

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Understanding virtual solvent through large-scale ligand discovery

**Permalink**

<https://escholarship.org/uc/item/3d8134c2>

**Author**

Stein, Reed

**Publication Date**

2020

Peer reviewed|Thesis/dissertation

Understanding virtual solvent through large-scale ligand discovery

by  
Reed Stein

DISSERTATION

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

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

*Brian K. Shoichet*

C8380A8942D641D...

Brian K. Shoichet

Chair

DocuSigned by:

*Matthew P. Jacobson*

DocuSigned by:

*Michael Grabe*

5B8B990D587B414...

Matthew P. Jacobson

Michael Grabe

Committee Members



## Acknowledgments

I would like to thank everyone that helped me throughout my graduate years. I have been very fortunate to learn from, work with, and mentor some truly brilliant people. First, I would like to thank my advisor, Brian Shoichet. His steadfast leadership, high expectations, and undaunted approach to asking hard questions enabled me to develop as a scientist. He gave me the freedom to fail and learn from my mistakes but was always available if I needed guidance. During those few times when I believed I had made a fatal mistake, completely sabotaged my project, and was too far gone to start over or continue, Brian convinced me that there was still hope, that things could be salvaged. One of my favorite communications with Brian was an e-mail that he sent me after I told him I was hesitant to speak at the annual UCSF QBC retreat:

“Apropos of speaking at the retreat, do you know the scene in Chapter 1 of *The Hobbit*, when Gandalf is talking to Bilbo:

‘For your old grandfather Took’s sake, and for the sake of poor Belladonna, I will give you what you asked for.’

‘I beg your pardon, I haven’t asked for anything!’

‘Yes, you have! Twice now. My pardon. I give it you. In fact I will go so far as to send you on this adventure.’

...Speaking in public is a big part of science; learning how to do it is simply part of your education. Seriously.”

It wasn’t until rereading the book that I understood the power of this passage. If you joined the Shoichet lab, you agreed to go on an adventure, to step outside of your



comfort zone, to challenge yourself, and this was because, to quote Gandalf again from *The Hobbit*:

“There is a lot more in him than you guess, and a deal more than he has any idea of himself.”

Brian believes in the people in his lab, even when they don't believe in themselves; even when they don't know what they are capable of. As a frightened first year graduate student with little programming and docking experience, Brian believed that I could contribute in his lab. Multiple people told me that I would learn how to think like a scientist if I joined the Shoichet lab, and I believe this approach to be a main reason why. I also thank Brian for his way with words. The following quote from Brian in an early morning model systems meeting sums up everything I learned about docking in my PhD:

“Every ligand has a story to tell. We just flip through the pages too quickly to listen.”

I would also like to thank my thesis committee, Matt Jacobson and Michael Grabe, who both contributed a rigorous physical perspective to my thesis meetings that I really appreciated and tried to incorporate as best I could in my work. I had multiple meetings with both of them, discussing electrostatics and the Poisson-Boltzmann equation, water thermodynamics, the nuts and bolts of the DOCK scoring function, and my career trajectory, all of which were extremely helpful.

Multiple people in the Shoichet lab helped me along the way, most notably my mentors, Trent Balias, who taught me how to dock, and Marcus Fischer, who taught me how to setup crystal trays. run binding assays, but also how important practical jokes are in the lab. Both were extremely supportive and were always willing to listen to any problems I had and offer advice. I consider them friends to this day. I thank Trent for his infinite patience in answering and explaining every dumb question I asked and for entertaining all my crazy ideas of how to make DOCK better. I'll always remember "board time", "water time" in which we discussed even more crazy DOCK ideas that we would never have the time to implement, and all the times he accused me of "delving too greedily and too deep" into the DOCK code when I tried to understand why certain problems arose during my graduate work. Remember, Trent:

"We are the music makers, and we are the dreamers of dreams."

I thank all other members of the Shoichet lab who have helped me over the years. Gabe Rocklin's excellent advice solidified my decision to attend UCSF and to join the Shoichet lab. Allison Doak's extremely honest and unbiased advice early in my graduate career, prepared me for the Shoichet lab, and has proven true repeatedly in the past 5+ years. I thank John Irwin, whose comments on the DUDE-Z benchmarking pipeline, helped me strengthen and make it more powerful, whose ligand building contributions to the lab ensured that all our projects are impactful, and whose contrarian-like comments helped me see things from a different perspective. Magdalena Korczynska, who, like Brian, was never quite satisfied with my answers as a first- and second-year graduate student, motivated me to read more, learn more, and ask myself

more challenging questions. She also taught me what was important in viewing protein-ligand interactions, for which I am extremely grateful. I thank Anat Levit, for always being willing to listen to any problems I had, scientific or otherwise, and offer advice, for her GPCR expertise that was always useful, and for offering her brilliant eye to any docking pose questions I had. Thank you to Inbar Fish for getting me set up in the experimental lab, and for showing me the rewarding and powerful lessons that the model systems could teach. Thank you to Josh Pottel, an awesome scientist who asked difficult questions and made the lab a fun place to be. I always looked forward to his group meetings to understand what research problem he was thinking about. Thank you to Xiaobo Wan for his never-ending laugh that brightened up the lab, and at unexpected times, relating very interesting scientific insights that I hadn't considered. I thank Jiankun Lyu for all the amazing science (methods development, GPCR, kinase, what have you) discussions we've had over the years, and like Trent, always entertaining a good, crazy idea for DOCK. I thank him for always sharing code, always lending a helping hand, and for his positivity. I thank Isha Singh for teaching me everything about AmpC  $\beta$ -lactamase – how to inhibit it, how to crystallize it, and how to shoot X-rays at it. I enjoyed the experimental component of my PhD more than I thought I would, and she was always willing to let me get as involved as I wanted. Thank you to Stefan Gahbauer, for his awesome questions and curiosity about docking and finding new ligands. I really enjoyed all our discussions about the DOCK code, what parameters to use, how things could go very wrong or very right with a simple parameter choice, and how deep the DOCK rabbit hole goes. Thank you to Ying Yang, for teaching me more about water thermodynamics, how powerful coding can be in chemistry, for always

lending a helping hand when I ran into molecular dynamics problems, and for being an excellent collaborator. I thank Parnian Lak and Yvonne Munchua for keeping the lab running smoothly, and making any problem I had, experimental or administrative, seem miniscule as they had it completely under control. Thank you to Elissa Fink for continuing to use AmpC as a model system, and for taking it to new heights with the first years. Lastly, I thank Chase Webb and Tia Tummino, for being awesome mentees, and being patient with me as I have tried to learn how to teach and mentor. I thank them for bringing in a fresh perspective and challenging my beliefs on docking, GPCRs, and the scientific approach. I hope they have learned as much from me as I have from them.

I thank our collaborators from outside UCSF including Grant Glatfelter, Anthony Jones, and Margarita Dubocovich from SUNY Buffalo, as well as Hye-Jin Kang, John McCorvy, Sam Slocum, Tao Che, XP Huang, Terry Kenakin, and Bryan Roth from UNC Chapel Hill. Additionally, I would like to thank Mossa Ghattas and Tom Kurtzman from CUNY Lehman, Ben Stauch, Linda Johansson, and Vadim Cherezov from USC, Christos Tsoutsouvas and Alexandros Makriyannis from Northeastern, and Georgios Skiniotis and Brian Kobilka from Stanford. This thesis wouldn't have been as exciting, impactful, or biologically meaningful without their amazing hard work and contributions.

I want to thank my graduate school mentors and friends. This includes our amazing former PSPG director, Deanna Kroetz, as well as Patsy Babbitt and several former graduate students including Adrian Stecula, John Bruning, Chelsea Hosey, Kyle DeFrees, Ken Hallenbeck, and Michael Martin. Thank you for your guidance, advice, and compassion. Thank you to Ilan "Jaxson" Chemmama, my first friend at UCSF, who has consistently offered great advice on science, graduate school, and bar, dining, and

hiking options. Thank you to my PSPG cohort, especially Christine Bowman, Megan Lo, and Darya Cheng for all the fun dinners, hikes, and drinks we've shared. Thank you to Ben Wong for our awesome, yet short-lived, tradition of going to Terror Tuesday every week at Alamo Drafthouse. We saw some real gems. And to all my fellow students and post-docs that I would randomly run into, or mainly saw at beer hour, recruitment, or the QBC retreat, and that helped me realize that there was more to graduate school than just work, thank you. This includes Ryan Muir, Clint Cario, Garrett Gaskins, Nima Emami, Seth Axen, Cole Helsell, Paul Thomas, and Joe Lobel, among many others.

I thank all the science teachers and professors that inspired or encouraged me to go to graduate school. This includes Maureen Leith and Heather Gebauer from Salpointe Catholic High School, and Thomas Poon, Emily Wiley, Jennifer Armstrong, Aaron Leconte, and Scot Gould from the Keck Science Department at the Claremont Colleges. I also thank Andrew Bordner for providing my first opportunity in docking, and for jumpstarting my interest in programming, chemistry, physics, molecular biology, and their intersection.

Thank you to Dr. Smudge E. Mo and Rachel Brunetti for supporting me in both the best and worst times of my graduate career. It would've been infinitely more difficult for me if you two weren't there for me throughout this long, strange trip.

Lastly, I want to thank my family, Stuart, Kathy, Steve, Jeff, and AJ, who, though they didn't always completely understand what I was doing, fully supported me and believed in the wild things I was doing on water, melatonin, and cannabinoids. Thank you to my father, Stuart, the true Dr. Stein, who instilled in me a passion for science early and planted the seeds of molecular docking and protein-ligand interactions into my

head all those years ago. Thank you to my mother, Kathy, for listening to my weird research ideas and for always keeping me on track with well-timed questions of when I would graduate and what I would do after. They helped me stay focused. I wouldn't have achieved anything near as much if it weren't for their continued support and love.

## Contributions

Chapter 1 of this thesis is a variation of the material as it appears in:

Balius, T.E.\*; Fischer, M.\*; Stein, R.M.; Adler, T.B.; Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T.; Shoichet, B.K., Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery. *Proc. Natl. Acad. Sci. U.S.A.*, **2017**, 114 (33) E6839-E6846. Doi: 10.1073/pnas.1703287114.

\*these authors contributed equally

Chapters 2 and 3 of this thesis were adapted from manuscripts in preparation:

Stein, RM; Singh, I; Ghattas, M; Kurtzman, T; Balius, TE; Shoichet, BK. Testing a faster Gaussian-based implementation of Inhomogeneous Solvation Theory in Large Scale Ligand Discovery.

Stein, RM; Yang, Y; Balius, TE; O'Meara, MJ; Lyu, JK; Young, J; Tang, K; Irwin, JJ; Shoichet, BK. Property-unmatched decoys in molecular docking.

Chapter 4 of this thesis is a variation of the material as it appears in:

Stein, RM\*; Kang, HJ\*; McCorvy, JD\*; Glatfelter, GC\*; Jones, AJ; Che, T; Slocum, S; Huang, XP; Savych, O; Moroz, YS; Stauch, B; Johansson, LC; Cherezov, V; Kenakin, T; Irwin, JJ; Shoichet, BK; Roth, BL; Dubocovich, ML. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* 579, 609–614 (2020).

\*these authors contributed equally

“They went to work designing computer molecules and computer brains.”

– Don DeLillo, *White Noise*



# **Abstract**

## **Understanding virtual solvent through large-scale ligand discovery**

Reed Stein

Predicting new ligands and their binding poses for a protein target relies on an understanding of the physical forces that exist between the water-submerged protein and ligand. The relative favorability of these molecular and atomic interactions between the protein and ligand compared with their interactions with water determine the binding affinity, which in turn can be converted into a binding free energy. Protein-ligand binding energetics are, with varying levels of success, encoded into scoring functions, which at their best, can only partially emulate the true binding affinity of a protein-ligand binding event. In the context of virtually screening millions or hundreds of millions of drug-like ligands, molecular docking algorithms take advantage of scoring functions to rank the binding energies of these molecules relative to one another to help prioritize the most promising ligands.

The focus of this dissertation is the balance between scoring function energy terms with an emphasis on water energetics, specifically the desolvation of the protein upon ligand binding. It is thought that our limited understanding of water is largely responsible for our limitations in discovering and designing drugs. This is due to the large number of roles that water can play, as well as its significant, and even dominant, contribution to protein-ligand binding energetics, which in the realm of molecular docking, is typically under-modeled or completely neglected.

First, I focus on the incorporation of receptor desolvation into the standard DOCK3.7 scoring function to more accurately model protein-ligand binding interactions by including further contributions of water. This is the original implementation of Grid Inhomogeneous Solvation Theory applied to the model cavity, cytochrome *c* peroxidase, and spearheaded by Trent Baliaus and Marcus Fischer. Second, I discuss an extension of GIST in DOCK3.7, a new implementation that relies on pre-computed Gaussian-weighted GIST receptor desolvation enthalpies. This results in negligible slowdown of the standard DOCK3.7 scoring function, similar performance to the original implementation of GIST, and the identification of new ligands for the drug-like model system, AmpC  $\beta$ -lactamase. The work on receptor desolvation contained within these two chapters inspires the name of this thesis, and were started in my rotation and have continued until the end. Third, I focus on the use of property-matched and property-unmatched decoys for use in retrospective enrichment calculations prior to running a large-scale molecular docking virtual screen. Decoy molecules share the same physical properties as ligands that bind a protein but are topologically dissimilar to ensure that they do not actually bind the protein. What we found was that charge mismatching between ligands and decoys could bias one's docking setup towards artifactually strong performance. Chapter 3 focuses on how we both decreased and increased the property space of decoys relative to ligands to safeguard against these docking setup biases. Fourth, I employ this knowledge of protein-ligand binding affinities to identify novel selective melatonin receptor ligands that are active in *in vivo* circadian rhythm assays. Finally, I discuss my current project on the CB<sub>1</sub> cannabinoid receptor in the context of analgesia, followed by future directions.

# Table of Contents

<b>Chapter 1: Testing IST in Structure-Based Ligand Discovery .....</b>	<b>19</b>
<b>References .....</b>	<b>49</b>
<b>Chapter 2: Testing a faster implementation of IST in Ligand Discovery.....</b>	<b>61</b>
<b>References .....</b>	<b>101</b>
<b>Chapter 3: Property-unmatched decoys in docking benchmarks .....</b>	<b>114</b>
<b>References .....</b>	<b>149</b>
<b>Chapter 4: Virtual discovery of MT receptor ligands to modulate circadian rhythms.....</b>	<b>160</b>
<b>References .....</b>	<b>178</b>
<b>Chapter 5: Large-scale docking on the CB<sub>1</sub> cannabinoid receptor .....</b>	<b>208</b>
<b>References .....</b>	<b>223</b>
<b>Chapter 6: Future Directions.....</b>	<b>229</b>
<b>References .....</b>	<b>236</b>
<b>Appendix A: Supplementary Figures and Tables .....</b>	<b>237</b>

## List of Figures

Figure 1.1. Receptor desolvation using GIST.....	25
Figure 1.2. Comparison of GIST and non-GIST screens.....	32
Figure 1.3. Comparison of experimental and predicted binding poses. ....	38
Figure 1.4. Three representative ligand binding curves.....	39
Figure 2.1. Scheme for incorporating grid inhomogeneous solvation theory.....	68
Figure 2.2. Comparing experiment to GIST-predicted hydration sites. ....	73
Figure 2.3. Comparison of large-scale docking molecule ranks.....	75
Figure 2.4. Comparison of pro-bGIST, anti-bGIST, and pose-changing molecules. ....	76
Figure 2.5. Representative inhibition curves for AmpC inhibitors.....	77
Figure 2.6. Crystallography of pose-changing molecules. ....	82
Figure 2.7. Comparison of GIST desolvation and reorganization enthalpies ....	83
Figure 3.1. Ligand desolvation and electrostatics weights alter enrichment.....	125
Figure 3.2. Proportion of charged molecules in DUD-E sets affects enrichment.....	128
Figure 3.3. Enrichment comparison between DUD-E and DUDE-Z.....	132
Figure 3.4. Enrichments and charge priority of DUDE-Z and Extrema.....	135
Figure 3.5: Enrichments and charge priority of DUDE-Z, Extrema, and Goldilocks. ....	139
Figure 3.6. Bootstrapping enrichment differences using different decoy backgrounds.....	143
Figure 4.1. Large library docking finds novel, potent melatonin receptor ligands. ....	164
Figure 4.2. Affinity, efficacy, and potency of MT <sub>1</sub> -selective inverse agonists ....	169
Figure 4.3. MT <sub>1</sub> -selective inverse agonists behave as agonists and inverse agonists ....	174

