Docking*; preprint; Bioinformatics, 2021. <https://doi.org/10.1101/2021.04.06.438722>.
45. O'Donnell, H. R.; Tummino, T. A.; Bardine, C.; Craik, C. S.; Shoichet, B. K. Colloidal Aggregators in Biochemical SARS-CoV-2 Repurposing Screens. *J. Med. Chem.* **2021**, *64* (23), 17530–17539. <https://doi.org/10.1021/acs.jmedchem.1c01547>.
46. McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* **2003**, *46* (20), 4265–4272. <https://doi.org/10.1021/jm030266r>.
47. Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065–6073. <https://doi.org/10.1021/acs.jcim.0c00675>.

48. Stein, R. M.; Yang, Y.; Balius, T. E.; O'Meara, M. J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B. K.; Irwin, J. J. Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model.* **2021**, *61* (2), 699–714. <https://doi.org/10.1021/acs.jcim.0c00598>.
49. Purkey, H. Abstract ND11: Discovery of GDC-6036, a Clinical Stage Treatment for KRAS G12C-Positive Cancers. *Cancer Res.* **2022**, *82* (12\_Supplement), ND11–ND11. <https://doi.org/10.1158/1538-7445.AM2022-ND11>.
50. Canon, J.; Rex, K.; Saiki, A. Y.; Mohr, C.; Cooke, K.; Bagal, D.; Gaida, K.; Holt, T.; Knutson, C. G.; Koppada, N.; Lanman, B. A.; Werner, J.; Rapaport, A. S.; San Miguel, T.; Ortiz, R.; Osgood, T.; Sun, J.-R.; Zhu, X.; McCarter, J. D.; Volak, L. P.; Houk, B. E.; Fakih, M. G.; O'Neil, B. H.; Price, T. J.; Falchook, G. S.; Desai, J.; Kuo, J.; Govindan, R.; Hong, D. S.; Ouyang, W.; Henary, H.; Arvedson, T.; Cee, V. J.; Lipford, J. R. The Clinical KRAS(G12C) Inhibitor AMG 510 Drives Anti-Tumour Immunity. *Nature* **2019**, *575* (7781), 217–223. <https://doi.org/10.1038/s41586-019-1694-1>.
51. Fairhurst, R. A.; Knoepfel, T.; Buschmann, N.; Leblanc, C.; Mah, R.; Todorov, M.; Nimsgern, P.; Ripoche, S.; Niklaus, M.; Warin, N.; Luu, V. H.; Madoerin, M.; Wirth, J.; Graus-Porta, D.; Weiss, A.; Kiffe, M.; Wartmann, M.; Kinyamu-Akunda, J.; Sterker, D.; Stamm, C.; Adler, F.; Buhles, A.; Schadt, H.; Couttet, P.; Blank, J.; Galuba, I.; Trappe, J.; Voshol, J.; Ostermann, N.; Zou, C.; Berghausen, J.; Del Rio Espinola, A.; Jahnke, W.; Furet, P. Discovery of Roblitinib (FGF401) as a Reversible-Covalent Inhibitor of the Kinase Activity of Fibroblast Growth Factor Receptor 4. *J. Med. Chem.* **2020**, *63* (21), 12542–12573. <https://doi.org/10.1021/acs.jmedchem.0c01019>.

52. Lin, C.; Kwong, A. D.; Perni, R. B. Discovery and Development of VX-950, a Novel, Covalent, and Reversible Inhibitor of Hepatitis C Virus NS3.4A Serine Protease. *Infect. Disord. Drug Targets* **2006**, *6* (1), 3–16. <https://doi.org/10.2174/187152606776056706>.
53. Venkatraman, S. Discovery of Boceprevir, a Direct-Acting NS3/4A Protease Inhibitor for Treatment of Chronic Hepatitis C Infections. *Trends Pharmacol. Sci.* **2012**, *33* (5), 289–294. <https://doi.org/10.1016/j.tips.2012.03.012>.
54. Metcalf, B.; Chuang, C.; Dufu, K.; Patel, M. P.; Silva-Garcia, A.; Johnson, C.; Lu, Q.; Partridge, J. R.; Patskovska, L.; Patskovsky, Y.; Almo, S. C.; Jacobson, M. P.; Hua, L.; Xu, Q.; Gwaltney, S. L.; Yee, C.; Harris, J.; Morgan, B. P.; James, J.; Xu, D.; Hutchaleelaha, A.; Paulvannan, K.; Oksenberg, D.; Li, Z. Discovery of GBT440, an Orally Bioavailable R-State Stabilizer of Sick Cell Hemoglobin. *ACS Med. Chem. Lett.* **2017**, *8* (3), 321–326. <https://doi.org/10.1021/acsmchemlett.6b00491>.
55. Vankadara, S.; Dawson, M. D.; Fong, J. Y.; Oh, Q. Y.; Ang, Q. A.; Liu, B.; Chang, H. Y.; Koh, J.; Koh, X.; Tan, Q. W.; Joy, J.; Chia, C. S. B. A Warhead Substitution Study on the Coronavirus Main Protease Inhibitor Nirmatrelvir. *ACS Med. Chem. Lett.* **2022**, *13* (8), 1345–1350. <https://doi.org/10.1021/acsmchemlett.2c00260>.
56. Adams, J.; Behnke, M.; Chen, S.; Cruickshank, A. A.; Dick, L. R.; Grenier, L.; Klunder, J. M.; Ma, Y.-T.; Plamondon, L.; Stein, R. L. Potent and Selective Inhibitors of the Proteasome: Dipeptidyl Boronic Acids. *Bioorg. Med. Chem. Lett.* **1998**, *8* (4), 333–338. [https://doi.org/10.1016/S0960-894X\(98\)00029-8](https://doi.org/10.1016/S0960-894X(98)00029-8).
57. Boike, L.; Henning, N. J.; Nomura, D. K. Advances in Covalent Drug Discovery. *Nat. Rev. Drug Discov.* **2022**. <https://doi.org/10.1038/s41573-022-00542-z>.

58. London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J. Covalent Docking of Large Libraries for the Discovery of Chemical Probes. *Nat. Chem. Biol.* **2014**, *10* (12), 1066–1072. <https://doi.org/10.1038/nchembio.1666>.
59. Wan, X.; Yang, T.; Cuesta, A.; Pang, X.; Balius, T. E.; Irwin, J. J.; Shoichet, B. K.; Taunton, J. Discovery of Lysine-Targeted EIF4E Inhibitors through Covalent Docking. *J. Am. Chem. Soc.* **2020**, *142* (11), 4960–4964. <https://doi.org/10.1021/jacs.9b10377>.
60. Wang, S.; Wacker, D.; Levit, A.; Che, T.; Betz, R. M.; McCorvy, J. D.; Venkatakrishnan, A. J.; Huang, X.-P.; Dror, R. O.; Shoichet, B. K.; Roth, B. L. D<sub>4</sub> Dopamine Receptor High-Resolution Structures Enable the Discovery of Selective Agonists. *Science* **2017**, *358* (6361), 381–386. <https://doi.org/10.1126/science.aan5468>.
61. Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against  $\beta$ -Lactamase. *J. Med. Chem.* **2008**, *51* (8), 2502–2511. <https://doi.org/10.1021/jm701500e>.
62. Korczynska, M.; Clark, M. J.; Valant, C.; Xu, J.; Moo, E. V.; Albold, S.; Weiss, D. R.; Torosyan, H.; Huang, W.; Kruse, A. C.; Lyda, B. R.; May, L. T.; Baltos, J.-A.; Sexton, P. M.; Kobilka, B. K.; Christopoulos, A.; Shoichet, B. K.; Sunahara, R. K. Structure-Based Discovery of Selective Positive Allosteric Modulators of Antagonists for the M<sub>2</sub> Muscarinic Acetylcholine Receptor. *Proc. Natl. Acad. Sci.* **2018**, *115* (10). <https://doi.org/10.1073/pnas.1718037115>.

63. Huang, X.-P.; Karpiak, J.; Kroeze, W. K.; Zhu, H.; Chen, X.; Moy, S. S.; Sadoris, K. A.; Nikolova, V. D.; Farrell, M. S.; Wang, S.; Mangano, T. J.; Deshpande, D. A.; Jiang, A.; Penn, R. B.; Jin, J.; Koller, B. H.; Kenakin, T.; Shoichet, B. K.; Roth, B. L. Allosteric Ligands for the Pharmacologically Dark Receptors GPR68 and GPR65. *Nature* **2015**, *527* (7579), 477–483. <https://doi.org/10.1038/nature15699>.
64. Mysinger, M. M.; Weiss, D. R.; Ziarek, J. J.; Gravel, S.; Doak, A. K.; Karpiak, J.; Heveker, N.; Shoichet, B. K.; Volkman, B. F. Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4. *Proc. Natl. Acad. Sci.* **2012**, *109* (14), 5517–5522. <https://doi.org/10.1073/pnas.1120431109>.
65. Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; Zvonok, N.; Jain, M. K.; Savych, O.; Radchenko, D. S.; Nikas, S. P.; Petasis, N. A.; Moroz, Y. S.; Roth, B. L.; Makriyannis, A.; Katritch, V. Synthon-Based Ligand Discovery in Virtual Libraries of over 11 Billion Compounds. *Nature* **2022**, *601* (7893), 452–459. <https://doi.org/10.1038/s41586-021-04220-9>.
66. Zhao, N.; Bardine, C.; Lourenço, A. L.; Wang, Y.-H.; Huang, Y.; Cleary, S. J.; Wilson, D. M.; Oh, D. Y.; Fong, L.; Looney, M. R.; Evans, M. J.; Craik, C. S. In Vivo Measurement of Granzyme Proteolysis from Activated Immune Cells with PET. *ACS Cent. Sci.* **2021**, *7* (10), 1638–1649. <https://doi.org/10.1021/acscentsci.1c00529>.
67. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>.

68. Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS ONE* **2013**, *8* (10), e75992. <https://doi.org/10.1371/journal.pone.0075992>.
69. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13* (4), 505–524. <https://doi.org/10.1002/jcc.540130412>.
70. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784. <https://doi.org/10.1021/ja00315a051>.
71. Gallagher, K.; Sharp, K. Electrostatic Contributions to Heat Capacity Changes of DNA-Ligand Binding. *Biophys. J.* **1998**, *75* (2), 769–776. [https://doi.org/10.1016/S0006-3495\(98\)77566-6](https://doi.org/10.1016/S0006-3495(98)77566-6).
72. Mysinger, M. M.; Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **2010**, *50* (9), 1561–1573. <https://doi.org/10.1021/ci100214a>.
73. Turlington, M.; Chun, A.; Tomar, S.; Egger, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J. S.; Dawson, E.; Saldanha, A.; Chase, P.; Baez-Santos, Y. M.; Lindsley, C. W.; Hodder, P.; Mesecar, A. D.; Stauffer, S. R. Discovery of N-(Benzo[1,2,3]Triazol-1-Yl)-N-(Benzyl)Acetamido)Phenyl) Carboxamides as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Inhibitors: Identification of ML300 and Noncovalent Nanomolar Inhibitors with an Induced-Fit Binding. *Bioorg. Med. Chem. Lett.* **2013**, *23* (22), 6172–6177. <https://doi.org/10.1016/j.bmcl.2013.08.112>.
74. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100–D1107. <https://doi.org/10.1093/nar/gkr777>.

75. Breidenbach, J.; Lemke, C.; Pillaiyar, T.; Schäkel, L.; Al Hamwi, G.; Dieltz, M.; Gedschold, R.; Geiger, N.; Lopez, V.; Mirza, S.; Namasivayam, V.; Schiedel, A. C.; Sylvester, K.; Thimm, D.; Vielmuth, C.; Phuong Vu, L.; Zyulina, M.; Bodem, J.; Gütschow, M.; Müller, C. E. Targeting the Main Protease of SARS-CoV-2: From the Establishment of High Throughput Screening to the Design of Tailored Inhibitors. *Angew. Chem. Int. Ed Engl.* **2021**, *60* (18), 10423–10429. <https://doi.org/10.1002/anie.202016961>.
76. Yang, S.; Chen, S.-J.; Hsu, M.-F.; Wu, J.-D.; Tseng, C.-T. K.; Liu, Y.-F.; Chen, H.-C.; Kuo, C.-W.; Wu, C.-S.; Chang, L.-W.; Chen, W.-C.; Liao, S.-Y.; Chang, T.-Y.; Hung, H.-H.; Shr, H.-L.; Liu, C.-Y.; Huang, Y.-A.; Chang, L.-Y.; Hsu, J.-C.; Peters, C. J.; Wang, A. H.-J.; Hsu, M.-C. Synthesis, Crystal Structure, Structure–Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *J. Med. Chem.* **2006**, *49* (16), 4971–4980. <https://doi.org/10.1021/jm0603926>.
77. Westberg, M.; Su, Y.; Zou, X.; Ning, L.; Hurst, B.; Tarbet, B.; Lin, M. Z. *Rational Design of a New Class of Protease Inhibitors for the Potential Treatment of Coronavirus Diseases*; preprint; Biochemistry, 2020. <https://doi.org/10.1101/2020.09.15.275891>.
78. Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Günther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B.; Chen, J.; Luo, H.; Yang, Y.; Yang, Y.; Zou, J.; Thiel, V.; Chen, K.; Shen, J.; Shen, X.; Jiang, H. Cinanserin Is an Inhibitor of the 3C-like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication in Vitro. *J. Virol.* **2005**, *79* (11), 7095–7103. <https://doi.org/10.1128/JVI.79.11.7095-7103.2005>.
79. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; Zhang, X. C.; Liao, M.; Bartlam, M.; Rao, Z. Structures of Two Coronavirus Main Proteases:



- Implications for Substrate Binding and Antiviral Drug Design. *J. Virol.* **2008**, *82* (5), 2515–2527. <https://doi.org/10.1128/JVI.02114-07>.
80. Ghosh, A. K.; Gong, G.; Grum-Tokars, V.; Mulhearn, D. C.; Baker, S. C.; Coughlin, M.; Prabhakar, B. S.; Sleeman, K.; Johnson, M. E.; Mesecar, A. D. Design, Synthesis and Antiviral Efficacy of a Series of Potent Chloropyridyl Ester-Derived SARS-CoV 3CLpro Inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18* (20), 5684–5688. <https://doi.org/10.1016/j.bmcl.2008.08.082>.
81. Kim, Y.; Lovell, S.; Tiew, K.-C.; Mandadapu, S. R.; Alliston, K. R.; Battaile, K. P.; Groutas, W. C.; Chang, K.-O. Broad-Spectrum Antivirals against 3C or 3C-like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. *J. Virol.* **2012**, *86* (21), 11754–11762. <https://doi.org/10.1128/JVI.01348-12>.
82. Wu, C.-Y.; King, K.-Y.; Kuo, C.-J.; Fang, J.-M.; Wu, Y.-T.; Ho, M.-Y.; Liao, C.-L.; Shie, J.-J.; Liang, P.-H.; Wong, C.-H. Stable Benzotriazole Esters as Mechanism-Based Inactivators of the Severe Acute Respiratory Syndrome 3CL Protease. *Chem. Biol.* **2006**, *13* (3), 261–268. <https://doi.org/10.1016/j.chembiol.2005.12.008>.
83. Zhang, J.; Huitema, C.; Niu, C.; Yin, J.; James, M. N. G.; Eltis, L. D.; Vederas, J. C. Aryl Methylene Ketones and Fluorinated Methylene Ketones as Reversible Inhibitors for Severe Acute Respiratory Syndrome (SARS) 3C-like Proteinase. *Bioorganic Chem.* **2008**, *36* (5), 229–240. <https://doi.org/10.1016/j.bioorg.2008.01.001>.
84. Chen, L.-R.; Wang, Y.-C.; Lin, Y. W.; Chou, S.-Y.; Chen, S.-F.; Liu, L. T.; Wu, Y.-T.; Kuo, C.-J.; Chen, T. S.-S.; Juang, S.-H. Synthesis and Evaluation of Isatin Derivatives as Effective SARS Coronavirus 3CL Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15* (12), 3058–3062. <https://doi.org/10.1016/j.bmcl.2005.04.027>.

85. Mukherjee, P.; Desai, P.; Ross, L.; White, E. L.; Avery, M. A. Structure-Based Virtual Screening against SARS-3CL(pro) to Identify Novel Non-Peptidic Hits. *Bioorg. Med. Chem.* **2008**, *16* (7), 4138–4149. <https://doi.org/10.1016/j.bmc.2008.01.011>.
86. Blanchard, J. E.; Elowe, N. H.; Huitema, C.; Fortin, P. D.; Cechetto, J. D.; Eltis, L. D.; Brown, E. D. High-Throughput Screening Identifies Inhibitors of the SARS Coronavirus Main Proteinase. *Chem. Biol.* **2004**, *11* (10), 1445–1453. <https://doi.org/10.1016/j.chembiol.2004.08.011>.
87. Rathnayake, A. D.; Zheng, J.; Kim, Y.; Perera, K. D.; Mackin, S.; Meyerholz, D. K.; Kashipathy, M. M.; Battaile, K. P.; Lovell, S.; Perlman, S.; Groutas, W. C.; Chang, K.-O. 3C-like Protease Inhibitors Block Coronavirus Replication in Vitro and Improve Survival in MERS-CoV-Infected Mice. *Sci. Transl. Med.* **2020**, *12* (557), eabc5332. <https://doi.org/10.1126/scitranslmed.abc5332>.
88. Yang, K. S.; Ma, X. R.; Ma, Y.; Alugubelli, Y. R.; Scott, D. A.; Vatansever, E. C.; Drelich, A. K.; Sankaran, B.; Geng, Z. Z.; Blankenship, L. R.; Ward, H. E.; Sheng, Y. J.; Hsu, J. C.; Kratch, K. C.; Zhao, B.; Hayatshahi, H. S.; Liu, J.; Li, P.; Fierke, C. A.; Tseng, C.-T. K.; Xu, S.; Liu, W. R. A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors\*. *ChemMedChem* **2021**, *16* (6), 942–948. <https://doi.org/10.1002/cmdc.202000924>.
89. Han, S. H.; Goins, C. M.; Arya, T.; Shin, W.-J.; Maw, J.; Hooper, A.; Sonawane, D. P.; Porter, M. R.; Bannister, B. E.; Crouch, R. D.; Lindsey, A. A.; Lakatos, G.; Martinez, S. R.; Alvarado, J.; Akers, W. S.; Wang, N. S.; Jung, J. U.; Macdonald, J. D.; Stauffer, S. R. Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 3CLpro). *J. Med. Chem.* **2022**, *65* (4), 2880–2904. <https://doi.org/10.1021/acs.jmedchem.1c00598>.