Figure 5.1. Comparison of properties of predicted and known CB1 ligands.....	214
Figure 5.2. Poses and functional dose response curves of novel ligands.....	215
Figure A.1.1. GIST Combinations.....	239
Figure A.1.2. GIST in docking is a good approximation.....	242
Figure A.1.3. Comparison of GIST combinations. ....	244
Figure A.1.4. GIST Weighting Factors.....	246
Figure A.1.5. Comparison of GIST grids from sub-trajectories. ....	248
Figure A.1.6. Enrichment analysis of CcP-ga and 25 DUD-E systems. ....	251
Figure A.1.7. Hydration of CcP-ga with the GIST enthalpy grid. ....	253
Figure A.1.8. Compound 14 with MES. ....	257
Figure A.1.9. Ligand binding curves. ....	258
Figure A.2.1. Correlations between GIST energies. ....	267
Figure A.2.2. Insufficient minimization scrambles best scoring poses. ....	269
Figure A.2.3. A new scoring scheme fixes insufficient minimization. ....	271
Figure A.2.4. Choosing molecules similar to known AmpC inhibitors ....	273
Figure A.2.5. Volume occupation of pro- and anti-bGIST molecules ....	274
Figure A.2.6. Parameter and solvent choice do not affect rank changing molecules. ....	274
Figure A.3.1. Examples of bootstrapping enrichment distribution. ....	286
Figure A.3.2. Bootstrapping on Binders/Nonbinders.....	287
Figure A.3.3. Bootstrapping Enrichment Differences.....	288
Figure A.3.4. Bootstrapping statistics for all 43 systems.....	290

Figure A.4.1. Concentration-response curves of initial 15 compounds. ....	296
Figure A.4.2. Concentration-response curves of interesting analogs.....	297
Figure A.4.3. Small ligand changes have large effects on activity and selectivity. ....	298
Figure A.4.4. MT <sub>1</sub> -selective inverse agonists decelerate re-entrainment rate <i>in vivo</i> . ....	299
Figure A.4.5. MT <sub>1</sub> -selective inverse agonists phase advance circadian activity at MT <sub>1</sub> .....	300
Figure A.4.6. Concentration-response curves of the inverse agonists .....	301
Figure A.4.7. Phase shift profiles of '7447, melatonin, and luzindole.....	302
Figure A.4.8. PRESTO-Tango GPCRome & off-target screening.....	306
Figure A.4.9. Dose-response curves for off-target receptors. ....	307
Figure A.4.10. Competition binding of inverse agonists against melatonin receptors.....	309
Figure A.4.11. Affinity, efficacy, and potency of MT <sub>2</sub> -selective agonist. ....	311

## List of Tables

Table 1.1. New candidate CcP ligands.....	34
Table 2.1. Adjusted logAUC values comparing GIST performance .....	70
Table 2.2. A selection of binding molecules. ....	80
Table 3.1. Enrichments for DOCK3.7 Scoring Coefficients over 43 Targets .....	126
Table 3.2: Ligand and Decoy Properties for 43 Protein Targets .....	130
Table 3.3: Average Enrichment log AUC values for Different Decoy Sets .....	132
Table 5.1. Active molecules from single-point radioligand displacement assay. ....	216
Table A.1.1. Comparison of GIST combinations.....	245
Table A.1.2. CcP-ga retrospective analysis for GIST weight. ....	247
Table A.1.3. Impact of modified sampling and subtrajectory on enrichment .....	249
Table A.1.4. DUD-E evaluation of GIST contribution on enrichment calculations. ....	250
Table A.1.5. Site energetics of subregions. ....	254
Table A.1.6. Detailed properties of selected molecules. ....	256
Table A.1.7. Ligand occupancies after automatic refinement. ....	259
Table A.1.8. Comparison of affinities for compounds with different interactions .....	259
Table A.1.9. DOCK3.7 run time slowdown with GIST referenced to non-GIST.....	260
Table A.1.10. CcP-ga and DUD-E simulation details.....	264
Table A.2.1. All molecules tested against AmpC.....	276
Table A.4.1. Active molecules from the initial docking screen. ....	291
Table A.4.2. Some of the potent analogs from initial hits.....	293

Table A.4.3. Pharmacokinetics of three melatonin receptor type-selective ligands.....	294
Table A.4.4: Probe pairs of <i>in vivo</i> tested molecules .....	295
Table A.4.5. Purity information of potent MT <sub>1</sub> /MT <sub>2</sub> compounds & probe pairs .....	303
Table A.4.6. Biased Analogs .....	308



## Introduction

In college, I was captivated by the interactions and reactions that occurred between molecules in my organic chemistry class. This led me to two summers at Mayo Clinic Scottsdale, where I was doing basic programming, molecular modeling, docking with Molsoft's ICM software, and judging docking performance by ICM's ability to reproduce crystallographic binding poses of ligands. My project was focused on judging the success of docking calculations after the incorporation of receptor flexibility, and was inspired by work done in the Abagyan lab<sup>1</sup>. I remember reading their work and attempting to understand such terms as "Biased Probability Monte Carlo stochastic optimizer", "0.5 Å spacing potential grid maps", "6-12 Lennard Jones potential", and "distance dependent dielectric constant". I felt I had only gotten a preview of the molecular modeler's techniques, and was determined to understand more, regardless of the difficulty. Those summers solidified my interest in modeling protein-ligand interactions and understanding the physics and forces involved in doing so. This pushed me to join UCSF with its strong foundation in computational chemistry, and the Pharmaceutical Sciences & Pharmacogenomics program, with its rigorous pharmacology and pharmacokinetics emphases so I could understand drug discovery and development.

After joining the Shoichet lab in 2015, I was already well underway in my project on receptor desolvation. But what is receptor desolvation? Why is it important in protein-ligand binding? Shape and electrostatic complementarity between the ligand and protein is not sufficient to predict binding as various players are involved in the binding

event including cofactors, ions, and hundreds to thousands of water molecules<sup>2</sup>. Thus, complex formation becomes a competition between favorable interactions between the ligand and bulk solvent, the protein binding site and solvent, and between the ligand and the protein binding site -- values large in magnitude whose difference is small and prone to error using our current modeling methods<sup>3,4</sup>.

The loss of ligand-solvent interactions when the ligand binds a protein has already been incorporated into the physics-based DOCK3.7 scoring function<sup>5</sup>, but the loss of protein-solvent interactions has not. This is because the energetics of binding site-bound water are difficult to calculate, though multiple programs have been employed to do so including 3D-RISM<sup>6</sup>, SPAM<sup>7</sup>, JAWS<sup>8</sup>, WaterMap<sup>9</sup>, and STOW<sup>10</sup>. It is also unclear which waters to displace, retain, or ignore when identifying potential ligands. A variety of algorithms are available that classify hydration sites as conserved or nonconserved, i.e. displaceable, and these include Consolv<sup>11</sup>, GRID<sup>12</sup>, PyWater<sup>13</sup>, and WaterFLAP<sup>14</sup>. These conserved waters are typically tightly bound, making multiple favorable interactions with the protein and are potentially necessary for protein structure or function. Displaceable waters, on the other hand, are high energy and mobile and thus, thought to be easy to displace<sup>2</sup>.

Additionally, the presence of bridging waters between protein and ligand is difficult to anticipate, though various docking programs have attempted to model them including FlexX 10<sup>15</sup>. Adding to the complexity is the fact that water that forms interactions with biological interfaces or that is confined in microenvironments such as binding sites exhibits different translational and rotational diffusion rates, residence times, hydrogen bond energies, polarity, pH, density, and viscosity compared to bulk

solvent<sup>16-19</sup>. Besides the water that mediates protein-ligand interactions, the energetics of water reorganizing around a ligand after binding is also difficult to evaluate, depends on the specific ligand binding, and can significantly affect the thermodynamics<sup>20</sup>. The quantity and different foci of these water-modeling algorithms highlights both the difficulty in modeling water's many behaviors accurately, but also how much research is still needed.

There are several examples from the literature regarding the importance of water and its multiple roles in protein-ligand binding. The classic example is that of cyclic urea inhibitors of HIV protease that were specifically designed to displace a conserved water molecule, while also maintaining the hydrogen bond that the water contributed<sup>21</sup>. This served to boost the potency more strongly than previous inhibitors, presumably due to an entropy increase from the release of this water into bulk solvent. However, this displace-and-replace approach isn't always successful, in some cases exhibiting no change in affinity<sup>22</sup>, or in others, decreasing affinity<sup>23,24</sup>.

Another example of water's difficult-to-model behavior is its involvement with the periplasmic oligopeptide binding protein OppA, which is capable of binding thousands of two to five amino acid peptides. Its ability to bind these diverse peptides is not due to protein conformational changes nor different protein-ligand contacts, but rather the inclusion of different numbers of water molecules that coordinate interactions between the peptide and protein<sup>25</sup>.

In addition to displace-and-replace and coordination, water can also affect the affinity of molecules by being destabilized. In the adenosine A<sub>2A</sub> G protein-coupled receptor, researchers found a correlation between the residence time of a series of

structurally related antagonists and the number and position of high energy trapped solvent molecules within the first shell of the ligand<sup>26</sup>. Similarly, a selective PI3K $\beta$  inhibitor was shown to be more selective over PI3K $\delta$  because of a less destabilized ligand-associated water that interacts with the charged Asp856 in PI3K $\beta$ , compared with the neutral Asn836 in PI3K $\delta$ <sup>27</sup> at the same position. In still other cases including R67 dihydrofolate reductase, it has been shown that it is water's reorganization after ligand binding that dominates the enthalpy of binding<sup>28</sup>.

This multifaceted process of protein desolvation, establishment of new bridging water-interactions between the protein and ligand, and reorganization of water and its associated energetics flies in the face of molecular docking methods, which are typically utilized for rapidly calculating binding energies between a rigid protein and a large library of compounds to narrow down the list to a set of plausible compounds with which to move forward. Molecular docking cannot usually calculate affinities accurately or reliably rank order high-scoring molecules due to tradeoffs between accuracy and speed<sup>29</sup>. This is likely why successful incorporation of all of water's roles into computational methods, especially molecular docking methods, hasn't been fully realized. Water models currently used focus on computational efficiency at the expense of accuracy, only partially modeling water's behavior, and thus fail at reproducing key properties of liquid water<sup>30</sup>.

Thus, the question becomes: how can we incorporate receptor desolvation in an approximate, quick way that will meaningfully account for water's role in protein-ligand binding? In addition, how can we fit this new scoring function term into the current DOCK3.7 scoring function, which is itself composed of a combination of different force

fields (Merck molecular force field 94, united atom AMBER force field to name a few), charge models (MMFF94 partial charges, AM1BCC, and united atom AMBER partial charges), and theories (Poisson-Boltzmann and Generalized Born Surface Area continuum electrostatics), and whose energies are approximate and entangled? If we can successfully incorporate this receptor desolvation term, how do we know that it is right for the right reasons? Do we know that the balance of energies in the DOCK3.7 scoring function as it is now is the best it can be, or can it be optimized further?

My first approach to incorporate receptor desolvation into DOCK3.7 involved taking advantage of the numerical Poisson-Boltzmann equation solver, QNIFFT, already utilized for calculating electrostatic energies. The goal was to pre-compute the receptor desolvation energy by placing a low-dielectric probe atom at each position in the binding site prior to docking and calculating the electrostatic energy of the system. By taking the difference in total electrostatic energy of the low-dielectric protein in high dielectric solvent and the low-dielectric protein and probe in high dielectric solvent, one could get the work associated with the change in charge-solvent interaction energies<sup>31</sup> – the electrostatic component of the solvation free energy. The electrostatic component of the solvation free energy is the interaction of a charged atom and the polarization it induces in the solvent. Thus, by placing a low-dielectric probe at each position in the binding site, we see how the lack of solvent, and thus, modified atomic-solvent interactions and solvent screening, at that position affects the total electrostatic energy of the system.

These electrostatic receptor desolvation energies were stored on a grid and read in during docking. This involved writing a new receptor desolvation scoring scheme in the DOCK source code that used trilinear interpolation, which is what the three other

scoring function terms use. What was assuring was that these electrostatic receptor desolvation energies, computed with the numerical Poisson-Boltzmann equation, for ligands were about the same magnitude as the ligand desolvation energies computed with AMSOL, which is based on Generalized Born theory. However, because pre-computing involved placing individual probes with radius 1.9 Å at each position in the binding site with a 1 Å grid spacing, this meant double-counting a substantial amount of the receptor desolvation energy when the individual atom's energies were summed up. In general, a ligand pose's receptor desolvation energies during docking reached 3-fold higher energies than the energies generated from the full ligands themselves, outside of docking. I tried a variety of different scaling factors to down-weight the desolvation energies, but in all cases, performance diminished relative to the standard scoring function.

The main problem was that introducing a probe, and thus, removing favorable protein-solvent interactions and solvent screening always increased the electrostatic energy of the system, and was therefore, always a penalty. Though this makes sense in terms of the theory, we know that water can be both favorable and unfavorable in protein cavities, and this seemed to be too big of an approximation. In terms of docking, this resulted in DOCK prioritizing molecules that were hanging out of the binding pocket as there would be fewer penalties there. What this suggested to me was that the ligand desolvation term was too strong, potentially compensating for the lack of accounting for receptor desolvation in the scoring function, and this is what motivated the parameter scanning in Chapter 3. Because of these issues, the PB-derived receptor desolvation

scoring scheme was abandoned, and a new implementation of the explicit solvent-based Grid Inhomogeneous Solvation Theory (GIST)<sup>32-34</sup> was pursued.

While I was working on the PB-derived receptor desolvation term, I was also running molecular dynamics simulations and GIST calculations on DUD-E systems to determine how GIST affected enrichment performance. My mentors, Trent Balias and Marcus Fischer, and I showed that incorporating GIST's receptor desolvation enthalpies into DOCK3.7 could be successful in a simplified model cavity with a single charged aspartate, in terms of prioritizing molecules that bind, predicting the correct binding pose, and predicting water-mediated interactions between ligand and protein<sup>35</sup>. This is highlighted in Chapter 1 of this thesis.

However, there were several drawbacks. The implementation of GIST required finding all GIST voxels (three-dimensional pixels) or grid points contained within the van der Waals radii of each atom in each pose of each ligand sampled during docking. With a 0.5 Å grid spacing for the GIST grids, this amounted to hundreds of voxels to identify on the fly, which slowed down the standard scoring function by 6-fold. As the lab was moving towards screening hundreds of millions of small molecules, this slowdown would guarantee that GIST could only be used for smaller screens. Second, the GIST term only accounted for 8% of the total docking energy, suggesting again that the ligand desolvation energy was likely too large, and potentially entangled with receptor desolvation. Lastly, a Simplex minimizer was incorporated into DOCK3.7<sup>36,37</sup>, that customarily went through up to 500 minimization steps before convergence, which, coupled with the slowdown caused by GIST, could not be incorporated into the current implementation.

Trent had been working on different implementations of GIST for scoring including a way to pre-compute GIST desolvation enthalpies by applying a Gaussian-weighting and summing up desolvation enthalpies at voxels contained within a pseudoatom that he called “blurry GIST”. We found that combining this blurry GIST scheme with the new trilinear interpolation receptor desolvation scoring I implemented into DOCK captured similar enrichment trends as the original displacement implementation of GIST, exhibited negligible slowdown in docking, and could be readily incorporated into Simplex minimization. This implementation of blurry GIST, its retrospective testing, and prospective testing after a 270 million screen to AmpC  $\beta$ -lactamase is described in Chapter 2.

## **Guide to the Chapters**

I have adapted one second author paper, one published co-first author paper, and two other soon-to-be published first author papers in the following chapters. Before each chapter, I introduce some context around the project with a short gloss.

In the first chapter, I discuss my contributions to Trent Balius’ and Marcus Fischer’s project on incorporating grid inhomogeneous solvation theory (GIST) into the DOCK3.7 scoring function. This work was performed on the cytochrome *c* peroxidase gateless mutant (CcP-ga) and was the first data that showed that receptor desolvation in DOCK could meaningfully improve prioritization of binding compounds and pose geometry in prospective screens. Next, I describe my efforts to incorporate the faster, Gaussian-weighted blurry GIST into DOCK3.7, and its application to identifying new ligands for AmpC  $\beta$ -lactamase. We ran retrospective enrichments on 40 DUD-E



systems with and without blurry GIST, ran a 300 million molecule prospective screen on AmpC, bought molecules prioritized and de-prioritized by GIST, and characterized them experimentally with kinetic binding assays and X-ray crystallography. In chapter 3, I describe parameter scanning of the DOCK scoring function on 41 DUD-E systems and DRD4 and MT1, which was an attempt to identify a better balance of the scoring function terms. This exercise came directly out of our work done on receptor desolvation as I hoped to find a better balance of the terms that would allow us to incorporate GIST more readily. In chapter 4, I move away from optimizing and extending the scoring function and describe our work on the melatonin receptors to identify type-selective molecules for the MT<sub>1</sub> receptor in collaboration with Bryan Roth's lab at University of North Carolina at Chapel Hill, Margarita Dubocovich's lab at the State University of New York at Buffalo, and Vadim Cherezov's lab at the University of Southern California. In that work, we were able to identify picomolar and nanomolar agonists and inverse agonists of the melatonin receptors and optimize these into MT<sub>1</sub>-selective inverse agonists that exhibited agonist and inverse agonist phenotypes *in vivo*. I then finish off with a description of what I have been working on in the last years of my PhD – docking to the CB<sub>1</sub> cannabinoid receptor for agonists that may be involved in analgesia – as well as a chapter focused on future directions and projects.

## References

1. Bottegoni, G., Kufareva, I., Totrov, M. & Abagyan, R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J Med Chem* **52**, 397-406, doi:10.1021/jm8009958 (2009).
2. Spyrikis, F. *et al.* The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J Med Chem* **60**, 6781-6827, doi:10.1021/acs.jmedchem.7b00057 (2017).
3. Shoichet, B. K., McGovern, S. L., Wei, B. & Irwin, J. J. Lead discovery using molecular docking. *Curr Opin Chem Biol* **6**, 439-446, doi:10.1016/s1367-5931(02)00339-3 (2002).
4. Brenk, R., Vetter, S. W., Boyce, S. E., Goodin, D. B. & Shoichet, B. K. Probing molecular docking in a charged model binding site. *J Mol Biol* **357**, 1449-1470, doi:10.1016/j.jmb.2006.01.034 (2006).
5. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
6. Beglov, D. & Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *The Journal of Physical Chemistry B* **101**, 7821-7826, doi:10.1021/jp971083h (1997).
7. Cui, G., Swails, J. M. & Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J Chem Theory Comput* **9**, 5539-5549, doi:10.1021/ct400711g (2013).

8. Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J Am Chem Soc* **131**, 15403-15411, doi:10.1021/ja906058w (2009).
9. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* **130**, 2817-2831, doi:10.1021/ja0771033 (2008).
10. Li, Z. & Lazaridis, T. Computing the thermodynamic contributions of interfacial water. *Methods Mol Biol* **819**, 393-404, doi:10.1007/978-1-61779-465-0\_24 (2012).
11. Raymer, M. L. *et al.* Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J Mol Biol* **265**, 445-464, doi:10.1006/jmbi.1996.0746 (1997).
12. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28**, 849-857, doi:10.1021/jm00145a002 (1985).
13. Patel, H., Gruning, B. A., Gunther, S. & Merfort, I. PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. *Bioinformatics* **30**, 2978-2980, doi:10.1093/bioinformatics/btu424 (2014).
14. Baroni, M., Cruciani, G., Sciabola, S., Perruccio, F. & Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J Chem Inf Model* **47**, 279-294, doi:10.1021/ci600253e (2007).

15. Rarey, M., Kramer, B. & Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34**, 17-28 (1999).
16. Nandi, N., Bhattacharyya, K. & Bagchi, B. Dielectric relaxation and solvation dynamics of water in complex chemical and biological systems. *Chem Rev* **100**, 2013-2046, doi:10.1021/cr980127v (2000).
17. Nandi, N. & Bagchi, B. Dielectric Relaxation of Biological Water. *The Journal of Physical Chemistry B* **101**, 10954-10961, doi:10.1021/jp971879g (1997).
18. Garcia-Sosa, A. T. Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J Chem Inf Model* **53**, 1388-1405, doi:10.1021/ci3005786 (2013).
19. Levy, Y. & Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct* **35**, 389-415, doi:10.1146/annurev.biophys.35.040405.102134 (2006).
20. Geschwindner, S. & Ulander, J. The current impact of water thermodynamics for small-molecule drug discovery. *Expert Opin Drug Discov* **14**, 1221-1225, doi:10.1080/17460441.2019.1664468 (2019).
21. Lam, P. Y. *et al.* Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380-384, doi:10.1126/science.8278812 (1994).
22. Seo, J. *et al.* Structure-based design and synthesis of N(omega)-nitro-L-arginine-containing peptidomimetics as selective inhibitors of neuronal nitric oxide

- synthase. Displacement of the heme structural water. *J Med Chem* **50**, 2089-2099, doi:10.1021/jm061305c (2007).
23. Andaloussi, M. *et al.* Design, synthesis, and X-ray crystallographic studies of alpha-aryl substituted fosmidomycin analogues as inhibitors of Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *J Med Chem* **54**, 4964-4976, doi:10.1021/jm2000085 (2011).
24. Mikol, V., Papageorgiou, C. & Borer, X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem* **38**, 3361-3367, doi:10.1021/jm00017a020 (1995).
25. Tame, J. R., Sleight, S. H., Wilkinson, A. J. & Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat Struct Biol* **3**, 998-1001, doi:10.1038/nsb1296-998 (1996).
26. Bortolato, A., Tehan, B. G., Bodnarchuk, M. S., Essex, J. W. & Mason, J. S. Water network perturbation in ligand binding: adenosine A(2A) antagonists as a case study. *J Chem Inf Model* **53**, 1700-1713, doi:10.1021/ci4001458 (2013).
27. Robinson, D. *et al.* Differential Water Thermodynamics Determine PI3K-Beta/Delta Selectivity for Solvent-Exposed Ligand Modifications. *J Chem Inf Model* **56**, 886-894, doi:10.1021/acs.jcim.5b00641 (2016).
28. Timson, M. J. *et al.* Further studies on the role of water in R67 dihydrofolate reductase. *Biochemistry* **52**, 2118-2127, doi:10.1021/bi301544k (2013).

29. Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* **59**, 4103-4120, doi:10.1021/acs.jmedchem.5b02008 (2016).
30. Onufriev, A. V. & Izadi, S. Water models for biomolecular simulations. **8**, e1347, doi:10.1002/wcms.1347 (2018).
31. Gilson, M. K. & Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* **4**, 7-18, doi:10.1002/prot.340040104 (1988).
32. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* **137**, 044101, doi:10.1063/1.4733951 (2012).
33. Nguyen, C. N., Cruz, A., Gilson, M. K. & Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J Chem Theory Comput* **10**, 2769-2780, doi:10.1021/ct401110x (2014).
34. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B* **102**, 3531-3541, doi:10.1021/jp9723574 (1998).
35. Balias, T. E. *et al.* Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc Natl Acad Sci U S A* **114**, E6839-E6846, doi:10.1073/pnas.1703287114 (2017).
36. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229, doi:10.1038/s41586-019-0917-9 (2019).

37. Gschwend, D. A. & Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* **10**, 123-132, doi:10.1007/bf00402820 (1996).

## List of my related publications

Balius, T.E.\*; Fischer, M.\*; Stein, R.M.; Adler, T.B.; Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T.; Shoichet, B.K., Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery. *Proc. Natl. Acad. Sci. U.S.A.*, **2017**, 114 (33) E6839-E6846. Doi: 10.1073/pnas.1703287114.

\*these authors contributed equally

Stein, R.M. \*; Kang, HJ\*; McCorvy, JD\*; Glatfelter, GC\*; Jones, AJ; Che, T; Slocum S; Huang, XP; Savych, O; Moroz, YS; Stauch, B; Johansson, LC; Cherezov, V; Irwin JJ; Shoichet BK; Roth, BL; Dubocovich, ML. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609-614, doi:10.1038/s41586-020-2027-0 (2020).

\*these authors contributed equally



## Gloss to Chapter 1

This chapter marks my introduction to receptor desolvation and the Shoichet lab. I remember sitting in Brian's office with Trent and Marcus in my first year to discuss a potential rotation and though I wanted to focus on experimental work at that time, I was quite intrigued by the idea of implementing and testing a new term in the scoring function. After hearing Brian's pitch of the project, I asked a question regarding the different roles of the scoring function terms, and though the conversation is fuzzy, I remember him answering that though the terms were modeled separately, they were intertwined in reality, competing, blending, and participating together. This was a fascinating insight to me and sparked a lot of the questions I had and have about the approximations we use in molecular modeling and how they all fit together to create a flawed, yet partially correct version of reality.

My contributions to this project involved choosing the 25 DUD-E systems to be used for retrospective enrichment calculations. This included scouring PDB structures and finding proteins with no missing loops, no cofactors, and that contained water molecules in their binding sites, running 50 nanosecond molecular dynamics simulations, then GIST calculations, and then docking to these proteins with different GIST weightings. Through this, I learned a lot about parameterizing systems for molecular dynamics and docking, and since there wasn't an easy way to prepare a protein for AMBER, or convert that protein from AMBER to DOCK format, a lot of this preparation was done manually. I spent many nights, several of these over winter break 2015, going through PDB files, adding charge-capping groups, disulfide bonds, and

ions, checking them in Chimera, as well as making sure the alignment between protein structures from the AMBER MD simulation were the same as those I prepared with our automated docking preparation pipeline, Blastermaster. I have pages and pages of explanations and commands in my first lab notebook on preparing these 25 DUD-E systems. After docking to the 25 DUD-E systems, we found that a -0.5 GIST weighting performed the best retrospectively with a mild enrichment improvement of +0.53, though we chose to use a weighting of -1.0 for the prospective screen as it had a larger contribution to the total DOCK score.

On the experimental side, I was responsible for dissolving compounds and running the binding assays on CcP-ga, as well as setting up some crystal trays to identify optimal conditions for growth, though this was mainly for practicing crystallography for my thesis project. With Marcus' help, I was also able to refine one of the crystal structures of CcP-ga in complex with a new ligand (PDB: 5UG2).

I gained both computational and experimental experience, but it was the computational side that really sparked my curiosity. Thus, the code became my focus during my PhD. Overall, we found that of the 14 molecules prioritized by GIST (Pro-GIST), 13 of these bound, while none of the 3 molecules deprioritized by GIST (Anti-GIST) bound. In terms of geometry, GIST predicted 6 of 9 crystallographic poses correctly, while the standard scoring function succeeded in 5 of 9 structures. Most exciting was the fact that a GIST-predicted water mediated the interaction between the ligand and protein in one of these correctly predicted crystal structures. Though GIST's contribution to the total docking score was small, it seemed to have a meaningful effect, which motivated the work featured in Chapter 2.

## Chapter 1: Testing IST in Structure-Based Ligand Discovery

Trent E. Balius,<sup>(a)#</sup> Marcus Fischer,<sup>(a)#‡</sup> Reed M. Stein,<sup>(b)</sup> Thomas B. Adler,<sup>(a)</sup> Crystal N. Nguyen<sup>(c)</sup>, Anthony Cruz,<sup>(d,e)</sup> Michael K. Gilson,<sup>(c)</sup> Tom Kurtzman,<sup>(d,e,f)</sup> and Brian K. Shoichet<sup>\*(a)</sup>

a. University of California, San Francisco, Department of Pharmaceutical Chemistry, San Francisco, California, 94158, United States of America

b. University of California, San Francisco, Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, San Francisco, California, 94158, United States of America

c. University of California, San Diego, Skaggs School of Pharmacy and Pharmaceutical Sciences, La Jolla, CA, 92093, United States of America

d. Lehman College Department of Chemistry, 250 Bedford Park Blvd West Bronx New York, 10468, United States of America

e. Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York New York, 10016, United States of America

f. Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York New York, 10016, United States of America

‡ Present address: Departments of Chemical Biology and Therapeutics, Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, United States.

# contributed equally

\* bshoichet@gmail.com -- to whom correspondence should be addressed.

The text of this chapter is adapted from:

Balius, T.E.\*; Fischer, M.\*; Stein, R.M.; Adler, T.B.; Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T.; Shoichet, B.K., Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery. *Proc. Natl. Acad. Sci. U.S.A.*, **2017**, 114 (33) E6839-E6846. Doi: 10.1073/pnas.1703287114.

\*these authors contributed equally

## 1.1 Abstract

Binding site water is often displaced upon ligand recognition, but is commonly neglected in structure-based ligand discovery. Inhomogeneous Solvation Theory (IST) has become popular to treat this effect, but it has not been tested in controlled experiments at atomic resolution. To do so, we turned to a Grid-based version of this method, GIST, readily implemented in molecular docking. Whereas the new term only improves docking modestly in retrospective ligand enrichment, it could be added without disrupting performance. We thus turned to prospective docking of large libraries to investigate GIST's impact on new ligand discovery, geometry, and water structure in a model cavity site well-suited to exploring these terms. Although top-ranked docked molecules with and without the GIST term often overlapped, many ligands were meaningfully prioritized or deprioritized; some these were selected for testing. Experimentally, 13/14 new molecules prioritized by GIST did bind while none of the molecules that it deprioritized were observed to bind. Nine crystal complexes were determined: in six the ligand geometry corresponded to that predicted by GIST, for one of these the pose without the GIST term was wrong, three crystallographic poses differed from both predictions. Notably, in one structure an ordered water molecule with a high GIST displacement penalty by GIST was observed to stay in place. Inclusion of this water-displacement term can substantially improve the hit rates and ligand geometries from docking screens, though the magnitude of its effects can be small, and its impact in drug binding sites merits further controlled studies.

## **1.2 Significance Statement.**

Water molecules play a crucial role in protein-ligand binding. Calculating the energetic consequences of displacing water upon ligand binding has challenged the field for many years. Inhomogeneous Solvation Theory (IST) is one of the most popular methods to distinguish favorable from unfavorable water molecules, but little controlled, prospective testing, at atomic resolution, has been done to evaluate the method. Here, we compare molecular docking screens with and without an IST term to gauge its impact on ligand discovery. We test predictions that include an IST term in prospective experiments for new ligands, using crystallography and direct binding.

### 1.3 Introduction

The treatment of receptor-bound water molecules, which are crucial for ligand recognition, is a widely recognized challenge in structure-based discovery.<sup>1-4</sup> The more tightly bound a water in a site, the greater the penalty for its displacement upon ligand binding, ultimately leading to its retention and the adoption of ligand geometries that do not displace it. More problematic still are when a new bridging water mediates interactions between the ligand and the receptor. Because the energetics of bound water molecules have been challenging to calculate, and bridging waters hard to anticipate, large-scale docking of chemical libraries have typically been conducted against artificially desolvated sites, or have kept a handful of ordered water molecules that are treated as part of the site, based on structural precedence.<sup>5-8</sup>

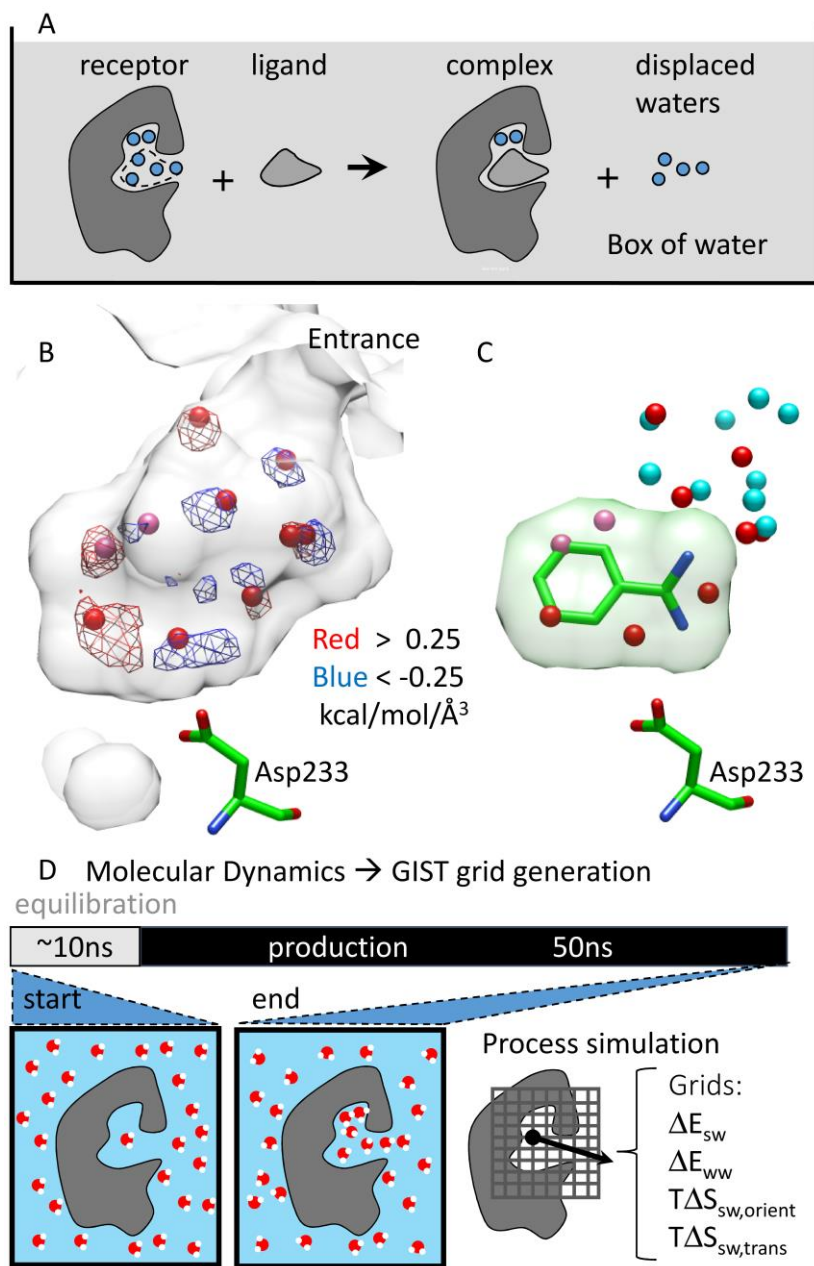
Recently, several relatively fast approaches, pragmatic for early discovery, have been advanced to account for the differential displacement energies of bound water molecules,<sup>9-20</sup> complementing more rigorous but computationally expensive approaches<sup>18-22</sup>. Among the most popular of these has been Inhomogeneous Solvation Theory (IST).<sup>23-25</sup> IST uses populations from molecular dynamics simulations on protein (solute) surfaces to calculate the cost of displacing individual water molecules (solvent) on that surface. IST has been used to calculate ligand SAR,<sup>26-29</sup> to map protein binding sites for solvent energetics,<sup>28,30,31</sup> to quantify the energetic contribution of structural waters,<sup>25,32</sup> and to understand water networks and how they rearrange in the presence of ligands.<sup>33</sup> There have been several implementations of IST including WaterMap<sup>26,27,31</sup> and STOW