90. Clyde, A.; Galanie, S.; Kneller, D. W.; Ma, H.; Babuji, Y.; Blaiszik, B.; Brace, A.; Brettin, T.; Chard, K.; Chard, R.; Coates, L.; Foster, I.; Hauner, D.; Kertesz, V.; Kumar, N.; Lee, H.; Li, Z.; Merzky, A.; Schmidt, J. G.; Tan, L.; Titov, M.; Trifan, A.; Turilli, M.; Van Dam, H.; Chennubhotla, S. C.; Jha, S.; Kovalevsky, A.; Ramanathan, A.; Head, M. S.; Stevens, R. High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor. *J. Chem. Inf. Model.* **2022**, *62* (1), 116–128. <https://doi.org/10.1021/acs.jcim.1c00851>.
91. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 125–132. <https://doi.org/10.1107/S0907444909047337>.
92. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50* (Pt 5), 760–763. <https://doi.org/10.1107/S0907444994003112>.
93. Evans, P. Scaling and Assessment of Data Quality. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62* (Pt 1), 72–82. <https://doi.org/10.1107/S0907444905036693>.
94. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (Pt 4), 658–674. <https://doi.org/10.1107/S0021889807021206>.
95. Costanzi, E.; Kuzikov, M.; Esposito, F.; Albani, S.; Demitri, N.; Giabbai, B.; Camasta, M.; Tramontano, E.; Rossetti, G.; Zaliani, A.; Storici, P. Structural and Biochemical Analysis of the Dual Inhibition of MG-132 against SARS-CoV-2 Main Protease (Mpro/3CLpro) and Human Cathepsin-L. *Int. J. Mol. Sci.* **2021**, *22* (21), 11779. <https://doi.org/10.3390/ijms222111779>.

96. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501. <https://doi.org/10.1107/S0907444910007493>.
97. Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (Pt 4), 355–367. <https://doi.org/10.1107/S0907444911001314>.
98. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (Pt 4), 352–367. <https://doi.org/10.1107/S0907444912001308>.
99. Allen, W. J.; Rizzo, R. C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *J. Chem. Inf. Model.* **2014**, *54* (2), 518–529. <https://doi.org/10.1021/ci400534h>.
100. 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.* **2021**, *64* (7), 4109–4116. <https://doi.org/10.1021/acs.jmedchem.0c02253>.

## Gloss to Chapter 5

Finally, the expansion of accessible chemical space and virtual databases used in molecular docking screens is considered here. Thanks to companies like Enamine, the number of molecules we can use for virtual screening is constantly growing because of their exploration of small building blocks and reliable reactions that can combine them into not already existing molecules. Reaching one billion molecules was a landmark achievement in the lab, and now we are approaching 10 billion. We hope this increase is useful with a ‘bigger is better’ approach, but to double check this is the case the lab has conducted experiments to simply ask “Is bigger really better?”

Here, I build off ideas from Jiankun Lyu’s work (Lyu et al. *Nature* 2019, Alon et al. *Nature* 2021, and Lyu et al. *Nat Chem Bio* 2022) in asking similar questions in context of a new use case of virtual screening and chemical databases. We are interested in finding compounds that modulate not just one protein (like ‘normal’) but instead a pair of desired proteins. The idea being molecules that can modulate multiple proteins or nodes in a complex disease state physiological network could have increased efficacy or benefit compared to only modulating one of the proteins alone; we define this as polypharmacology. The disease network of pain is the primary application as the lab and field look to discover new analgesic therapies with improved efficacies without the opioid side effects, or other side effects associated with a single protein therapeutic.

The work in the following chapter is in its preliminary stages but will take two paths: 1) Retrospective analysis performed on model systems will be used to probe how successful virtual screening for polypharmacology might be, and 2) prospective test cases for pairs of proteins

implicated in pain will be explored. The work has given me an opportunity to learn larger coding and data management skills, and work on many more protein systems. It brings the total number of proteins I have performed some type of computational work with up to ~22, and I am likely forgetting some. The work also allowed me to think about the technique differently than a typical discovery project of applying the pipeline of setup optimization, virtual screen, and compound selection.

Overall, the work starts to showcase the growth of databases beyond 500 million increases the number of shared top-ranked molecules that we could consider in looking for polypharmacology ligands; this is true for both model systems and for a prospective test pair of proteins involved in pain. As the work continues, it holds great opportunity to be applied in the continued search for nonopioid (or at least, not only opioid) receptor pain therapeutics development. I sincerely hope one day society has safer analgesics available and sees some relief to the opioid epidemic.

## Chapter 5

### Polypharmacology and expanding virtual chemical space

#### Contributing Authors

Elissa A. Fink<sup>1,2</sup>, Laura Shub<sup>1,3</sup>, Yujin Wu<sup>1</sup>, John J. Irwin<sup>1,\*</sup>, Brian K. Shoichet<sup>1,\*</sup>

<sup>1</sup> Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, CA, USA

<sup>2</sup> Graduate Program in Biophysics, University of California-San Francisco, San Francisco, CA, USA

<sup>3</sup> Graduate Program in Bioinformatics, University of California-San Francisco, San Francisco, CA, USA

\* Corresponding author. Email: [jir322@gmail.com](mailto:jir322@gmail.com) (J.J.I.); [bshoichet@gmail.com](mailto:bshoichet@gmail.com) (B.K.S.)

## 5.1 Introduction

Traditionally, small molecule drug discovery projects desire highly selective drugs, active only on one protein target of interest. Often this is because additional activities on other proteins may produce side effects. However, there are existing therapeutics that modulate multiple proteins for a net efficacious effect that could not be achievable with a single protein activity. Their polypharmacology activities modulate a set of proteins that are typically nodes of related cellular circuitry for increased, either synergistic or additive, effects (1–3).

There are many instances where drugs with polypharmacology are beneficial. For example, dual or triple reuptake inhibitors modulate multiple neurotransmitter transporters for increased psychological effects (4), including the FDA-approved drug duloxetine. This is also true for complex pathways in cancer biology, where chemotherapy drugs will inhibit multiple subtypes of histone deacetylases (HDACs) (5), kinases, or enzymes in the folate synthesis pathway. Another use case for polypharmacology is offsetting side effects conferred by activity on a single target. Atypical antipsychotics antagonize the dopamine-2 receptor (D<sub>2</sub>R) for the desired efficacy, but this produces D<sub>2</sub>R-caused dyskinesia side effects; however, with the additional serotonin-2A receptor (5HT<sub>2A</sub>R) antagonism the D<sub>2</sub>R-caused side effects are reduced without affecting efficacy (6). Opioid receptor modulators also are known to have polypharmacology, including the dual  $\mu$ -opioid receptor ( $\mu$ OR) and  $\kappa$ -opioid receptor ( $\kappa$ OR) agonist morphine, and the dual  $\mu$ OR agonist and  $\kappa$ OR antagonist buprenorphine (7). The dual receptor agonism of morphine increases analgesic efficacy, and the mixed agonist-antagonist profile of buprenorphine enables pain and opioid use disorder treatment while offsetting side effects of morphine (or other dual opioid receptor agonists)-induced mood disorders and addiction.



Virtual screening ultra-large chemical databases has been successful for a range of protein targets (8–12), typically with 20-60% hit rates and new chemotypes with nM to  $\mu$ M activities (9, 11–14). The technique has been applied to polypharmacology only a few times with limited success (15), and only smaller libraries of 3 million molecules were screened. Currently ZINC22 (16) is over 5B and growing quickly towards 10B, with chemical space existing at much higher order of magnitudes (12). Here, we simulate how this increase in chemical space could affect the use of virtual screening to discovery molecules with designed polypharmacology. In addition, we apply the technique in a prospective polypharmacology docking campaign targeting the serotonin transporter (SERT) and  $\mu$ OR, proteins involved in the complex pain neuroscience network.

## 5.2 Results

### Database growth enriches for shared top-ranked molecules in model systems

Previous work has utilized model systems (11), or proteins with well characterized active molecules, to study the effect of perturbations to docking parameters. We looked to do the same and identified polypharmacology model systems that have ligands or FDA-approved drugs with multiple activities contributing to efficacy or offsetting side effects (**Table 5.1**). The protein pairs differ on axes of therapeutic application, protein class (transmembrane receptor, transporter, enzyme), binding sites, and overall protein similarity. Docking setups were prepared and 900 million ZINC22 molecules were virtually screened against each protein using DOCK3.8. All molecules scored in each of the protein docking screen were considered in this simulation of database growth and analysis (see Methods). Database sizes started at 10 thousand molecules and increased 5-fold up to 500 million molecules for a total of 10 sizes. When analyzing docking screens and selecting compounds to purchase, typically only the best scoring results are considered because the hit rate will fall as docked scores increase (get worse) (12, 13). We applied that logic here and only considered the top-ranked molecules by rank number or by percent, typically not looking beyond the best-ranked 10%.