<sup>32</sup>, and the approach has been integrated into library docking programs such as Glide<sup>34,35</sup>, DOCK3.5.54,<sup>36</sup> and Autodock.<sup>37</sup>

Notwithstanding its popularity, IST has rarely been tested in prospective library screens for its ability to predict new ligands, their bound geometries, and the water molecules that they either do or do not displace.<sup>4</sup> Here, we do so in a model cavity in Cytochrome c Peroxidase (CcP-ga), a highly-defined buried site, but partially open to bulk solvent, that binds small heterocyclic monocations. We and others have used this and related cavities as model systems for docking, owing to their small size, the dominance of one or two interaction terms in ligand binding, and the existence of thousands of plausible ligands among commercially available, dockable small molecules.<sup>38-41</sup>

The CcP-ga cavity is particularly well suited to explore the impact of ordered waters on the prospective discovery of novel ligands (**Figure 1.1**). On binding, ligands displace between three and eight waters observed in apo-structures,<sup>38,39</sup> while new waters can be recruited to bridge between the cavity and the ligands. The limited number of these waters and the tight definition of the site makes exploration of the problem tractable. Also, the affinities of newly predicted ligands may be determined quantitatively and their structures may be determined to high resolution, making atomic resolution testing plausible.





### Figure 1.1. Receptor desolvation using GIST.

(A) Upon ligand binding, ordered water can be displaced, remain unaffected, or bridge between ligand and protein. (B) The CcP-gateless apo cavity (transparent surface) is filled with 9 crystallographic water molecules (red spheres, pink spheres indicate half occupancy) (4NVA) and compared to GIST enthalpy grid maps representing unfavorable water positions (red mesh,  $>0.25$  kcal/mol/Å<sup>3</sup>) and favorable water positions (blue mesh,  $<-0.25$  kcal/mol/Å<sup>3</sup>). (C) Ligand benzamidine (4NVC) displaces four apo cavity waters (red spheres) and reorders several of the remaining waters (cyan

spheres) about the ligand. **(D)** The GIST grids are calculated by post-processing a molecular dynamics (MD) simulation of a restrained apo protein in a box of water.

We integrated GIST, the grid implementation of IST,<sup>30</sup> into DOCK3.7. In GIST, MD simulations of the hydrated receptor are analyzed to yield spatially resolved information about water density and thermodynamics over the voxels (cubic grid cells) of a three-dimensional grid covering the protein binding site (**Figure 1.1**). The grid basis of GIST lends itself to docking because water displacement energies can be pre-calculated and stored on a lattice of points, supporting the rapid scoring necessary for large library screens. These water energies can then be combined with the other terms of the DOCK3.7 physics-based scoring function.

We first tested including GIST in retrospective controls against 26 targets drawn from the DUD-E benchmark<sup>42</sup>, composed of about 6600 annotated ligands and 400,000 property matched decoys<sup>42</sup>. These enrichment calculations investigate the weighting of the new GIST term ( $E_{rec,desol}$ ) with other DOCK3.7 terms<sup>43</sup>: van der Waals ( $E_{vdW}$ ), electrostatic ( $E_{es}$ ), ligand desolvation ( $E_{lig,desol}$ ), and protein conformational energies ( $E_{rec,conf}$ ) (eq 1.1).

$$E_{score} = E_{rec,desol} + E_{vdw} + E_{es} + E_{lig,desol} + E_{rec,conf} \quad (\text{Equation 1.1})$$

These retrospective calculations helped calibrate the new term, assess its computational cost, and establish that it could be used without disrupting the balance of the other scoring terms.

More illuminating are prospective tests that we prosecuted against the model cavity. In screens of between 0.2 to 1.8 million compounds, we prioritize molecules by

three criteria: 1) they are previously untested, 2) they rank substantially better or worse with the GIST term than without it, or 3) they bind differently due to the displacement of GIST-defined water molecules. A total of 17 new molecules were purchased and tested experimentally for binding, and nine ligand-CcP-ga crystal structures were determined. From these studies, several advantages of IST for ligand discovery emerge; the method meaningfully improved the selection of new ligands, and was often right for the right reasons, correctly capturing the role of displaceable or implicitly bridging water. Still, and notwithstanding the great advantages of IST seen in other studies,<sup>26-29,34</sup> in controlled prospective discovery, at atomic resolution its liabilities also emerge.

## 1.4 Results

Inhomogeneous Solvation Theory methods use a molecular mechanics potential energy function and water occupancies to calculate thermodynamic properties of water in the context of the receptor. In GIST, the energies of solute-water enthalpy ( $E_{s,w}$ ), water-water enthalpy ( $E_{w,w}$ ), translational ( $TS_{trans}$ ), and orientational ( $TS_{orient}$ ) entropy are represented spatially into grid voxels. The receptor desolvation cost is calculated by summing the voxels displaced by a docked ligand and added to the DOCK3.7 scoring function (cf. eq 1.1). To investigate how the new GIST energies are best weighted, and which GIST terms are most useful— as there are questions on this point in the literature<sup>28,37</sup> — we began with retrospective calculations against the CcP-ga cavity, docking 46 known ligands against 3,338 property matched decoys. We explored four different combinations of the GIST grids: (1) unscaled Free Energy ( $E_{GIST} = E_{s,w} + E_{w,w} + TS_{trans} + TS_{orient}$ ), (2) unscaled Enthalpy ( $E_{GIST} = E_{s,w} + E_{w,w}$ ), (3) scaled Free Energy

( $E_{\text{GIST}} = E_{\text{s,w}} + 2 \times E_{\text{w,w}} + TS_{\text{trans}} + TS_{\text{orient}}$ ), and (4) scaled Enthalpy ( $E_{\text{GIST}} = E_{\text{s,w}} + 2 \times E_{\text{w,w}}$ ); both with the water-water term scaled by two (**Figure A.1.3**, and **Table A.1.1**). Here, enthalpy was not normalized by occupancy, in contrast to previous studies,<sup>28,37</sup> but still referenced to bulk water energy, as this produced the best enrichments. Following convention negative GIST energies reflect favorable, costly-to-displace waters. We used Adjusted Log AUC to measure docking enrichment,<sup>43-47</sup> this metric weights each factor of ten in docking rank order equally, beginning from the top 0.1%, prioritizing the performance of the very top-ranking ligands or decoys in the docking screen.<sup>44</sup> Scaled Enthalpy performed the best (Adjusted log AUC of  $57.46 \pm 1.84$ ), closely followed by unscaled Free Energy ( $56.08 \pm 1.42$ ). Enthalpy alone performed the worst with ( $49.50 \pm 1.34$ ). Setting  $E_{\text{GIST}} = E_{\text{s,w}} + 2 \times E_{\text{w,w}}$  sets aside several GIST terms, but has precedence in earlier studies.<sup>28,30</sup>

We next explored the receptor desolvation term and the best scaling factor ( $\alpha$ , eq A.1.8) to bring the GIST value into balance with the other terms in eq 1 (**Figure A.1.4** and **Table A.1.2**). Staying with the CcP-ga system, we considered eight scaling factors ranging from -8.0 to +8.0 for the weighting of  $E_{\text{GIST}}$ . Reassuringly, we found that the scaling factors of -1.0 (log AUC =  $57.46 \pm 1.84$ ) and -0.5 (log AUC =  $56.54 \pm 2.10$ ) behave better than overweighting the term by a factor of -8.0 (log AUC =  $36.91 \pm 1.52$ ) or +8.0 (log AUC =  $46.94 \pm 2.07$ ). At a scaling of -1.0, the absolute GIST energy averaged 1.99 kcal/mol for the top-ranking 100 docked molecules, about 8% of the value of the overall docking energy score in this cavity. Here, as in all calculations in this study, we based the GIST energies on MD simulations of 50 ns. These appeared to be sufficiently

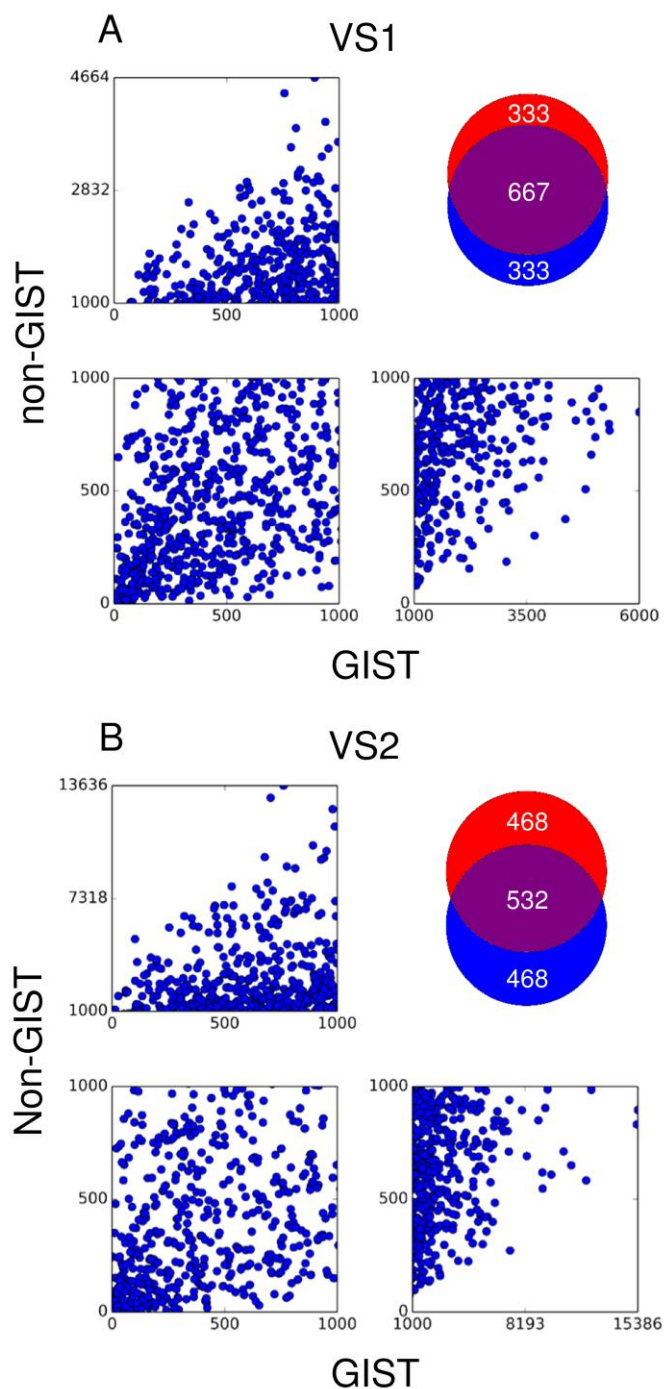
converged for docking, based on the small variance in performance using GIST grids from each of ten 5 ns sub-trajectories (**Figure A.1.5** and **Table A.1.3**).

Using the same GIST terms used in the cavity (equation 1), we examined the impact of scaling factors on 25 DUD-E systems for which solvation likely plays a role. These 25 targets bind a diverse range of cationic (CXCR4, ACES, TRY1), anionic (PUR2, AMPC, PTN1), and neutral ligands (ITAL, KITH, and HS90a), and make water-mediated interactions (AMPC, EGFR). In these systems, we noticed that there were a very few voxels in the GIST grids—on average 58 out of 210,000 total voxels—with extremely high magnitude absolute energies, ranging from 14.6 to 119.7 kcal/mol/Å<sup>3</sup>, between 101 and 391  $\sigma$  (standard deviations) away from the mean voxel energies. These extrema seem to reflect the restrained MD simulations used for the GIST calculations, as when we allowed even side chains to move in the MD, they were much attenuated or entirely eliminated. Accordingly, we truncated the maximum absolute magnitude of the GIST grids at 3 kcal/mol/Å<sup>3</sup> in these retrospective calculations (a value still on average 12  $\sigma$  away from the mean voxel energies); we also scaled the GIST energy by -0.5 when combining it with the other terms in the DOCK3.7 scoring function, which we found to perform slightly better than a simple weighting of 1.0 (**Table A.1.4** further describes the origins of the energy extrema and the retrospective docking performance under different weighting of the GIST term). In the retrospective docking screens, 13 of the 25 DUD-E systems had better enrichment versus docking without the GIST term, 6 had worse enrichment, and 6 were within +/- 0.5 Log AUC difference (unchanged). The average log AUC difference over all systems is 0.53 better than no-GIST (**Table A.1.4**, and **Figure A.1.6**). To get a sense of the impact of the GIST

energies, the absolute value of the GIST term was about 6 kcal/mol for the top 100 ranked docked molecules in the 25 DUD-E targets, about 12% of the total docking score for these molecules. For the CcP-ga cavity, to which we will turn for prospective screens, the absolute GIST energy was about 8% of the total docking score for the top 100 docked molecules. The overall impact of GIST on the DUD-E benchmarks is modest, and perhaps the most important result to emerge from these retrospective controls is that the GIST term may be added without disrupting the docking scoring function, retaining physically sensible results.

We next turned to prospective docking screens against the CcP-ga cavity, with and without an unweighted (-1.0) GIST term, looking to predict new cavity ligands and their geometries. The GIST grids identified four favorable water sites in the pocket, including one close to Asp233, and three unfavorable water sites, including two regions close to the heme, and one near Gly178, a residue that can hydrogen bond with ligands through its backbone (**Figure A.1.7**, and **Table A.1.5**). We docked two purchasable fragment libraries, one straight from ZINC of ~200,000 molecules prepared at pH 6.4 (VS1), and 1.8 million molecules built at a pH of 4.0 (VS2), which favors positively charged molecules typically recognized by the cavity Asp233. We sampled, in VS1, 462.5 million orientations of the library molecules and ~15 billion scored conformations; 95,000 of the 200,000 molecules could be fit in the site. From the larger VS2 screen 5.9 billion orientations and about 319 billion scored conformations were sampled; 1.09 million molecules could be fit in the site. To isolate the effect of the GIST term on our screening performance we ran each screen twice, with and without the GIST term.

Most of the top-ranking 1000 molecules are shared between the GIST and non-GIST screens, 667 are shared in VS1 while 532 are shared in the larger VS2 (**Figure 1.2**), reflecting the comparatively small magnitude of the GIST energies relative to the overall docking score (below). We focused on those molecules that experienced rank changes of a half-log (3.16-fold) or better. For instance, a molecule that changed rank from 30<sup>th</sup> to 100<sup>th</sup>, or from 400<sup>th</sup> to 1300<sup>th</sup> on including the GIST term would be prioritized. From the smaller screen (VS1) 217 docking hits improved ranks by at least half-a-log order with the GIST term while 282 had ranks that were better by at least this amount without the GIST term. For the larger VS2 screen, 2421 had half-log improved ranks with GIST while 2869 had ranks that improved by at least half-a-log order without it. There were also several molecules for which the inclusion of the GIST term greatly changed the docked geometry; these we also considered for testing.