We first asked how many molecules are randomly shared in the docking results at different thresholds and how it changes as the databases increase (**Fig. 5.1**). The simulations demonstrated a low percent of molecules are shared, with the highest number or percent shared occurring in the 500 million database and top 50 million ranked molecules. These values were used to establish a baseline to which further analysis could be compared back to.

For the following non-random analysis, each database chunk is sorted by DOCK scores. Here, as the database size increases from 10 thousand to 500 million, the number of shared top-ranked docked results increases (**Fig. 5.2**). In the top-ranked 50 million subset of the 500 million database the number of shared molecules is 10-16 million in the three model systems (or 17 to 21%), and in the top-ranked 5 million the number is 400 to 840 thousand (or 8-17%). Extrapolation using the linear regression of the log-log plots (**Table 5.S1**), in larger 10 billion databases the number of shared compounds may increase to 7 to 16 million in the top-ranked 5 million, 18 to 39 million in the top-ranked 10 million, and 166 to 290 million in the top-ranked 50 million. In past single target docking campaigns, typically <1% of the top-ranked results are analyzed in post-screen filtering steps. Logically, we may need to look at more top-ranked compounds as they will be double filtered (against two proteins) and up to 10% may be considered but merits further investigation.

As a control, the enrichment of shared top-ranked results was compared back to the random overlap baseline values. At each top-ranked subset of all database sizes, the number of shared molecules is larger than observed in the random results (**Fig. 5.3**). For the 5HT<sub>2A</sub>R-D<sub>2</sub>R results, the highest enrichments are in the 500 million database top 50 thousand and top 100 thousand ranked results with 779- and 674-fold more shared molecules versus random, respectively. The enrichment is next highest in the 500 million database top 10 thousand and top 500 thousand results with 187- and 136-fold more than random, respectively. The same four top-ranked subsets of the  $\mu$ OR- $\kappa$ OR and SERT-DAT 500 million databases are also the highest enriched, but lower in the 50 thousand and 100 thousand top-ranked results with all never higher than 200-fold enrichment over random.

The highest enrichment is occurring in the mid-top-ranked results and not at the highest ranked results (1,000 or 5,000) or when including a larger subset of top-ranked results (from 1 to 50 million). This could indicate the larger databases have more molecules that score well on a single target, here in the top 1,000 or 5,000 with low enrichment and decreases the number of top-ranked molecules that are scoring well on both proteins; this has previously been observed as a phenomenon that occurs with larger chemical databases (17). However, with the highest enrichment occurring in the mid-top-ranked results and not in the larger 1 to 50 million subsets, this indicates the database growth and increased single-target scoring molecules do not push the shared molecules completely to the bottom results to which we would consider for post-docking screen analysis. If that were the case, the highest enrichment would occur in the 1 to 50 million databases.

As previously mentioned, historically docking results look at top results by the top number of ranked compounds, usually 300 thousand, and usually is <1% of the total size of the chemical library docked. The analysis is converted now to looking at the top-ranked percent of results, starting at similar top-ranked thresholds of 0.01 to 1% and including a larger threshold up to 10%. The percent of shared molecules increases with database growth, also seen above (**Fig. 5.4**). The slopes of the log-log plots for 5HT<sub>2A</sub>R-D<sub>2</sub>R are very similar with the largest being the top-ranked 10%; this is also true for  $\mu$ OR- $\kappa$ OR and SERT-DAT with even larger slopes of the top-ranked 10%. There are some deviations from the linear regression, including at the top-ranked 10% of the 10 thousand database of 5HT<sub>2A</sub>R-D<sub>2</sub>R, the top-ranked 10% of the 50 million databases for all three model systems, and at all top-ranked percent thresholds of the three model systems. The reason for the deviation is unclear.

We return to ensuring the increase in top-ranked results is occurring at a rate higher than the random overlap using the fold-enrichment of the top-ranked docking results at different percent-thresholds. A surprising trend emerges in the 5HT<sub>2A</sub>R-D<sub>2</sub>R simulation, where an increase in the fold-enrichment hits a plateau after 3 5-fold increases in the database size, for example the increase the top 1% of results starting at 100 thousand hits the plateau at 5 million (**Fig. 5.5, Table 5.S2**). The trend is present in  $\mu$ OR- $\kappa$ OR and SERT-DAT, but less of a distinct plateau is present in the top 1% ranked data. For all three of model systems, the first plateau to emerge with database growth is the top-ranked 10% shared molecules. This could be worrying, as increasing the databases from 5 million to 500 million does not increase the enrichment of shared molecules versus random overlap and could be a continuing trend as databases increase further towards 10 and 20 billion molecules.

### **Characterizing prospective polypharmacology pair SERT- $\mu$ OR for pain**

Finally, with some initial understanding into the effect of chemical library growth in polypharmacology docking campaigns, we pursue a prospective test-case with application in pain therapeutics. Combinations of protein pairs are prioritized for many reasons, including a) each protein is involved in pain physiology, b) the proteins are considered ‘druggable’ and even further, ‘dockable’ with a high-confidence modeled or experimentally protein structure, and c) share similar ligand preferences (including, but not limited to, formal charge, cLogP, and molecular weight). One pain polypharmacology pairing prioritized is SERT- $\mu$ OR, and there is some precedence for pain polypharmacology ligands for these in the literature (18, 19). When comparing the binding site, they are similar with key recognition of cationic ligands by D98 in SERT and D147

in  $\mu$ OR. SERT ligand recognition is primarily driven by hydrophobic interactions and pi-pi interactions with F335, F341, and Y176, along with additional hydrophobic residues A169, V343, and L443 in the pocket.  $\mu$ OR has hydrophobic residues to complement those in SERT, such as Y148 which is proximal to the D147 in  $\mu$ OR as Y176 in SERT is with D98, Y326, W133, W318, and many valines and isoleucines.  $\mu$ OR also has more hydrophilic residues primed for hydrogen bonding in the binding site whereas SERT has less available to be similarly exploited. The experimental crystal structures were solved in SERT with a small molecule and in  $\mu$ OR with a small synthetic peptide causing the  $\mu$ OR site to be larger. Other  $\mu$ OR small molecule-bound experimental structures have been determined and may be considered for a more compatible template with SERT. (**Fig. 5.6A-B**).

The above analysis discussed above was also applied with SERT- $\mu$ OR (**Fig. 5.6C-E**). The random overlap with increasing database sizes remains low and constant, whereas the shared top-ranked results do increase with database growth. When comparing the enrichment of the shared molecules compared to the random overlap, the largest enrichment growths are seen at the mid-tier top-ranked thresholds of 10 to 100 thousand similar to the polypharmacology model systems. For the 500 million database, the top 1 million to 50 million ranked results contains 420 thousand to 12 million shared molecules. This will be the first prospective case where these molecules are filtered as normally performed in a single target docking screen and then selected for synthesis. It will be determined here how many remain viable at two targets compared to on a single target alone.

### 5.3 Discussion

From this retrospective analysis, notable conclusions emerge. First, the number of molecules shared in top-ranked docking results between pairs of polypharmacology model systems increases as databases increase from 10 thousand to 500 million (**Fig. 5.2**). This increase appears to be greater than the chance for random overlap (**Fig. 5.1**), indicating continued expansion of the chemical databases would be beneficial to increasing the number of shared molecules for potential selection for synthesis and purchasing.

Second, the largest enrichment of molecules compared to random is at the middle tiers of top-ranked results, usually 10,000 to 100,000 and not at the highest top-ranked threshold of 1,000 nor the lowest at 10 to 50 million (**Fig. 5.3**). This aligns with previous work (17), showing the larger databases improve DOCK scores in the very-top-ranked results, in that the polypharmacology enrichment would be negatively impacted and thus the top 1 or 5 thousand results would not see the highest polypharmacology enrichment. Also in previous work (12, 13), the experimental hit rate falls as you go down the docking ranked list by DOCK energy. The analysis here does not look past the top 10% ranked molecules with this decrease in ‘hit’ prevalence. It is promising the polypharmacology enrichment is not occurring at the highest rate for the top-ranked 10 or 50 million subsets as it likely approaches the region of hit rate decrease.

Third, the same analysis applied to the prospective pair of SERT- $\mu$ OR reveals similar trends as observed in the polypharmacology model systems (**Fig. 5.6**). In contrast to the model systems, these top-ranked results will be filtered, selected for 3D-complementarity against the two targets,

and purchased for make-on-demand synthesis. These results will extend the retrospective analysis towards a case study of using molecular docking for polypharmacology ligand discovery.