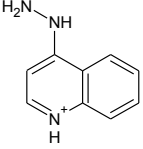
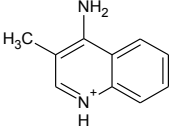
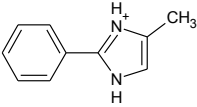
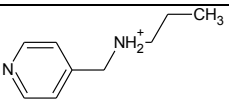
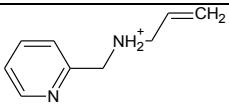
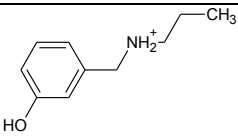
**Figure 1.2. Comparison of GIST and non-GIST screens.**

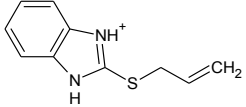
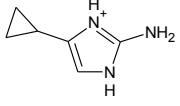
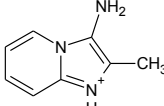
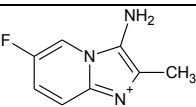
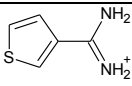
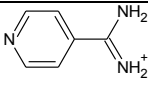
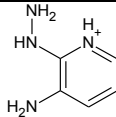
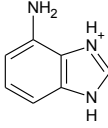
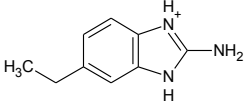
**(A)** Results from the virtual screen (VS) 1 of 200,000 molecules. **(B)** Results from VS2 of 1.8 million molecules. Top right panel shows a Venn diagram of the top 1000 ranked molecules from the GIST screen in red and non-GIST in blue. Bottom left panel is the overlapping region.

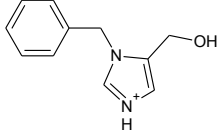
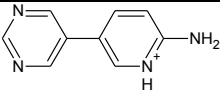


Based on these criteria, 17 molecules were acquired for experimental testing. Compounds **3** to **14** were selected because their ranks improved with GIST (Pro-GIST), while compounds **15** to **17** were selected because of better ranks without the GIST term (Anti-GIST) (**Table 1.1**). We also looked for molecules where a substantial pose change occurred between the two scoring functions (e.g. compounds **1** and **2**, **Tables 1.1** and **A.1.6**). Finally, we considered implicit water-mediated interactions to be favorable regions in the GIST grid within hydrogen-bonding distance to ligand and protein, though no explicit water molecules were used. This occurred with compounds **3**, **4**, **5**, and **6** (**Table 1.1**). In selecting these compounds, we were sometimes led to compounds that we expected, based on past experience with this cavity, to be GIST failures. For instance, compounds **3** through **6** adopted an unusual geometry in the site, giving up a direct ion-pair with Asp233 to hydrogen bond with backbone carbonyls, owing to a large implicit desolvation cost for docked orientations where the ion pair was formed. These poses were relatively favored by the GIST term, but we expected them either not to bind or to bind to form the ion pair. Conversely, we expected the molecules deprioritized by GIST to bind, in contrast with the new term, also based on precedence of other molecules. For both classes of molecules it was the GIST prediction that was confirmed, to our surprise.

**Table 1.1.** New candidate CcP ligands

Cmpd #	ZINC id	Structure	GIST Rank	Non-GIST rank	GIST energy (kcal/mol) <sup>a</sup>	K <sub>d</sub> (μM) <sup>b</sup>	RMSD to xray
<b>Compounds with different docked geometries</b>							
<b>1</b>	2564381		490	180	1.46	n.d.	G = 1.90 Å NG = 3.00 Å
<b>2</b>	6557114		664	740	2.03	154 ±19	G = 0.28 Å NG = 3.19 Å
<b>Compounds prioritized by GIST</b>							
<b>3</b>	4705523		13	249	-1.67	3472 ±172	1.34 Å
<b>4</b>	6869116		112	464	0.60	809 ±99	--
<b>5</b>	6855945		869	2550	-0.07	1606 ±287	--
<b>6</b>	19439634		91	355	0.86	3435 ±860	--

Cmpd #	ZINC id	Structure	GIST Rank	Non-GIST rank	GIST energy (kcal/mol) <i>a</i>	K <sub>d</sub> (μM) <sup>b</sup>	RMSD to xray
7	1827502		5	19	2.12	114 ±20	--
8	42684308		601	1916	0.04	1962 ±554	0.79 Å
9	20357620		98	745	-0.65	522 ±21	1.72 Å
10	74543029		1128	4923	0.46	~712 ±231	1.81 Å
11	161834		358	1212	0.28	1.30 ±0.03	0.44 Å
12	2389932		118	645	-0.02	619 ±63	0.60 Å
13	39212696		147	1462	-1.82	n.d.	--
14	112552		747	4380	0.01	29.6 ±2.5	0.46 Å
<b>Compounds prioritized by non-GIST</b>							
15	2534163		9487	906	8.56	NB	--

Cmpd #	ZINC id	Structure	GIST Rank	Non-GIST rank	GIST energy (kcal/mol) <i>a</i>	K <sub>d</sub> (μM) <sup>b</sup>	RMSD to xray
<b>16</b>	156254		1482 8	1657	8.70	NB	--
<b>17</b>	22200625		6000	577	8.09	n.d.	--

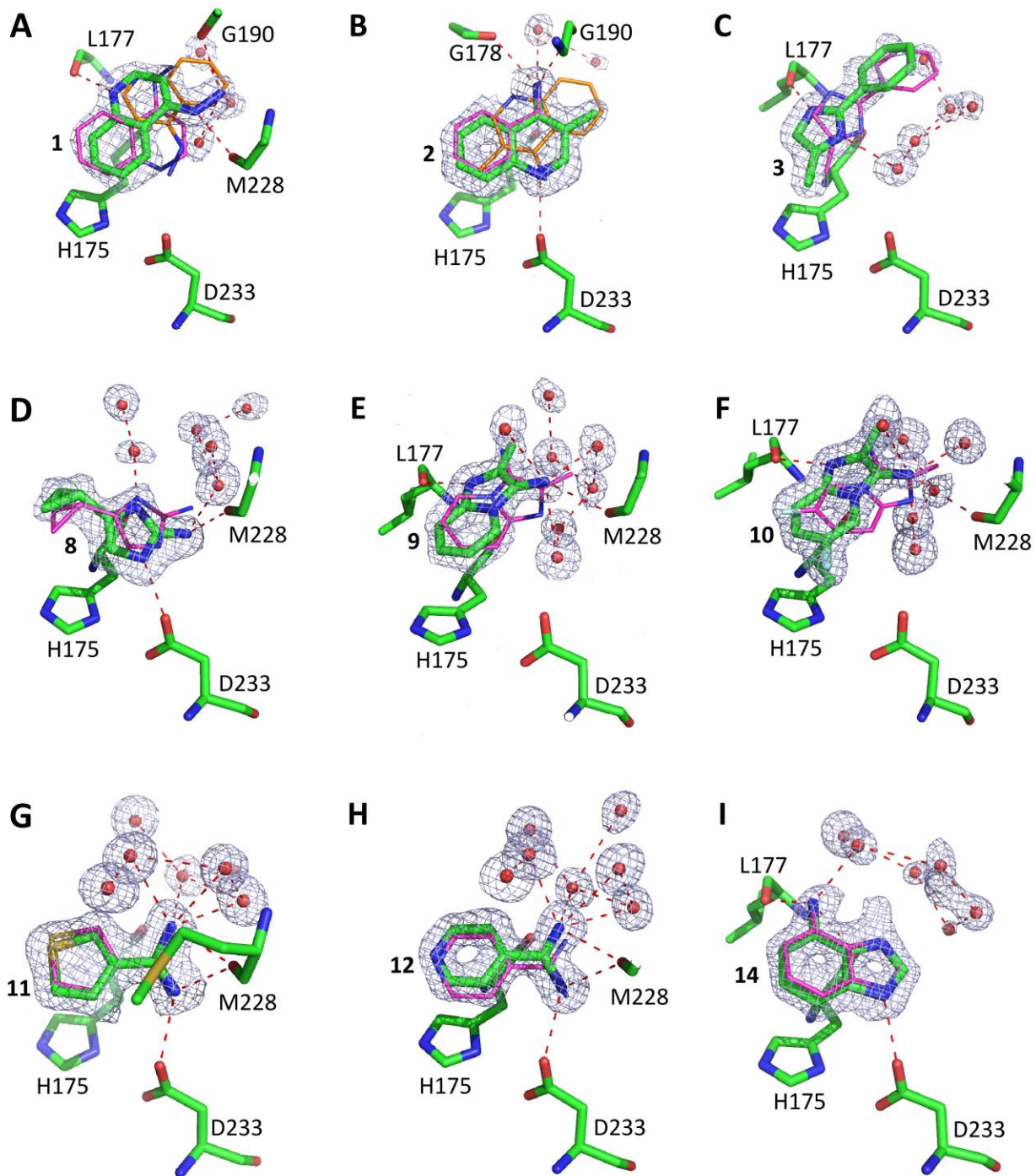
<sup>a</sup> positive GIST values are penalties.

<sup>b</sup> **n.a.**, not available - molecule not in assayable form. **n.d.**, not determinable - compound interferes with absorbance peaks. **NB**, non-binder <5mM. “~”, assay interference of compound **10** before saturation was reached.

<sup>c</sup> RMSDs are calculated with the Hungarian algorithm (lower bound): GIST pose G, non-GIST pose NG, “--” no crystal structure available, single values for same G vs NG pose.

**Pro-GIST.** We tested the binding of 14 GIST-favored molecules, determining X-ray crystal structures for nine of them. All crystallographic datasets were collected to at least 1.6 Å resolution and refined to R<sub>free</sub> values under 20%, indicating good global model quality. Locally, electron density maps for the ligands in the cavity were unambiguous as early as unrefined initial Fo-Fc maps. Final 2mFo-DFc composite omit maps<sup>48</sup> show unbiased electron density for the binding site ligand and water molecules (**Figure 1.3**). This allowed ready placement of the ligands and ordered water molecules in the final stages of refinement. Automatic refinement of ligand and water occupancies showed that ligands are unequivocally present in the binding site (between 88-93% occupancy); the complex with compound **14** refined to 73% occupancy in the presence of 26% MES from the crystallization buffer (**Figure A.1.8** and **Table A.1.7**). We

modeled all ligands in a single conformation, with only compound **2** showing difference density for an alternative ligand conformation. Electron densities of binding site waters are generally well defined (**Figure 1.3**), indicating extensive water networks that interact with both ligand and protein.

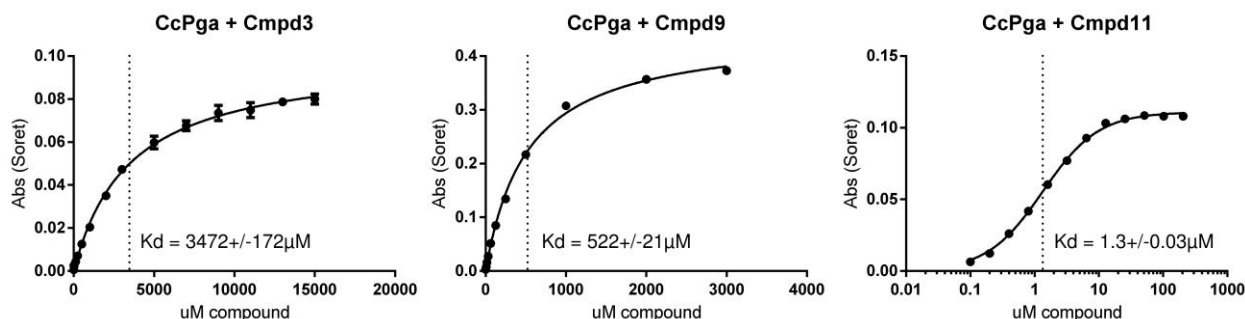


**Figure 1.3. Comparison of experimental and predicted binding poses.**

Superposition of crystallographic (green) and predicted ligand poses (GIST docking poses in purple; differential non-GIST docking poses for compounds **1** and **2** in orange). 2mFo-DFc omit electron density maps (blue mesh) are shown at 1 $\sigma$  for binding site ligand and water molecules (red spheres), with hydrogen bonds shown as red dashed lines. Nine compounds are shown (with PDB-IDs): (A) compound **1**, 5u60; (B)

compound **2**, 5u5w; (C) compound **3**, 5u5z; (D) compound **8**, 5u61; (E) compound **9**, 5u5y; (F) compound **10**, 5ug2; (G) compound **11**, 5u5x; (H) compound **12**, 5u5u; and (I) compound **14**, 5u5v. For clarity, co-crystallized MES for compound **14** is omitted (cf. Figure A.1.7).

Of the 14 docked molecules favored by the GIST term, 13 (93%) could be shown to bind, typically by a UV-Vis Soret band perturbation assay (**Figure 1.4** and **Figure A.1.9**).<sup>49</sup> Affinities for 11 ligands were determined at least in duplicate and fit to a one-site binding model with  $R^2$  values of at least 95%. Two molecules were only observed bound in their co-complexed crystal structures, owing to assay interference (**Table 1.1**). The  $K_d$  values of the GIST-prioritized molecules ranged from 1.3  $\mu\text{M}$  to 3.5 mM, with eight better than 1 mM. For these fragments the ligand efficiencies (LEs) ranged from 1.0 to 0.28 kcal/mol/atom.



**Figure 1.4. Three representative ligand binding curves.**

The Soret band shift is shown as a function of ligand concentration ( $\mu\text{M}$ ). The plots for compounds **3** and **9**) are on a linear scale while, for clarity, the x-axis of the plot for compound **11** is on the log-scale. The dashed line indicates the  $K_d$ . The circles and bars are the mean and estimated error of two observations.

Compound **11**, ranking 358 with GIST but 1212 without GIST, had a  $K_d$  value of 1.3  $\mu\text{M}$ . Compound **11** has a slightly unfavorable GIST energy of 0.28 kcal/mol, owing

to its calculated displacement of a bound water. Nevertheless, its rank improved relative to the non-GIST docking screen, reflecting even larger penalties for other, formerly higher-ranking molecules. On determination of its structure to 1.54 Å resolution, the crystallographic geometry corresponded closely to that predicted by docking, with an RMS deviation of 0.44 Å (**Table 1.1**, **Figure 1.3**). Similar effects were seen for compounds **8**, **10**, **12**, and **14**, whose energy scores were only modestly affected by GIST, and for which docking well-predicted the subsequently determined crystallographic geometry.

Unexpectedly, compounds **3** through **6** were predicted by the GIST docking to interact indirectly with the critical Asp233 via an implicitly ordered water molecule (i.e., an area with a high water displacement penalty). Such a geometry, though not unprecedented for CcP cavity ligands, is rare, as cationic ligands typically ion-pair with this aspartate. In the apo-cavity this aspartate is solvated by one bound water<sup>39,40</sup> whose displacement by cationic moieties, though typical, undoubtedly has an energy cost. Indeed, according to GIST such penalty is incurred by molecules like **7**, which dock to maximally displace these waters and ion-pair with the aspartate. Conversely, compounds **3** through **6** dock so as to retain these waters, and compound **3**, instead of ion-pairing with Asp233, the molecule flips its imidazole to hydrogen bond with the carbonyl oxygen of Leu177 and only interacts, via the other side of the imidazole, with Asp233 through a water network. This surprising prediction was confirmed crystallographically: the imidazole interacts with the Leu177 and an ordered water molecule is unambiguously present in the electron density (**Figure 1.3**). Indeed, even



the placement of this bridging water substantially agrees with the GIST calculation, differing only by 0.7 Å. The relatively poor ranks of molecules like **3** when the GIST term is left out is explained by their more distant electrostatic interaction with Asp233 versus molecules that ion pair with it, uncompensated by the advantage of leaving the ordered water molecules undisplaced—a term only modeled by including the GIST penalty. That said, inclusion of the GIST term did not always get this balance correct. Compounds **1** and **9**, though predicted to interact directly with the aspartate, also flip to interact with the Leu177 carbonyl crystallographically (**Figure 1.3**); i.e., even with the GIST term, the correct balance between ion-pairing and water displacement was not achieved. We also note that compounds that do ion-pair with Asp233 typically bind 10-fold tighter than those that bind via water-mediated interactions (**Table A.1.8**).

Compounds **1** and **2** were chosen because inclusion of the GIST term changed their docked geometries. Compound **2** docks to hydrogen-bond with Asp233 while only partly impinging on what are, according to GIST, hard-to-displace water molecules (still incurring a GIST penalty of 2 kcal/mol). In the non-GIST docking, conversely, **2** flips and shifts such that its quinolone nitrogen hydrogen-bonds with the backbone oxygen of Gly178 while its amine hydrogen-bonds with Asp233 and its methyl occupies an unfavorable water site near the heme. The two poses differ by an RMSD of 3.2 Å. In the subsequently determined CcP-ga/2 crystal structure, **2** adopts a geometry that closely agrees with the GIST pose (RMSD of 0.3 Å), but differs by 3.2 Å from the non-GIST docking pose (**Figure 1.3, Table 1.1**). For three compounds, **1**, **9** and **10**, however, we consider the crystallographic complexes to be different from either the

with- or the without-GIST docking pose, although none exceed the commonly-used cut-off of 2 Å RMSD (**Table 1.1**, **Figure 1.3**).

**Anti-GIST.** Compounds **15**, **16**, and **17** ranked much better *without* the GIST term than with it, and their GIST-based ranks, between 6000 and 15,000, would have put them outside the range normally considered as viable for screens of this size; all three sterically complemented the binding site well. Whereas we could determine neither an affinity nor a crystal structure under high soaking concentrations for compound **17**, compounds **15** and **16** either bound very weakly, worse than 5 mM, or undetectably. This is consistent with their GIST-based deprioritization, owing to their displacement of well-bound water molecules from the cavity. It is interesting to note that the benzimidazole of **15** and the imidazole of **16** are both common among CcP-ga ligands (**Table 1.1** and previous studies<sup>38,39,41</sup>). Hence, this anti-prediction is not simply a matter of trivial functional group bias or ionization, indeed, we ourselves expected these molecules to bind, but seems to reflect detailed assessment of fit and presumably water displacement.

## 1.5 Discussion

Inhomogeneous Solvation Theory (IST) has been enthusiastically greeted as a way to model the role of bound water molecules in ligand discovery<sup>25,27,28,31</sup>; it has been widely incorporated into discovery methods.<sup>34-37</sup> Despite its successes,<sup>4,26,27,29</sup> the method has not been tested in prospective, controlled discovery screens at atomic resolution. Three key observations emerge from this study. First, the inclusion of a

water displacement energy noticeably improved the prospective docking screens. Of the molecules prioritized by the water-displacement term, 13 of 14 bound when tested, and one of these, compound **11**, was the most potent ligand yet found for the CcP-ga cavity, with a  $K_d$  of 1.3  $\mu\text{M}$  (ligand efficiency of 1.0 kcal/mol/atom). Correspondingly, of the three molecules ranked higher by the non-GIST versus the GIST docking, none could be shown to bind. Second, the newly-predicted molecules were often right for the right reasons. The docking poses that were based on the water-displacement term corresponded closely to the crystallographic results in six of nine structures. Compellingly, in the CcP-ga/**3** complex, the ligand adopts an unusual pose that does not interact directly with the crucial Asp233, but rather docks to conserve a hard-to-displace, bridging water, as predicted by the GIST energetics. Third, and notwithstanding these favorable results, the IST term, at least in this implementation, had a modest effect in overall ranking, and can introduce its own errors. The term had little effect on retrospective enrichment against the DUD-E benchmark, and there remained remarkable overlap between the top 1000 docking-ranked ligands with and without the term in the CcP-ga screens (**Figure 1.2** Venn diagrams). Also, in three of the nine new crystal structures there were important differences between the GIST-based docking poses and the experimental results. While several of the newly predicted molecules were potent both by the standards of the site and by ligand efficiency, several others were of modest affinity compared to other ligands previously discovered for this cavity.

The ability to prioritize new molecules and to deprioritize unlikely ones is among the strongest results to emerge from this study. Compellingly, 13/14 molecules selected using GIST bind, while none of the GIST-deprioritized molecules did so. Including the GIST term accounts for penalties of displacing water upon ligand binding, which can change both rank and pose. These changes can reveal molecules that would otherwise not have been prioritized for testing. Such molecules include those that replace the hallmark hydrogen bond with Asp233 with an alternative pose that exploits a costly-to-displace water to mediate this ionic interaction, as for compounds **3**, **9** and **10**. Just as important, including the GIST term deprioritizes decoys we would otherwise have ranked highly, like molecules **15-17**.

Often, the GIST-predicted molecules were right for the right reasons; six of nine crystal structures corresponded closely to the docking predictions. This is most striking in those structures in which the GIST term correctly predicted an ordered water molecule that would be costly to displace, favoring a ligand geometry where such a water would be included in the complex with the ligand. Two notable examples are compound **2**, where the GIST-predicted pose differed substantially from that without the GIST term, and was confirmed by subsequent crystallography, and compound **3**, whose crystal structure confirms a water-mediated interaction with Asp233 and an unusual interaction with the carbonyl oxygen of Leu177 (**Figure 1.3**). The water site that **3** retains is one of the most favorable in the cavity; summing up the voxels that contribute to it leads to 4.3 kcal/mol in the GIST calculation. Similarly, compounds **8**, **11**, and **12** interact with a water network toward the pocket entrance that is implicitly predicted by

the GIST grids (**Figure A.1.7**, regions s5-s7); in the CcP-ga/8 complex three crystallographic waters correspond to regions s5-s7 from those predicted by GIST.

Notwithstanding these successes, inclusion of an inhomogenous solvation term only improves docking so far. The GIST term failed to correctly predict the poses of compounds **9** and **10**, and several compounds prioritized by GIST, like **3**, **5**, **6**, and **8**, had  $K_d$  values  $>1\text{mM}$ , which is weak for cavity ligands, if still decent by ligand efficiency (**Tables 1.1**). Retrospectively, at best a modest improvement in enrichment was observed in the benchmarking screens on 25 DUD-E<sup>42</sup> targets (**Figure A.1.6**), and there was substantial overlap among the top-scoring ligands in docking screens with and without the GIST term (**Figure 1.2**). Partly these effects reflect the small magnitude of the net GIST energies: for the top 100 docked molecules from a library screen the term averaged 12% of the overall DOCK3.7<sup>43</sup> energy score in these systems (6 kcal/mol at a 0.5 GIST weighting). This is small enough that the term could be overwhelmed by the errors in other docking terms,<sup>50</sup> reducing its impact. Intriguingly, its beneficial effects were greatest in those benchmarking sets that had a mixture of favorable and unfavorable water sites. Mechanically, at least as implemented here, the GIST term is costly, increasing the time of a docking screen by on average six-fold (**Table A.1.9**), though there may be ways to avoid this cost.

These caveats should not distract from the main observations of this study – the ability of GIST to meaningfully improve large library docking screens. The inclusion of a water displacement term successfully prioritized molecules that did bind on testing, and

it deprioritized those that were found not to, in the teeth of high rankings from the identical scoring function that did not include the GIST term and even our own expectations. Overall, docking with the GIST term led to a 93% hit-rate, with 6-of-9 crystallographic structures in agreement with the docking predictions. The contrast between successful prospective and mediocre retrospective prediction partly reflects the biases towards good performance already baked into the benchmarking sets, however unintentionally. It also reflects our reluctance to optimize the weighting of the scoring function terms for optimal retrospective performance, aware of the oft-described trade-offs between retrospective optimization and prospective prediction.<sup>51</sup> Finally, it is worth noting that in implementing GIST we only considered the energetic consequences of displacing ordered waters, and did not model the specific interactions between ligands and such waters, which play a role in most protein-ligand complexes.<sup>6,7,38,52,53</sup> Here, such interacting waters, which can appear with a ligand to bridge between it and the protein surface, were only implicitly modeled as high-energy, hard-to-displace regions. Including bridging waters explicitly would add new favorable interactions to ligand recognition, adding to the currently small magnitude water term. Even without such bridging waters, this study does support the pragmatism of including a displaceable water energy term like IST, which can materially improve the success of docking ligand prediction and geometry.

## 1.6 Methods

**Experimental affinities and structures.** The protein was purified and crystallized as described<sup>39</sup>. The crystallographic protein-ligand complexes were deposited at the PDB as 5U60 (**1**), 5U5W (**2**), 5U5Z (**3**), 5U61 (**8**), 5U5Y (**9**), 5UG2 (**10**), 5U5X (**11**), 5U5U (**12**), 5U5V (**14**). Affinities were measured at least in duplicate, by monitoring the shift of the heme Soret band.

**Molecular dynamics.** MD was conducted and analyzed with AMBER14.<sup>54</sup> The program *tleap* was used to prepare all proteins for the simulations: the protein systems were placed in a box of TIP3P water, such that all atoms were at least 10 Å from the boundary of the box. For CcP-ga, 10 crystallographically-observed waters were included in or near the binding site. The heme was parameterized as previously<sup>55</sup> (**Table A.1.10**).

The module PMEMD.cuda<sup>56</sup> was used to carry out simulations on GPUs (GeForce GTX 980). The equilibration run consisted of two minimizations of up to 6000 steps followed by six 20 ps runs at constant volume where the temperature of the simulation was raised from 0 K to 298.15 K (**Figure 1.1D**). Langevin dynamics<sup>57</sup> were used to maintain the temperature of the simulation with a collision frequency of 2.0 ps<sup>-1</sup>. Next, a constant pressure (NPT) run allowed the volume of the box adjust for 5 ns to maintain 1 bar of pressure. Finally, constant volume (NVT) simulations were performed for 5 ns, under the same conditions as the subsequent production simulations. Production NVT simulations were performed for 50ns. All protein heavy atoms were restrained with a 5 kcal/mol/Å<sup>2</sup> force constant. The Shake algorithm<sup>58</sup> was used with a

2 fs time step. Periodic boundary conditions were applied and the Particle Mesh Ewald<sup>59</sup> method was used to calculate long-range electrostatics.

**GIST grids.** GIST grids were generated using the Cpptraj<sup>60,61</sup> trajectory analysis program from Am bertools14<sup>54</sup> by processing the 50 ns trajectories with a grid spacing of 0.5 Å. The grids were combined using python scripts that are available at [https://github.com/tbalius/GIST\\_DX\\_tools](https://github.com/tbalius/GIST_DX_tools) and will be made available with the next DOCK release.

**Docking.** Scripts and programs in the DOCK3.7 distribution<sup>43</sup> were used to prepare the receptors and ligand databases for docking and to carry out the library screens. Blastermaster.py was used to prepare the protein. For GIST, proteins were aligned using Chimera<sup>62</sup> into the simulation's frame of reference before DOCK preparation. Root-mean-square deviations (RMSDs) were calculated with the Hungarian algorithm in DOCK6.6.<sup>63</sup>



## References

1. Ringe, D. What makes a binding site a binding site? *Curr Opin Struct Biol* **5**, 825-829 (1995).
2. Mattos, C. *et al.* Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J Mol Biol* **357**, 1471-1482, doi:10.1016/j.jmb.2006.01.039 (2006).
3. Landon, M. R. *et al.* Detection of ligand binding hot spots on protein surfaces via fragment-based methods: application to DJ-1 and glucocerebrosidase. *J Comput Aided Mol Des* **23**, 491-500, doi:10.1007/s10822-009-9283-2 (2009).
4. Bodnarchuk, M. S. Water, water, everywhere... It's time to stop and think. *Drug Discovery Today* **21**, 1139-1146, doi:<http://dx.doi.org/10.1016/j.drudis.2016.05.009> (2016).
5. Österberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. & Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Structure, Function, and Bioinformatics* **46**, 34-40, doi:10.1002/prot.10028 (2002).
6. Verdonk, M. L. *et al.* Modeling Water Molecules in Protein–Ligand Docking Using GOLD. *Journal of medicinal chemistry* **48**, 6504-6515, doi:10.1021/jm050543p (2005).
7. Huang, N. & Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **51**, 4862-4865, doi:10.1021/jm8006239 (2008).
8. Spyrikis, F. & Cavasotto, C. N. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site

- water molecules description. *Arch. Biochem. Biophys.* **583**, 105-119, doi:<http://dx.doi.org/10.1016/j.abb.2015.08.002> (2015).
9. Bayden, A. S., Moustakas, D. T., Joseph-McCarthy, D. & Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *Journal of Chemical Information and Modeling* **55**, 1552-1565, doi:10.1021/ci500746d (2015).
  10. Sindhikara, D. J. & Hirata, F. Analysis of Biomolecular Solvation Sites by 3D-RISM Theory. *The Journal of Physical Chemistry B* **117**, 6718-6723, doi:10.1021/jp4046116 (2013).
  11. Kovalenko, A. & Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **290**, 237-244, doi:[http://dx.doi.org/10.1016/S0009-2614\(98\)00471-0](http://dx.doi.org/10.1016/S0009-2614(98)00471-0) (1998).
  12. Beglov, D. & Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **101**, 7821-7826, doi:10.1021/jp971083h (1997).
  13. Dzubiella, J., Swanson, J. M. J. & McCammon, J. A. Coupling nonpolar and polar solvation free energies in implicit solvent models. *J. Chem. Phys.* **124**, 084905, doi:<http://dx.doi.org/10.1063/1.2171192> (2006).
  14. Zhou, S. *et al.* Variational Implicit-Solvent Modeling of Host–Guest Binding: A Case Study on Cucurbit[7]uril. *Journal of Chemical Theory and Computation* **9**, 4195-4204, doi:10.1021/ct400232m (2013).

15. Fennell, C. J., Kehoe, C. W. & Dill, K. A. Modeling aqueous solvation with semi-explicit assembly. *Proc. Natl. Acad. Sci. USA* **108**, 3234-3239, doi:10.1073/pnas.1017130108 (2011).
16. Baroni, M., Cruciani, G., Sciabola, S., Perruccio, F. & Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *Journal of Chemical Information and Modeling* **47**, 279-294, doi:10.1021/ci600253e (2007).
17. Mason, J. S. *et al.* High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. *In Silico Pharmacol.* **1**, 23, doi:10.1186/2193-9616-1-23 (2013).
18. Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *The Journal of Physical Chemistry B* **113**, 13337-13346, doi:10.1021/jp9047456 (2009).
19. Gerogiokas, G. *et al.* Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *J. Chem. Theory Comput.* **10**, 35-48, doi:10.1021/ct400783h (2014).
20. Michel, J. *et al.* Evaluation of Host–Guest Binding Thermodynamics of Model Cavities with Grid Cell Theory. *Journal of Chemical Theory and Computation* **10**, 4055-4068, doi:10.1021/ct500368p (2014).
21. Jorgensen, W. L. & Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *Journal of Chemical Theory and Computation* **4**, 869-876, doi:10.1021/ct800011m (2008).

22. Ross, G. A., Bodnarchuk, M. S. & Essex, J. W. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *Journal of the American Chemical Society* **137**, 14930-14943, doi:10.1021/jacs.5b07940 (2015).
23. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B* **102**, 3531-3541, doi:10.1021/jp9723574 (1998).
24. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *The Journal of Physical Chemistry B* **102**, 3542-3550, doi:10.1021/jp972358w (1998).
25. Li, Z. & Lazaridis, T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *Journal of the American Chemical Society* **125**, 6636-6637, doi:10.1021/ja0299203 (2003).
26. Abel, R. *et al.* Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **6**, 1049-1066, doi:10.1002/cmdc.201000533 (2011).
27. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **130**, 2817-2831, doi:10.1021/ja0771033 (2008).
28. Nguyen, C. N., Cruz, A., Gilson, M. K. & Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J Chem Theory Comput* **10**, 2769-2780, doi:10.1021/ct401110x (2014).

29. Horbert, R. *et al.* Optimization of Potent DFG-in Inhibitors of Platelet Derived Growth Factor Receptor $\beta$  (PDGF-R $\beta$ ) Guided by Water Thermodynamics. *Journal of medicinal chemistry* **58**, 170-182, doi:10.1021/jm500373x (2015).
30. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* **137**, 044101, doi:10.1063/1.4733951 (2012).
31. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proceedings of the National Academy of Sciences* **104**, 808-813, doi:10.1073/pnas.0610202104 (2007).
32. Li, Z. & Lazaridis, T. in *Computational Drug Discovery and Design* Vol. 819 *Methods in Molecular Biology* (ed Riccardo Baron) Ch. 24, 393-404 (Springer New York, 2012).
33. Snyder, P. W. *et al.* Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proceedings of the National Academy of Sciences* **108**, 17889-17894, doi:10.1073/pnas.1114107108 (2011).
34. Murphy, R. B. *et al.* WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand–Receptor Docking. *Journal of medicinal chemistry* **59**, 4364-4384, doi:10.1021/acs.jmedchem.6b00131 (2016).
35. Repasky, M. P. *et al.* Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *Journal of*

- Computer-Aided Molecular Design* **26**, 787-799, doi:10.1007/s10822-012-9575-9 (2012).
36. Sun, H., Zhao, L., Peng, S. & Huang, N. Incorporating replacement free energy of binding-site waters in molecular docking. *Proteins* **82**, 1765-1776, doi:10.1002/prot.24530 (2014).
37. Uehara, S. & Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **21**, 1604 (2016).
38. Barelier, S. *et al.* Roles for ordered and bulk solvent in ligand recognition and docking in two related cavities. *PLoS One* **8**, e69153, doi:10.1371/journal.pone.0069153 (2013).
39. Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **6**, 575-583, doi:10.1038/nchem.1954
40. Fischer, M., Shoichet, B. K. & Fraser, J. S. One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *ChemBioChem* **16**, 1560-1564, doi:10.1002/cbic.201500196 (2015).
41. Rosenfeld, R. J., Hays, A. M., Musah, R. A. & Goodin, D. B. Excision of a proposed electron transfer pathway in cytochrome c peroxidase and its replacement by a ligand-binding channel. *Protein Sci* **11**, 1251-1259, doi:10.1110/ps.4870102 (2002).
42. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better

- Benchmarking. *Journal of medicinal chemistry* **55**, 6582-6594, doi:10.1021/jm300687e (2012).
43. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992, doi:10.1371/journal.pone.0075992 (2013).
  44. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
  45. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582-6594, doi:10.1021/jm300687e (2012).
  46. Mysinger, M. M. *et al.* Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc Natl Acad Sci U S A* **109**, 5517-5522, doi:10.1073/pnas.1120431109
  47. Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem* **6**, 575-583, doi:10.1038/nchem.1954
  48. Terwilliger, T. C. *et al.* Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallographica Section D* **64**, 515-524, doi:doi:10.1107/S0907444908004319 (2008).
  49. Fitzgerald, M. M., Churchill, M. J., McRee, D. E. & Goodin, D. B. Small Molecule Binding to an Artificially Created Cavity at the Active Site of Cytochrome c Peroxidase. *Biochemistry* **33**, 3807-3818, doi:10.1021/bi00179a004 (1994).