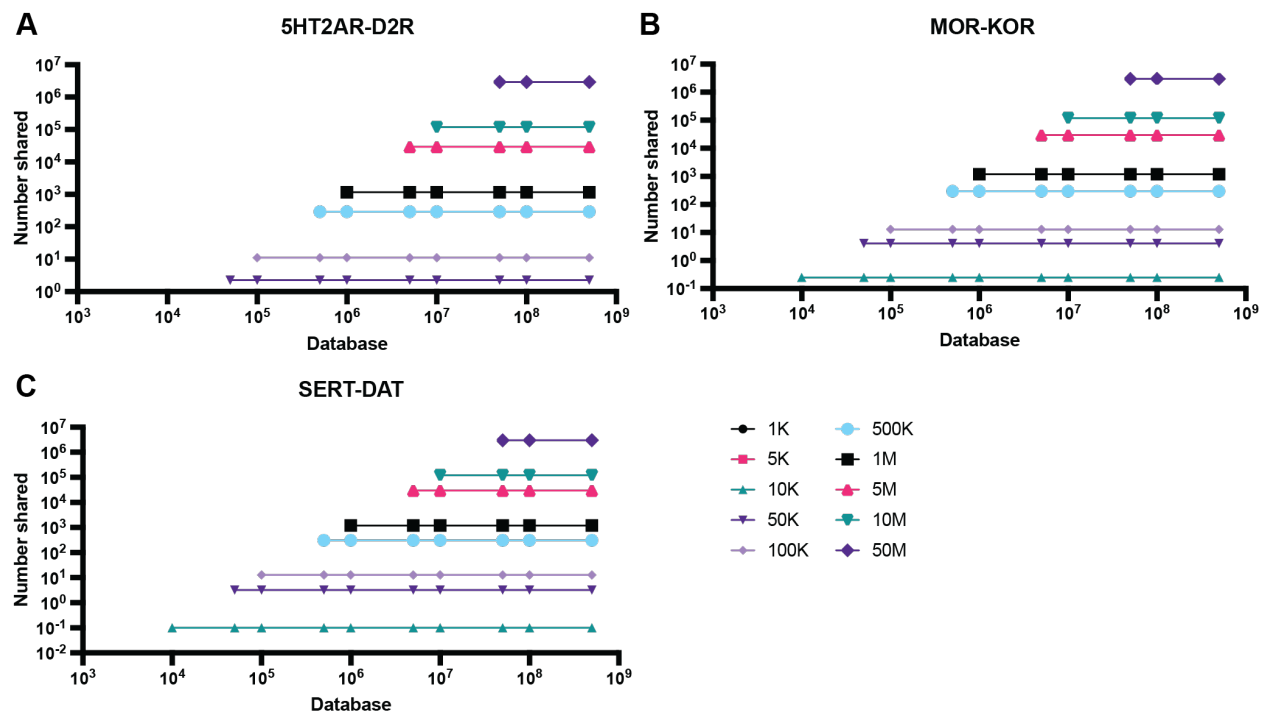


## 5.4 Future Directions

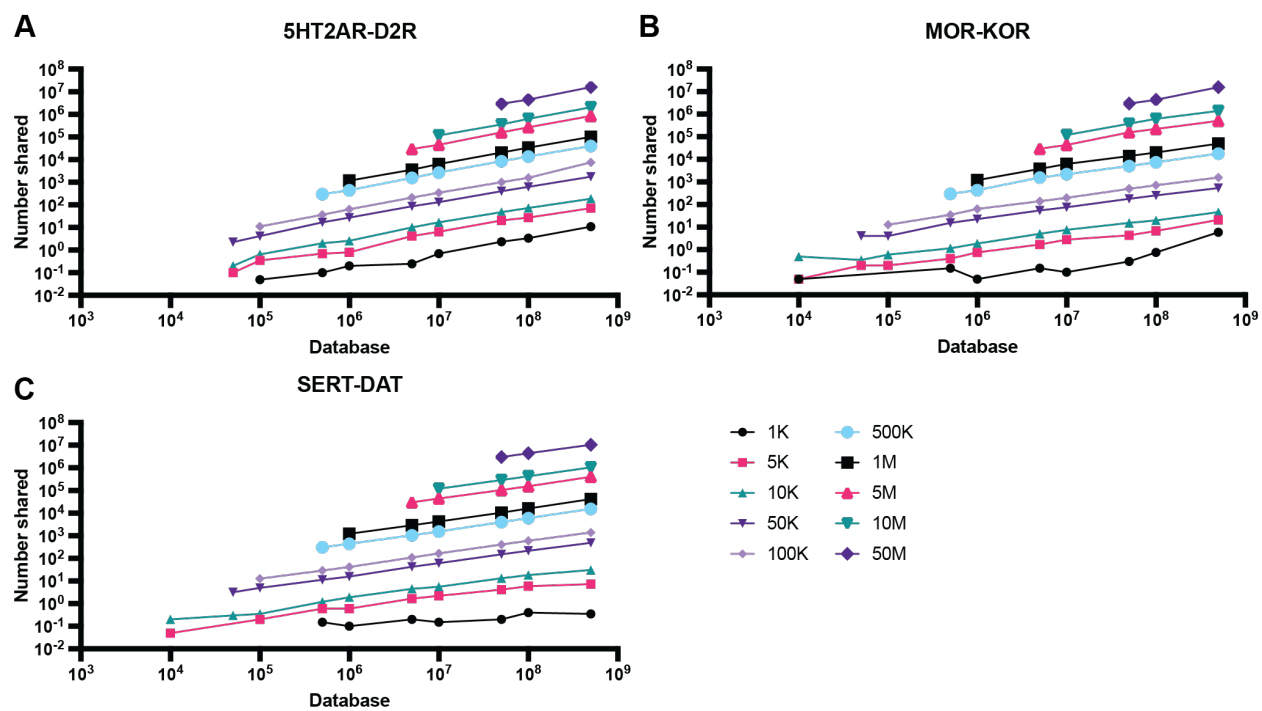
Finally, future directions are discussed here for the most immediate work to follow the culmination of this overall dissertation. For the retrospective analysis, this will be considered against the remaining model systems in **Table 5.1**. that move away from GPCRs and transporters. Additional characterization of the top-ranked results will consider chemotypes rather than just identical molecules, ensuring a large number of diverse chemical scaffolds are being enriched. We will also see if any known polypharmacology-like ligands are appearing in the top-ranked results as a positive control, and if those increase with database growth. We will also consider how many shared molecules or chemotypes might be needed to pursue further filtering and selection of compounds for synthesis. Finally, across the model systems we will explore how hit rates change from a single target to a pair of targets.

For prospective polypharmacology for pain, we currently are pursuing the SERT- $\mu$ OR system as well as  $\alpha_{2A}$ AR-SERT and  $\alpha_{2A}$ AR- $\mu$ OR. All three receptors have implications in pain, recognize similar cationic ligands, and have successful single target docking campaigns in which we can build off. Following these case studies, using molecular docking for polypharmacology discovery could be applied in other disease states where perturbations of multiple proteins is beneficial over a single target, including chemotherapy, antibiotic, antiviral, and other psychiatric therapeutics.