50. Tirado-Rives, J. & Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *Journal of medicinal chemistry* **49**, 5880-5884 (2006).
51. van Drie, J., H. Pharmacophore Discovery - Lessons Learned. *Current Pharmaceutical Design* **9**, 1649-1664, doi:<http://dx.doi.org/10.2174/1381612033454568> (2003).
52. Barillari, C., Taylor, J., Viner, R. & Essex, J. W. Classification of Water Molecules in Protein Binding Sites. *Journal of the American Chemical Society* **129**, 2577-2587, doi:10.1021/ja066980q (2007).
53. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat Rev Drug Discov* **14**, 95-110, doi:10.1038/nrd4486
54. AMBER 14 (University of California, San Francisco, 2014).
55. Rocklin, G. J. *et al.* Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *Journal of Molecular Biology* **425**, 4569-4583, doi:<http://dx.doi.org/10.1016/j.jmb.2013.07.030> (2013).
56. Götz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **8**, 1542-1555, doi:10.1021/ct200909j (2012).
57. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics* **65**, 1409-1419, doi:10.1080/00268978800101881 (1988).
58. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics



- of n-alkanes. *Journal of Computational Physics* **23**, 327-341,  
doi:[http://dx.doi.org/10.1016/0021-9991\(77\)90098-5](http://dx.doi.org/10.1016/0021-9991(77)90098-5) (1977).
59. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092,  
doi:<http://dx.doi.org/10.1063/1.464397> (1993).
60. Ramsey, S. *et al.* Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *Journal of Computational Chemistry* **37**, 2029-2037,  
doi:10.1002/jcc.24417 (2016).
61. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095, doi:10.1021/ct400341p (2013).
62. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612,  
doi:10.1002/jcc.20084 (2004).
63. Allen, W. J. & Rizzo, R. C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *Journal of Chemical Information and Modeling* **54**, 518-529, doi:10.1021/ci400534h (2014).

## Gloss to Chapter 2

Though the original implementation of GIST seemed to be successful, there were several drawbacks that limited its regular usage in the lab. These drawbacks were its slowdown of the standard scoring function by 6-fold on average, its inability to be incorporated into Simplex minimization, whose addition significantly improved docking performance, and its small magnitudes that were dwarfed by the other scoring function terms. Trent had different implementations of GIST that he was working on including what he called “blurry sphere GIST”. In this implementation, a new “blurry GIST” grid is generated before docking that takes in the GIST grid as input. In the blurry GIST grid, each voxel, instead of containing the receptor desolvation enthalpy at that individual position, contains the sum of Gaussian-weighted receptor desolvation enthalpies of its neighboring voxels contained within some sphere radius. In this way, the enthalpy at voxels is what a ligand atom would see during docking in the original displacement implementation, but with a Gaussian weighting so that the enthalpies at voxels closer to the center were weighted more, and voxels further away were weighted less heavily. This was done to reduce double-counting of GIST receptor desolvation enthalpies. We decided to generate two blurry GIST grids, one for heavy atoms (1.8 Å radius) and one for hydrogens (1.0 Å radius) to be consistent with the ligand desolvation grids, but also because using two grids instead of one exhibited better agreement with the displacement GIST enthalpies. Because the receptor desolvation enthalpies were pre-computed on the blurry GIST grid, we could use the trilinear interpolation scheme I had implemented for my Poisson-Boltzmann-derived receptor desolvation method described

in the Introduction, resulting in almost no slowdown in docking. Similarly, the quick trilinear interpolation scoring scheme for blurry GIST ensured that it was readily incorporated into the Simplex minimization scheme, which scored molecules by trilinear interpolation of the other three scoring function terms. Thus, we had successfully solved two of the issues of the original displacement GIST scheme – the slowdown in docking time and implementation of GIST into Simplex minimization. However, though we had reached similar magnitudes of GIST enthalpies and similar performance compared with GIST in enrichments, the blurry GIST enthalpies were still quite small relative to the other scoring function energies. This is what inspired Chapter 3 of this thesis.

In further tests, I noticed that when comparing screens of molecules scored with the standard and blurry GIST scoring functions that the addition of Simplex minimization resulted in different minimization paths, such that a substantial amount of molecules could find their best scoring pose in the opposite scoring function. To correct this, I rewrote the DOCK source code so that each pose of each molecule was scored by both scoring functions in a single docking run, thereby cutting the screening time in half by running one virtual screen instead of two, as was done with displacement GIST.

In the following chapter, we run enrichments on 40 DUD-E systems using this fast Gaussian-weighted blurry GIST implementation, compare it to the original displacement GIST scheme in terms of performance and speed, where it performs similarly with no slowdown in docking time. Then we then run a 300 million molecule large scale docking screen on the drug-like model system, AmpC  $\beta$ -lactamase, and experimentally characterize molecules that score better in the blurry GIST scoring function and worse in the standard scoring function (Pro-bGIST), molecules that score

better in the standard scoring function and worse in the blurry GIST scoring function (Anti-bGIST), and molecules that rank in the top 10,000 of either scoring function but whose geometries change. What we find is that the Anti-bGIST and pose-changing molecules have a higher hit rate than the Pro-bGIST molecules, suggesting that only incorporating for desolvation may be insufficient to account for the water energetics in the solvent-exposed AmpC binding site. However, we did have success in using blurry GIST to predict the correct binding geometry for at least one molecule, as confirmed by X-ray crystallography. Further such studies are now underway, and I hope to complete them in the next several weeks. Were this trend to continue over more molecules, without substantial confounds, it may suggest that blurry GIST can be helpful with geometric fidelity of the docking predictions, even though it struggles to improve prioritizing molecules as likely binders. Proper accounting of water energetics may require including the reorganization of water around the protein-ligand complex, which we do in collaboration with Tom Kurtzman's lab at CUNY Lehman. Thus, further work remains in determining how much of water's behavior needs to be modeled in the DOCK3.7 scoring function, and how this relates to the properties of the protein binding site targeted.

## **Chapter 2: Testing a faster implementation of IST in Ligand Discovery**

Reed M Stein<sup>1</sup>, Isha Singh<sup>1</sup>, Mossa Ghattas<sup>2</sup>, Tom Kurtzman<sup>\*2</sup>, Trent E Balius<sup>\*3</sup>, Brian K Shoichet<sup>1\*</sup>

1. Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158
2. Lehman College Department of Chemistry, 250 Bedford Park Blvd West Bronx, NY 10468
3. Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc. PO Box B, Frederick, MD, 21702

\*Corresponding authors

## 2.1 Abstract.

Ordered water in protein binding sites is both displaced and rearranges upon ligand binding, but capturing this behavior is challenging in structure-based ligand discovery. To do so, the statistical mechanics-based inhomogeneous solvation theory has found wide use but the method has seen limited prospective testing. In one set of prospective tests in a simple model cavity, the method did show promise. Here, we extend our previous implementation of a grid-based version of this method, GIST, making it amenable to ultra large library docking, and testing it in a more relevant, drug-like binding site, that of AmpC  $\beta$ -lactamase. This optimized version of GIST, which we call blurry GIST, relies on Gaussian-weighting to precompute GIST desolvation energies prior to docking and can recapitulate the behavior of our previous implementation of GIST with a 12-fold speed up in docking time. While retrospective enrichment was only moderately improved with the addition of blurry GIST, we turned to prospective docking of over 300 million molecules on AmpC to understand how blurry GIST impacts ligand discovery and geometry in this difficult, solvent-exposed site. We selected molecules that were both prioritized and deprioritized on addition of the blurry GIST term for testing. In activity assays, 2/31 molecules prioritized by blurry GIST were found to bind, 8/18 molecules deprioritized by blurry GIST were found to bind, and 9/18 molecules highly ranked in both scoring functions but exhibiting different docking geometries were found to bind. Two crystal structures have been determined with one pose corresponding to that predicted by blurry GIST, whereas the second structure differed from both predictions. While the incorporation of receptor desolvation via blurry GIST may not substantially improve hit rates in complicated solvent-exposed binding

sites, it may accurately predict binding geometries, a topic which we are delving into further by solving more crystal structures and performing protein-ligand complex GIST calculations to understand differences in solvation free energies between docked poses.

## 2.2 Introduction.

Water molecules play significant roles in protein-ligand binding events, contributing to the hydrophobic effect<sup>1-3</sup>, stabilization of protein-ligand complexes through water-mediated interactions<sup>4-6</sup>, as well as entropy-enthalpy compensation through burial and displacement<sup>7-10</sup>. There is a plethora of computational approaches to characterize the location and energetics of water around proteins and ligands including WaterMap<sup>11,12</sup>, STOW<sup>13</sup>, GIST<sup>14-16</sup>, JAWS<sup>17</sup>, and SPAM<sup>18</sup>, and they have been used to predict water structure and compute energetics in and around protein binding sites<sup>19-23</sup>, to characterize changes in ligand potency and selectivity<sup>24-28</sup>, and to predict water reorganization location and energies upon ligand binding<sup>29,30</sup>. Although these methods have been incorporated into docking programs including WaterMap into Glide<sup>31</sup>, solvent properties analysis (SPA) into DOCK3.5.54<sup>32</sup>, grid inhomogeneous solvation theory (GIST) into AutoDock<sup>33</sup>, and recently in our lab, GIST into DOCK3.7<sup>34</sup>, they have not been used in ultra-large prospective library docking screens in drug-like cavities<sup>35</sup>.

One barrier to GIST's incorporation into ultra-large library docking screens is its speed, which we found slows down the performance of DOCK3.7 by 6-fold on average. Another potential issue is the magnitudes of GIST energies, which comprise only 8-12% of the total docking score over 25 Directory of Useful Decoys – Enhanced (DUD-E)<sup>36</sup> systems and cytochrome *c* peroxidase. We wondered whether these small energies

would be meaningful in drug-like cavities where there are more complicated surfaces, polar and charged residues, and water dynamics, or whether they would wash out in the noise of the other three scoring function terms. Lastly, code changes in DOCK3.7 included the incorporation of a Simplex minimization procedure<sup>35,37</sup>, an upgrade that we believed would significantly increase docking time if GIST was incorporated into this scheme.

Given GIST's success in predicting correct binders and poses in the mostly buried model site, cytochrome *c* peroxidase gateless mutant, where 13 of 14 GIST predicted molecules bound, and six of nine crystal structures of ligand complexes corroborated the geometries predicted by GIST, we felt it a worthwhile endeavor to correct these drawbacks. Here, we have implemented a new GIST scoring scheme into DOCK3.7 termed blurry GIST, that eliminates the slowdown and speeds up docking time by 2-fold, matches the magnitudes of GIST energies of the original implementation, and that is readily incorporated into Simplex minimization. We applied this new GIST implementation to 40 DUD-E systems to quantify its effect on ligand enrichment, and then prospectively screened over 300 million molecules against the bacterial enzyme, AmpC  $\beta$ -lactamase, a rigid, solvent-exposed active site containing several polar and charged residues that binds anions. We have previously used this as a model system for understanding new docking methods and for identifying new ligands<sup>35,38-43</sup>. This presents a more challenging system than the mostly buried site of cytochrome *c* peroxidase gateless mutant as AmpC's charged active site directly interacts with solvent, and allows us to determine whether the static representation of water from



blurry GIST can both account for the behavior of water in this site, but also integrate successfully into the current DOCK3.7 scoring function.

After our prospective screen, we purchased molecules that: i) were previously untested at AmpC, ii) experienced substantial rank changes upon addition of the blurry GIST term, or iii) changed geometries after the addition of blurry GIST, resulting in a total of 68 molecules being tested and two crystal structures being solved. We find that molecules that are highly ranked in the docking hit lists have a higher likelihood of binding AmpC, and these molecules are typically penalized and deprioritized upon addition of blurry GIST, but that poses predicted by blurry GIST may be more accurate over the current DOCK scoring function. Regardless, our results suggest that desolvation alone may be insufficient to quantify solvent effects in open sites, and solvent reorganization effects may need to be incorporated.

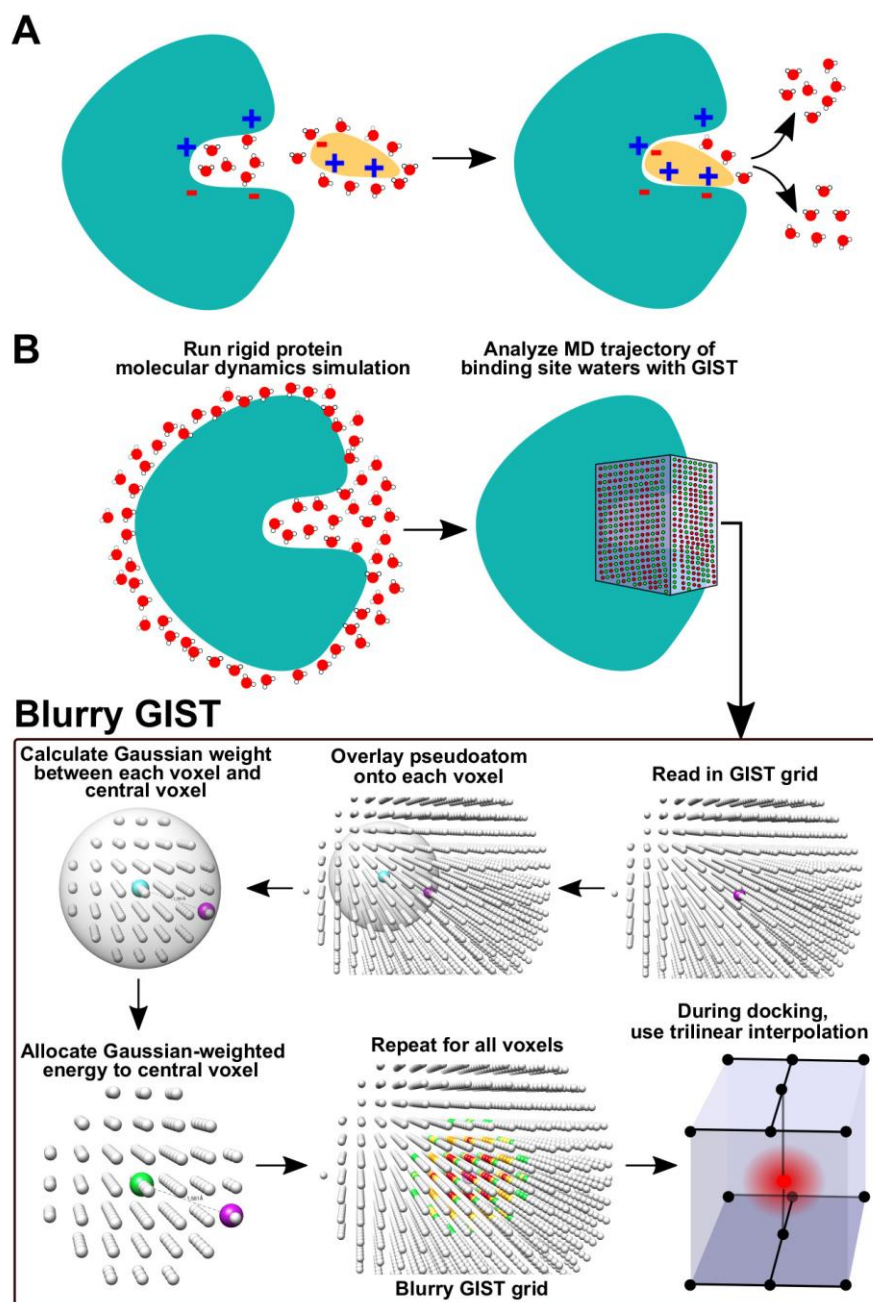
### **2.3 Results.**

Inhomogeneous solvation theory methods rely on energies from a molecular mechanics potential function and snapshots from MD trajectories to calculate water thermodynamics in and around a protein. GIST represents the water thermodynamics of solute-water enthalpy ( $E_{s,w}$ ), water-water enthalpy ( $E_{w,w}$ ), translational entropy ( $TS_{trans}$ ), and orientational entropy ( $TS_{orient}$ ), by discretizing them onto a three-dimensional grid (**Figure 2.1**). In the original implementation of GIST into DOCK3.7<sup>34</sup>, the total receptor desolvation of each molecular pose was calculated by identifying the voxels contained within the van der Waals radii, and then summing up the energies stored at those voxels. The GIST grids most useful in producing the best enrichments included the

solute-water enthalpy and water-water enthalpy grids, with the GIST term set as  $E_{\text{GIST}} = E_{\text{s,w}} + 2 \times E_{\text{w,w}}$ , which includes favorable interactions between water and the protein ( $E_{\text{s,w}}$ ), as well as between pairs of waters within the context of the protein binding site ( $E_{\text{w,w}}$ ), which is referenced to the density-weighted bulk solvent water-water energy for each voxel (see Methods). The water-water energies are multiplied by two to account for the fact that energies at each voxel contain only half the water-water interaction energy, and thus, need to be multiplied by two to recover the full interaction energy. This term does not include the entropy energies, but it has been suggested that the enthalpy terms are more predictive and meaningful<sup>10,16</sup>. As before, the maximum absolute magnitudes of GIST voxels were capped at 3 kcal/mol/Å<sup>3</sup> to reduce the effect of extreme GIST energies and to enhance performance.