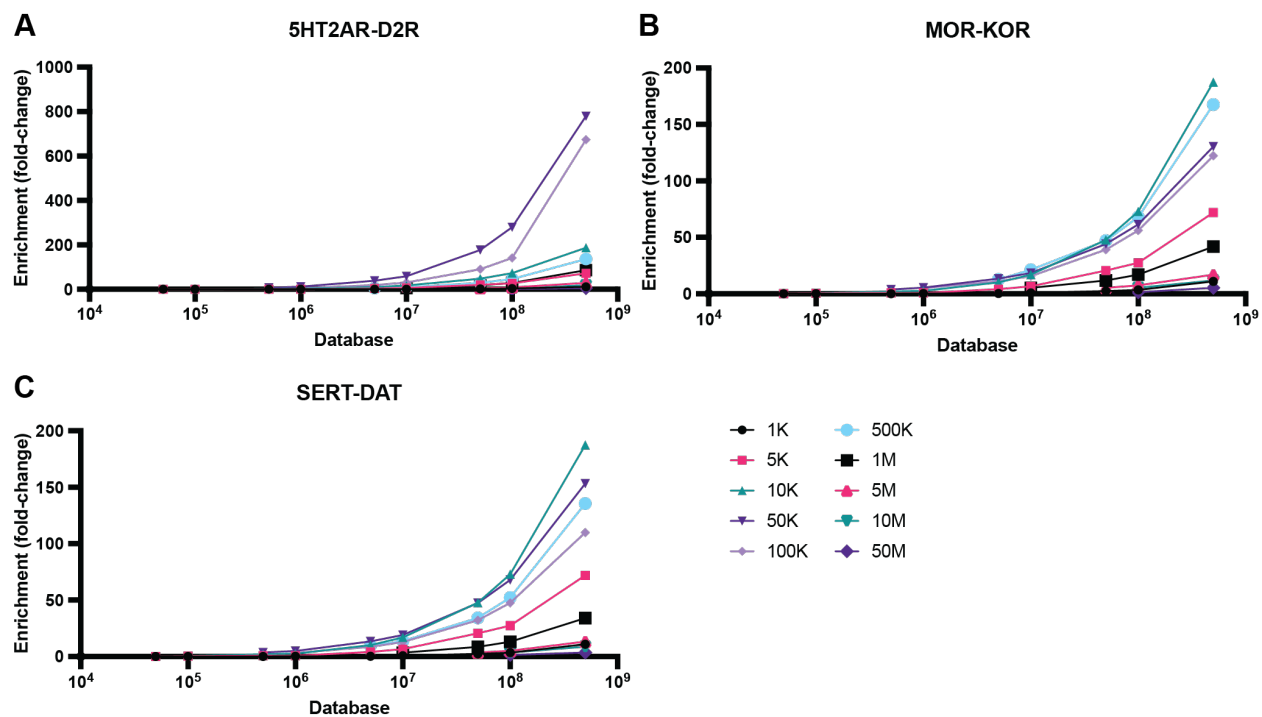
## 5.5 Figures



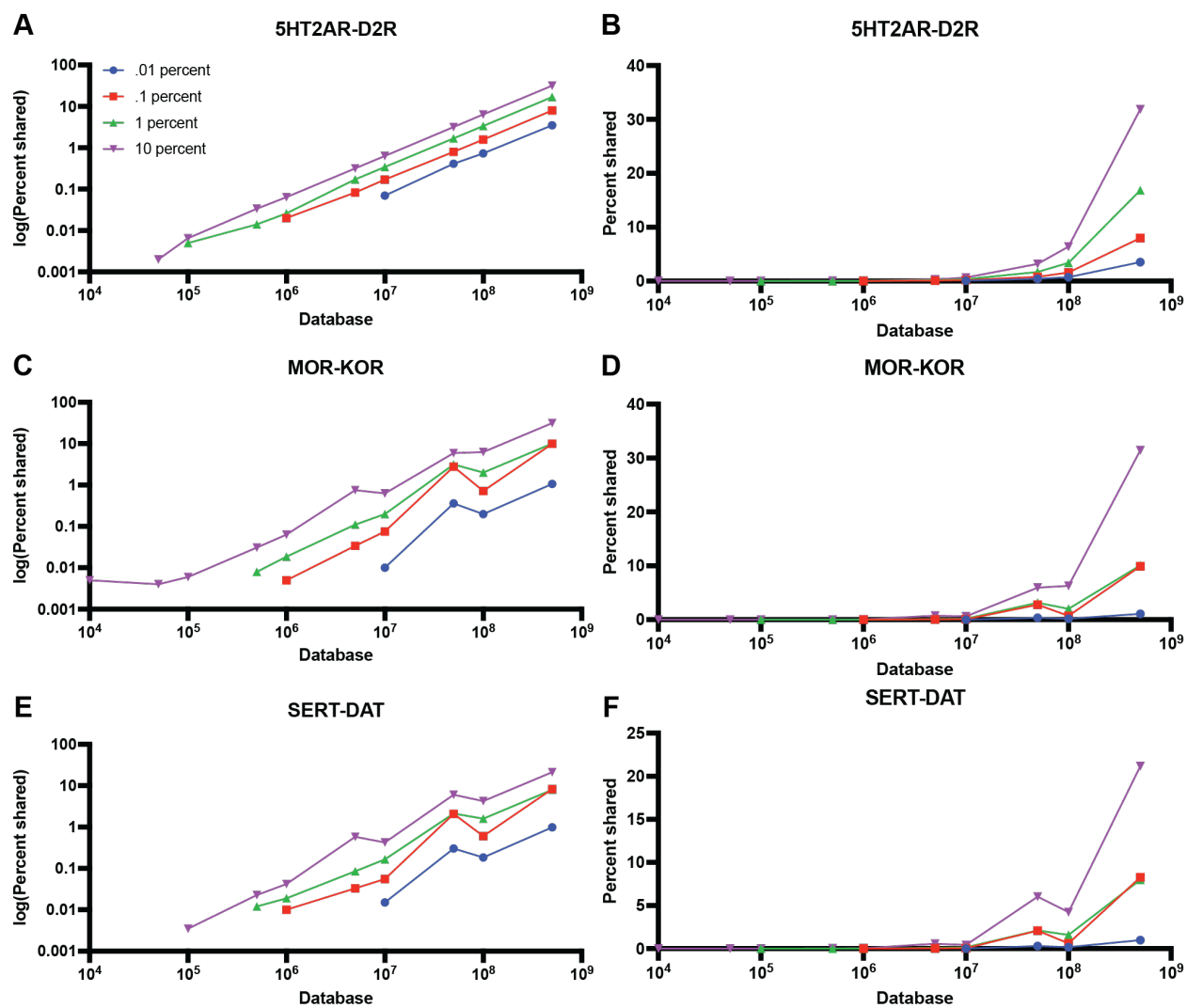
**Figure 5.1. Random overlap of shared top-ranked molecules are not enriched with database growth.** Polypharmacology model systems 5HT<sub>2A</sub>R-D<sub>2</sub>R (A),  $\mu$ OR- $\kappa$ OR (B), and SERT-DAT (C) with database growth from 10K to 500M molecules. The number of shared molecules at different thresholds (1K to 50M) of randomized docking data does not increase as databases grow.



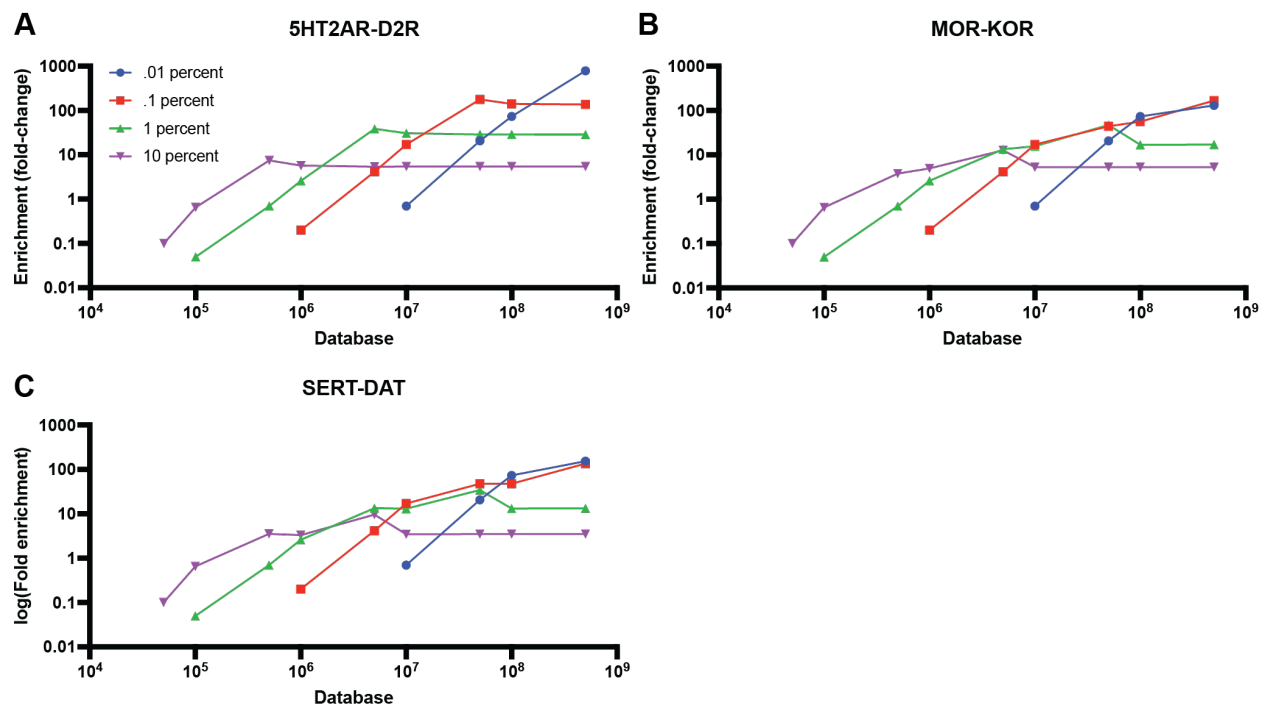
**Figure 5.2.** The number of shared top-ranked molecules increase with database size. Polypharmacology model systems 5HT<sub>2A</sub>R-D<sub>2</sub>R (A),  $\mu$ OR- $\kappa$ OR (B), and SERT-DAT (C) with database growth from 10K to 500M molecules. The number of shared molecules at different thresholds (1K to 50M) of top-ranked docking data increases with database growth.



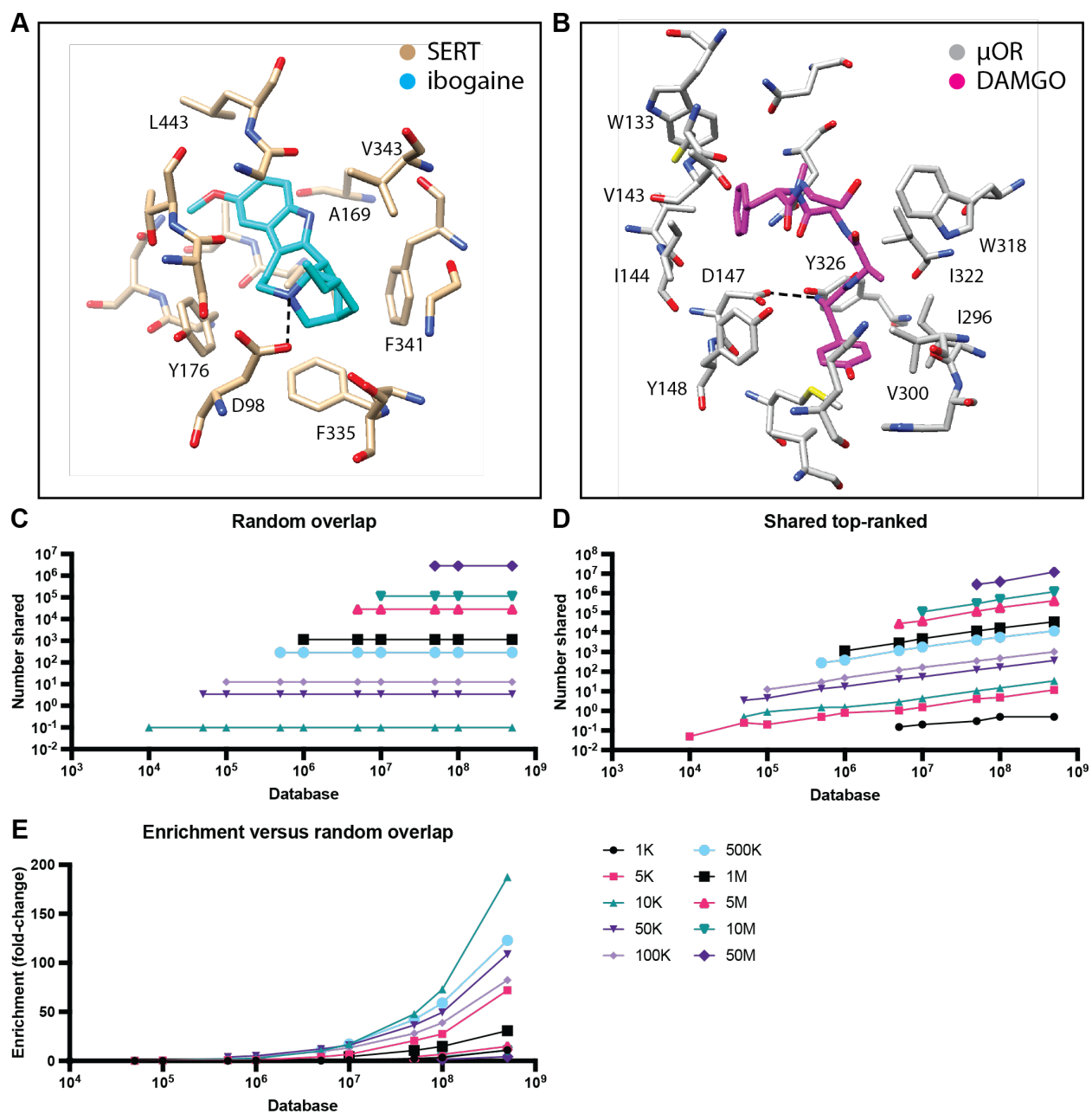
**Figure 5.3. Enrichment of shared molecules is greatest at mid-top-ranked docking results.** Polypharmacology model systems 5HT<sub>2A</sub>R-D<sub>2</sub>R (A),  $\mu$ OR- $\kappa$ OR (B), and SERT-DAT (C) number of shared molecules enriched over random overlap increases as databases grow. The largest increase is at the 10K to 500K top-ranked molecules.



**Figure 5.4. Docking results in the top-ranked percent thresholds increases with database size.** Polypharmacology model systems 5HT<sub>2A</sub>R-D<sub>2</sub>R (A),  $\mu$ OR-KOR (B), and SERT-DAT (C) with the percent shared of top-ranked molecules in different thresholds of percent (0.01% to 10%) increases with database growth.



**Figure 5.5. Enrichment of shared molecules over random reaches a plateau at different thresholds of percent top-ranked results.** Polypharmacology model systems 5HT<sub>2A</sub>R-D<sub>2</sub>R (A),  $\mu$ OR- $\kappa$ OR (B), and SERT-DAT (C) show increased enrichment of shared top-ranked molecules compared to random until a plateau is reached with each subset. Raw data located in Table 5.S2.



**Figure 5.6. Database growth analysis with prospective polypharmacology pair SERT- $\mu$ OR.** Binding sites of experimental determined structures for SERT bound to ibogaine (A) and  $\mu$ OR bound to DAMGO (B). Residues within 5Å are shown as sticks. Salt bridge indicated with dashed line in both protein-ligand complex. (C,D,E) Polypharmacology analysis with increasing databases show similar trends as model systems.

## 5.6 Tables

**Table 5.1. Polypharmacology model systems.**

Protein pair	Ligand with polypharmacology	Binding site identify (%)	Binding site similarity (%)	Protein similarity
5HT2AR D2R	Risperidone, atypical antipsychotic	62	77	60
MOR KOR	Morphine, analgesic (or Bup)	74	96	69
SERT DAT	Duloxetine, antidepressant	83	92	74
DHFR TS	Methotrexate, chemotherapy	44	67	na
EGFR HER2	Lapatinib, chemotherapy	94	100	63
HDAC2 HDAC8	Vorinostat, chemotherapy	70	88	66

na, not applicable



**Table 5.S1. Enrichment of shared molecules linear regression results.**

<b>Model system</b>	<b>Regression Best-fit Values</b>	<b>Top 5M</b>	<b>Top 10M</b>	<b>Top 50M</b>
<b>5HT2AR-D2R</b>	Slope	0.001596	0.003848	0.02880
	Y-intercept	56869	160581	1542416
<b>SERT-DAT</b>	Slope	0.0008921	0.002364	0.02838
	Y-intercept	72543	239130	1536024
<b>MOR-KOR</b>	Slope	0.0007080	0.001768	0.01632
	Y-intercept	52455	179508	2458835

**Table 5.S2. Fold enrichment of top-ranked percent thresholds over random.**

<b>5HT2AR-D2R</b>				
<b>Database size</b>	<b>Top-ranked threshold</b>			
	<b>Top 0.01%</b>	<b>Top 0.1%</b>	<b>Top 1%</b>	<b>Top 10%</b>
10K	na	na	na	0.0
50K	na	na	na	0.1
100K	na	na	0.1	0.7
500K	na	na	0.7	7.5
1M	na	0.2	2.6	5.7
5M	na	4.2	38.4	5.4
10M	0.7	17.0	30.5	5.4
50M	20.6	177.1	28.8	5.4
100M	72.9	141.0	28.8	5.4
500M	779.4	136.0	28.7	5.4
<b>MOR-KOR</b>				
<b>Database size</b>	<b>Top-ranked threshold</b>			
	<b>Top 0.01%</b>	<b>Top 0.1%</b>	<b>Top 1%</b>	<b>Top 10%</b>
10K	na	na	na	0.0
50K	na	na	na	0.1
100K	na	na	0.1	0.7
500K	na	na	0.7	7.5
1M	na	0.2	2.6	5.7
5M	na	4.2	38.4	5.4
10M	0.7	17.0	30.5	5.4
50M	20.6	177.1	28.8	5.4
100M	72.9	141.0	28.8	5.4
500M	779.4	136.0	28.7	5.4
<b>SERT-DAT</b>				
<b>Database size</b>	<b>Top-ranked threshold</b>			
	<b>Top 0.01%</b>	<b>Top 0.1%</b>	<b>Top 1%</b>	<b>Top 10%</b>
10K	na	na	na	0.0
50K	na	na	na	0.1
100K	na	na	0.1	0.7
500K	na	na	0.7	7.5
1M	na	0.2	2.6	5.7
5M	na	4.2	38.4	5.4
10M	0.7	17.0	30.5	5.4
50M	20.6	177.1	28.8	5.4
100M	72.9	141.0	28.8	5.4
500M	779.4	136.0	28.7	5.4

na, not applicable

## 5.7 Materials and Methods

### Model system characterization

For protein binding site identity and binding site similarity, the pair of receptors were aligned and residues within 5 Å of the antagonists were used for calculations. Binding site identity was determined by the number of residues identical/total binding site residues. Binding site similarity classified residues into the following groups: 1) nonpolar, aliphatic amino acids: Gly, Ala, Val, Leu, Met, Ile; 2) polar, uncharged amino acids: Ser, Thr, Cys, Pro, Asn, Gln; 3) aromatic residues: Phe, Tyr, Trp; 4) positively charged: Lys, Arg, His; and 5) negatively charged residues: Asp, Glu. The percent similarity was determined by the number of similar residues/total binding site residues. BLAST (20) of the two protein sequences was used to determine overall protein similarity.

### Molecular docking

The 5HT<sub>2A</sub>R with risperidone (PDB code 6A93) (21), D<sub>2</sub>R with haloperidol (PDB code 6LUQ) (22), and DAT with cocaine (PDB 4XPB) (23) were used for docking calculations. SERT (10), μOR and κOR (24) setups were used from previous work. Receptors were protonated with REDUCE (25) for DAT and in Maestro (version 2019-4, Schrödinger, Inc.) for D<sub>2</sub>R and 5HT<sub>2A</sub>R; the latter two were also minimized in Maestro Protein Preparation Wizard protocol using OPLS4 force field. Control calculations used IUPHAR-BPS database (26) annotated antagonists and property-matched decoys (11), and were optimized by logAUC and quality of docked poses, as described before (27). 900 M ZINC22 (16) screen using DOCK3.8 were performed for all 6 receptors with 613 to 840 million scoring on each receptor. All molecules that received DOCK scores were used for database simulations.

## **Retrospective Simulations**

The docking results for each receptor were shuffled 20 times. From those, subsets of 10,000 to 5 million molecules were selected to simulate database growth. The ‘random subsets’ for each database were selected at random for 1,000 to 50 million (if possible). Each database was sorted by DOCK scores (most negative = top-ranked). The top-ranked subsets were selected for 1,000 to 50 million molecules (is possible). Only up to the top-ranked 10% is evaluated, except for the 500 million database, where up to 20% is used to enable linear regressions of the log-log plots. All of this was done for each single receptor. Then overlap in the results for the pairs of receptors were identified, SERT-DAT, 5HT<sub>2A</sub>R-D<sub>2</sub>R,  $\mu$ OR- $\kappa$ OR, and prospective pair SERT- $\mu$ OR. Simple linear regressions were run on data in log-log plots to determine the slopes for comparison of different datasets using GraphPad Prism (v. 9.1.1).

## **Code availability**

DOCK3.8 is freely available for non-commercial research; commercial licenses available from the UC Regents. Open-source web-based DOCK is located at <http://blaster.docking.org>.

## **5.8 Acknowledgements**

This work was supported by This work was supported by DARPA grant HR0011-19-2-0020 (B.K.S., J.J.I.) and by NIH grant R35GM122481 (B.K.S., J.J.I.). We thank OpenEye Software for Omega and related tools, and Schrodinger, Inc. for the Maestro package.

## **5.9 Author Contributions**

E.A.F. prepared the 5HT<sub>2A</sub>R and D<sub>2</sub>R setups, conducted the model system analysis, and database simulations and analysis. L.S. prepared the DAT docking setup. Y.W. assisted in model system setup preparation and analysis. All was conducted under the supervision and guidance of J.J.I. and B.K.S. E.F. wrote the paper with input from all authors. B.K.S. conceived the project.

## 5.10 References

1. A. L. Hopkins, Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690 (2008).
2. B. L. Roth, D. J. Sheffler, W. K. Kroeze, Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* **3**, 353–359 (2004).
3. Z. A. Knight, H. Lin, K. M. Shokat, Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer.* **10**, 130–137 (2010).
4. M. A. M. Subbaiah, Triple Reuptake Inhibitors as Potential Therapeutics for Depression and Other Disorders: Design Paradigm and Developmental Challenges. *J. Med. Chem.* **61**, 2133–2165 (2018).
5. A. R. de Lera, A. Ganesan, Epigenetic polypharmacology: from combination therapy to multitargeted drugs. *Clin. Epigenetics.* **8**, 105 (2016).
6. P. Fallon, S. M. Dursun, A naturalistic controlled study of relapsing schizophrenic patients with tardive dyskinesia and supersensitivity psychosis. *J. Psychopharmacol. Oxf. Engl.* **25**, 755–762 (2011).
7. H. Pathan, J. Williams, Basic opioid pharmacology: an update. *Br. J. Pain.* **6**, 11–16 (2012).
8. M. Schuller, G. J. Correy, S. Gahbauer, D. Fearon, T. Wu, R. E. Díaz, I. D. Young, L. Carvalho Martins, D. H. Smith, U. Schulze-Gahmen, T. W. Owens, I. Deshpande, G. E. Merz, A. C. Thwin, J. T. Biel, J. K. Peters, M. Moritz, N. Herrera, H. T. Kratochvil, QCRG Structural Biology Consortium, A. Aimon, J. M. Bennett, J. Brandao Neto, A. E. Cohen, A. Dias, A. Douangamath, L. Dunnett, O. Fedorov, M. P. Ferla, M. R. Fuchs, T. J. Gorrie-Stone, J. M. Holton, M. G. Johnson, T. Krojer, G. Meigs, A. J. Powell, J. G. M. Rack, V. L.

- Rangel, S. Russi, R. E. Skyner, C. A. Smith, A. S. Soares, J. L. Wierman, K. Zhu, P. O'Brien, N. Jura, A. Ashworth, J. J. Irwin, M. C. Thompson, J. E. Gestwicki, F. von Delft, B. K. Shoichet, J. S. Fraser, I. Ahel, Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).
9. E. A. Fink, J. Xu, H. Hübner, J. M. Braz, P. Seemann, C. Avet, V. Craik, D. Weikert, M. F. Schmidt, C. M. Webb, N. A. Tolmachova, Y. S. Moroz, X.-P. Huang, C. Kalyanaraman, S. Gahbauer, G. Chen, Z. Liu, M. P. Jacobson, J. J. Irwin, M. Bouvier, Y. Du, B. K. Shoichet, A. I. Basbaum, P. Gmeiner, Structure-based discovery of nonopioid analgesics acting through the  $\alpha$ 2A-adrenergic receptor. *Science*. **377**, eabn7065 (2022).
10. I. Singh, A. Seth, C. B. Billesbølle, J. Braz, R. M. Rodriguiz, K. Roy, B. Bekele, V. Craik, X.-P. Huang, D. Boytsov, P. Lak, H. O'Donnell, W. Sandtner, B. L. Roth, A. I. Basbaum, W. C. Wetsel, A. Manglik, B. K. Shoichet, G. Rudnick, "Structure-based Discovery of Conformationally Selective Inhibitors of the Serotonin Transporter" (preprint, *Biochemistry*, 2022), , doi:10.1101/2022.06.13.495991.
11. R. M. Stein, Y. Yang, T. E. Balius, M. J. O'Meara, J. Lyu, J. Young, K. Tang, B. K. Shoichet, J. J. Irwin, Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model.* **61**, 699–714 (2021).
12. J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, Ultra-large library docking for discovering new chemotypes. *Nature*. **566**, 224–229 (2019).
13. A. Alon, J. Lyu, J. M. Braz, T. A. Tummino, V. Craik, M. J. O'Meara, C. M. Webb, D. S. Radchenko, Y. S. Moroz, X.-P. Huang, Y. Liu, B. L. Roth, J. J. Irwin, A. I. Basbaum, B. K.



- Shoichet, A. C. Kruse, Structures of the  $\sigma_2$  receptor enable docking for bioactive ligand discovery. *Nature*. **600**, 759–764 (2021).
14. A. Manglik, H. Lin, D. K. Aryal, J. D. McCorvy, D. Dengler, G. Corder, A. Levit, R. C. Kling, V. Bernat, H. Hübner, X.-P. Huang, M. F. Sassano, P. M. Giguère, S. Löber, Da Duan, G. Scherrer, B. K. Kobilka, P. Gmeiner, B. L. Roth, B. K. Shoichet, Structure-based discovery of opioid analgesics with reduced side effects. *Nature*. **537**, 185–190 (2016).
15. D. R. Weiss, J. Karpiak, X.-P. Huang, M. F. Sassano, J. Lyu, B. L. Roth, B. K. Shoichet, Selectivity Challenges in Docking Screens for GPCR Targets and Antitargets. *J. Med. Chem.* **61**, 6830–6845 (2018).
16. B. Tingle, K. Tang, J. Castanon, J. Gutierrez, M. Khurelbaatar, C. Dandarchuluun, Y. Moroz, J. Irwin, “ZINC-22 - A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery” (preprint, Chemistry, 2022), , doi:10.26434/chemrxiv-2022-82czl.
17. J. Lyu, J. Irwin, B. Shoichet, “Modeling the expansion of virtual screening libraries” (preprint, Chemistry, 2022), , doi:10.26434/chemrxiv-2022-6lv34-v2.
18. R. Arakawa, A. Takano, C. Halldin, Serotonin and Norepinephrine Transporter Occupancy of Tramadol in Nonhuman Primate Using Positron Emission Tomography. *Int. J. Neuropsychopharmacol.* **22**, 53–56 (2019).
19. D. M. Marks, M. J. Shah, A. A. Patkar, P. S. Masand, G.-Y. Park, C.-U. Pae, Serotonin-norepinephrine reuptake inhibitors for pain control: premise and promise. *Curr. Neuropharmacol.* **7**, 331–336 (2009).
20. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

21. K. T. Kimura, H. Asada, A. Inoue, F. M. N. Kadji, D. Im, C. Mori, T. Arakawa, K. Hirata, Y. Nomura, N. Nomura, J. Aoki, S. Iwata, T. Shimamura, Structures of the 5-HT<sub>2A</sub> receptor in complex with the antipsychotics risperidone and zotepine. *Nat. Struct. Mol. Biol.* **26**, 121–128 (2019).
22. L. Fan, L. Tan, Z. Chen, J. Qi, F. Nie, Z. Luo, J. Cheng, S. Wang, Haloperidol bound D<sub>2</sub> dopamine receptor structure inspired the discovery of subtype selective ligands. *Nat. Commun.* **11**, 1074 (2020).
23. K. H. Wang, A. Penmatsa, E. Gouaux, Neurotransmitter and psychostimulant recognition by the dopamine transporter. *Nature.* **521**, 322–327 (2015).
24. C. M. Webb, F. Hetzer, B. K. Shoichet, A. Manglik, Novel  $\mu$ OR agonists with tuned pharmacology, signaling and in vivo efficacy (In preparation).
25. J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation 1 Edited by J. Thornton. *J. Mol. Biol.* **285**, 1735–1747 (1999).
26. S. D. Harding, J. F. Armstrong, E. Faccenda, C. Southan, S. P. H. Alexander, A. P. Davenport, A. J. Pawson, M. Spedding, J. A. Davies, NC-IUPHAR, The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acids Res.* **50**, D1282–D1294 (2022).
27. B. J. Bender, S. Gahbauer, A. Lutten, J. Lyu, C. M. Webb, R. M. Stein, E. A. Fink, T. E. Balius, J. Carlsson, J. Irwin, B. K. Shoichet, A Practical Guide to Large-Scale Docking. *Nat Protoc* (2021), doi:10.1038/s41596-021-00597-z.

## 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.

DocuSigned by:

*Elissa Fink*

0C9D23E337774AE...

\_\_\_\_\_  
Author Signature

12/13/2022

\_\_\_\_\_  
Date