To decrease the average time of docking, we devised a scheme for which the receptor desolvation energies could be pre-computed prior to docking, and these pre-computed receptor desolvation energies could be stored on a new grid, which we call the blurry sphere GIST grid (bGIST). In this scheme, the GIST grid serves as an input (**Figure 2.1**). For each voxel in the original GIST grid, a sphere with radius 1.8 Å (representing a heavy atom) or 1.0 Å (representing a hydrogen atom) is overlaid onto the voxel. For each voxel contained within this pseudo-atom, we calculate the distance between that voxel and the central voxel and calculate a Gaussian scaling factor. The receptor desolvation energy of each voxel within the pseudo-atom is scaled by the Gaussian scaling factor and then added onto the central voxel. Thus, each voxel becomes a sum of Gaussian-weighted receptor desolvation energies contained within a pseudo-atom of a specified radius. We use a Gaussian distribution to reduce the

amount of double counting of voxels, as the 0.5 Å grid spacing ensures that voxels will be within the volume of multiple pseudo-atoms' radii. Once the new blurry sphere GIST grids are computed, they can be read in during docking, and the GIST energies can be calculated using trilinear interpolation on the heavy atom blurry sphere GIST grid for heavy atoms, and the hydrogen atom blurry sphere GIST grid for hydrogen atoms. We tried various values for  $\sigma$  and found that the radius divided by 1.3 and with a weighting of -2.0 for blurry GIST in DOCK3.7 provided the best agreement with dGIST energies (**Figure A.2.1**). Since the implementation of displacement GIST (dGIST), we incorporated a Simplex minimizer into DOCK3.7<sup>35</sup>. Given the simplicity of the new bGIST scoring scheme by utilizing trilinear interpolation, we also ensured that all poses for each molecule would be minimized with blurry GIST energies in addition to van der Waals, electrostatics, and ligand desolvation.



**Figure 2.1. Scheme for incorporating grid inhomogeneous solvation theory. A)** Water fills protein binding sites and surrounds ligands, and must be displaced, or coordinate protein and ligand upon complex association. **B)** As part of GIST, a 50ns molecular dynamics simulation is run on a rigid protein, and the MD trajectory is analyzed to output a GIST grid containing densities, enthalpies, or entropies at voxel positions. In the blurry GIST scheme, a GIST grid is read in as an input and a Gaussian weighting scheme is used (see Methods) to store GIST receptor desolvation energies at voxels. During docking, trilinear interpolation is used to score each atom, and the atomic blurry GIST desolvation energies are summed.

## Retrospective DUD-E results

We had previously prepared 25 DUD-E systems for enrichment calculations and extended this to 40 DUD-E systems for which we identified water molecules in the binding site. In the retrospective docking screens, the standard scoring function without minimization reached an average adjusted log AUC of 17.56, whereas dGIST with a weighting of -0.5 improving upon this by 0.47 with an average adjusted log AUC of 18.03 (**Table 2.1**). Blurry GIST with a weighting of -1.0 in DOCK3.7 and without minimization improved enrichment by 0.39 over the standard scoring function with an average adjusted log AUC of 17.95. After including minimization in the standard scoring function, enrichment improved to 20.26 average adjusted log AUC, while bGIST with a weighting of -1.0 and minimization improved to an average adjusted log AUC of 20.75, an improvement of 0.49. Thus, blurry GIST improvement is additive with the improvement from Simplex minimization. Blurry GIST with minimization improves over the original standard scoring function without minimization by 3.19, and the original dGIST implementation, which isn't compatible with Simplex minimization, by 2.72.

**Table 2.1. Adjusted logAUC values comparing GIST performance**

DOCK Type	Better (>1%)	Same	Worse (<1%)	Average adjusted logAUC	Average adjusted logAUC relative to STD without minimization
STD no min				17.56	
Displacement GIST (-0.5x)	14	9	17	18.03	+0.47
STD + min	28	5	7	20.26	+2.70
bGIST no min	16	7	17	17.95	+0.39
bGIST + min	34	1	5	20.75	+3.19
STD combinatorial (1x)	29	5	6	20.26	+2.70
STD combinatorial (2x)	31	4	5	20.44	+2.89
bGIST combinatorial (1x)	34	2	4	20.79	+3.23
bGIST combinatorial (2x)	31	4	5	20.50	+2.94

After docking, we noticed that when including minimization, some molecules that were scored using the standard scoring function, which does not include blurry GIST energies, could attain better energetic poses after rescoring with blurry GIST than the same molecule when scored using the blurry GIST scoring function during docking (**Figure A.2.2**). We found that this was due to the Simplex minimization, as this effect does not occur with the minimization turned off, and this was likely due to the energy landscape changing with the incorporation of blurry GIST. To potentially correct this, we attempted Monte Carlo optimization using the Metropolis criterion<sup>44</sup> instead of Simplex minimization, but found that it suffered from the same issues, though it could reduce the number of high energy difference outliers. We then modified DOCK3.7 to score each pose of each molecule for both scoring functions in a single docking run (see Methods and **Figure A.2.3**). The benefits of this scheme are two-fold: one, it ensures that the

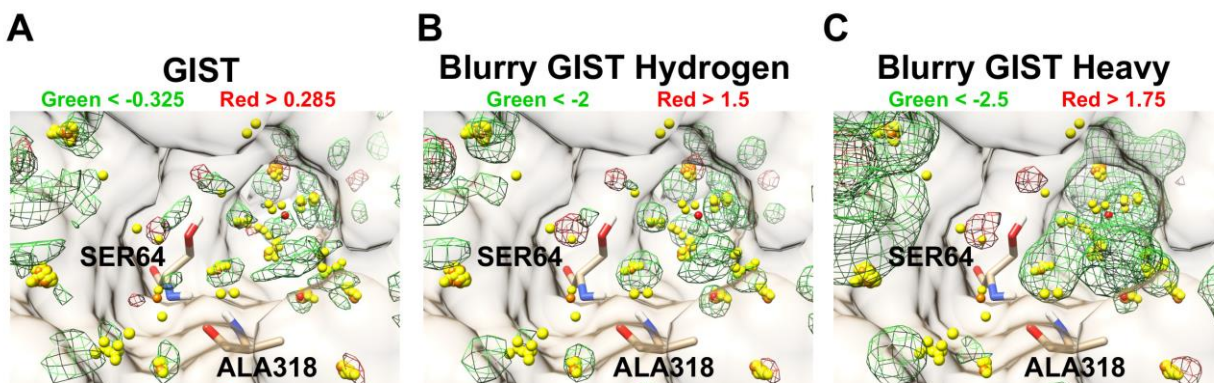
best scoring pose for both scoring functions is chosen, regardless of whether the pose was originally generated in the standard or blurry GIST docking; two, it speeds up the docking calculation by two-fold, making it so we only need to run one docking screen, instead of two separate screens for the two scoring functions, as we did previously. After incorporating this change, the retrospective docking was performed again with -1.0 and -2.0 docking weights. The combinatorial standard scoring function with minimization reached an average adjusted log AUC of 20.26 and the combinatorial blurry GIST scoring function with -1.0 weighting and with minimization reached an average adjusted log AUC of 20.79, a 0.53 improvement, while with the -2.0 weighting, the improvement was almost negligible at 0.05 average adjusted log AUC. The absolute value of the blurry GIST term at a weighting of -1.0 was about 4.4 kcal/mol for the top 100 ranked docked molecules in the 40 DUD-E targets, amounting to about 5% of the total docking score of these molecules, while for a blurry GIST weighting of -2.0, the absolute value was 7.8 kcal/mol which amounts to about 9% of the total docking score for these molecules. Thus, though the energetic contribution of blurry GIST remains similar to the original implementation of displacement GIST, blurry GIST's improvement in average adjusted log AUC mirrored that of dGIST's modest improvement.

## Prospective AmpC results

Given the fact that the new blurry GIST did not diminish performance, and that GIST was now fast enough to use for large scale docking, we chose to perform an ultra-large library docking screen on the bacterial enzyme, AmpC, to predict novel ligands and their geometries. This protein has been heavily studied for mechanism and biophysics, and we have consistently used it to understand ligand binding in a drug-like cavity<sup>39,41-43,45</sup>. The binding site is open to solvent, contains anionic and cationic residues, and binds anionic ligands, many containing a carboxylate or phenolate moiety interacting with the oxyanion hole, which would allow us to determine if the new blurry GIST energies were in balance with the electrostatics, van der Waals, and ligand desolvation energies in the standard scoring function. Multiple crystal structures have been determined of AmpC, and waters from 96 of these structures were collected (**Figure 2.2**), showing that most of these water clusters are well-predicted by GIST including the water site coordinated by the backbone amides of Ser64 and Ala318 termed the “oxyanion hole”, where anionic charges of AmpC ligands bind. Interesting to note is that due to the polar and charged nature of the active site, almost all these GIST-predicted water sites are more favorable enthalpically than bulk solvent, such that ligands that displace waters in the AmpC active site will be penalized by GIST. We found that even with the majority penalty site from displacement and blurry GIST, we could improve enrichment by over 2% adjusted log AUC relative to the standard scoring function. In the prospective screen, we utilized the combinatorial scoring function with a -2.0 bGIST weighting as this exhibited a higher improvement in enrichment (+2.31) than the -1.0 bGIST weighting for AmpC retrospective results, and the magnitude of the



bGIST energies were larger, which we reasoned, would generate larger differences between the two scoring functions. For a fair comparison and to understand the specific contribution of the blurry GIST term to docking, we compare standard combinatorial (2x) and blurry GIST combinatorial (2x) in our prospective screen molecule ranks, which have a small 0.05 difference in enrichment retrospectively (**Table 2.1**).



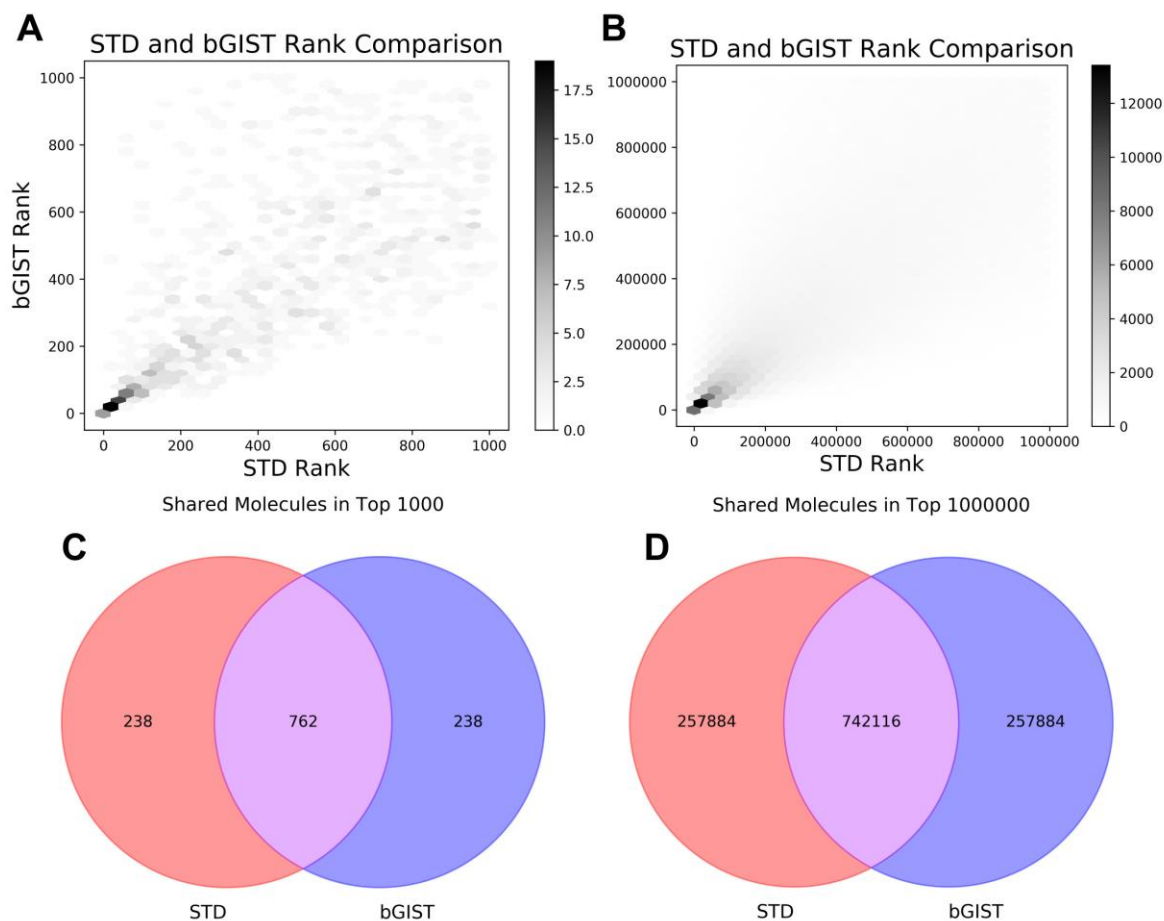
**Figure 2.2. Comparing experiment to GIST-predicted hydration sites.**

A) The GIST enthalpy ( $E_{s,w} + 2 E_{w,w}$ ) grid referenced to bulk solvent. Red spheres, orange spheres, and yellow spheres are crystallographic water oxygens from 96 AmpC  $\beta$ -lactamase crystal structures with B-factors less than  $10 \text{ \AA}^2$ , between  $10$  and  $20 \text{ \AA}^2$ , and between  $20$  and  $30 \text{ \AA}^2$ , respectively. Green mesh represents favorable GIST enthalpies and red mesh represents unfavorable GIST enthalpies, relative to bulk solvent. Units are in  $\text{kcal/mol/\AA}^3$ . B) The blurry GIST hydrogen grid using a pseudo-atom radius of  $1.0 \text{ \AA}$  using the GIST enthalpy grid referenced to bulk solvent as an input. C) The blurry GIST heavy grid using a pseudo-atom radius of  $1.8 \text{ \AA}$  using the GIST enthalpy grid referenced to bulk solvent as an input.

We docked a subset of the ZINC15 (<http://zinc15.docking.org>) that had favorable physical properties ( $\text{cLogP} \leq 3.5$  and  $\text{MW} \leq 400 \text{ Da}$ ) with the combinatorial scoring scheme, which minimizes poses generated from the standard and blurry GIST scoring functions, rescores them against the opposite scoring function, and chooses the best scoring pose for each molecule and scoring function. This library contained over 300 million molecules, most of which were make-on-demand compounds from the Enamine

REAL set from ZINC15<sup>46</sup>. Of these, more than 271 million molecules successfully scored. An average of 4082 orientations, and for each orientation, an average of 563 conformations was sampled, amounting to over 198 trillion protein-ligand complexes, that were scored against both scoring functions. The calculation time was 161,230 core hours, or 4.49 calendar days on 1,500 cores.

In the top 1,000 molecules of the standard screen, 762 of these molecules were also found in the top 1,000 of the blurry GIST screen, while in the top 1 million molecules of the standard screen, over 740,000 molecules were shared, though the correlation in the molecules' ranks was weak, and the two scoring functions share similar ranks only within the top 100 molecules (**Figure 2.3**). We focused on molecules that experience rank changes of a half-log (3.16-fold) or better, such that a molecule whose rank changes from 35,000<sup>th</sup> to 6345<sup>th</sup>, or from 38,055<sup>th</sup> to 9,121<sup>st</sup> after addition of the blurry GIST term would be prioritized. When considering only the top 1% of the screen (2.7 million molecules), upon addition of blurry GIST, 154,256 molecules were prioritized, while 159,071 molecules were de-prioritized. Additionally, we focused on molecules that ranked in the top 10,000 from either scoring function, and whose geometries changed between the scoring functions.

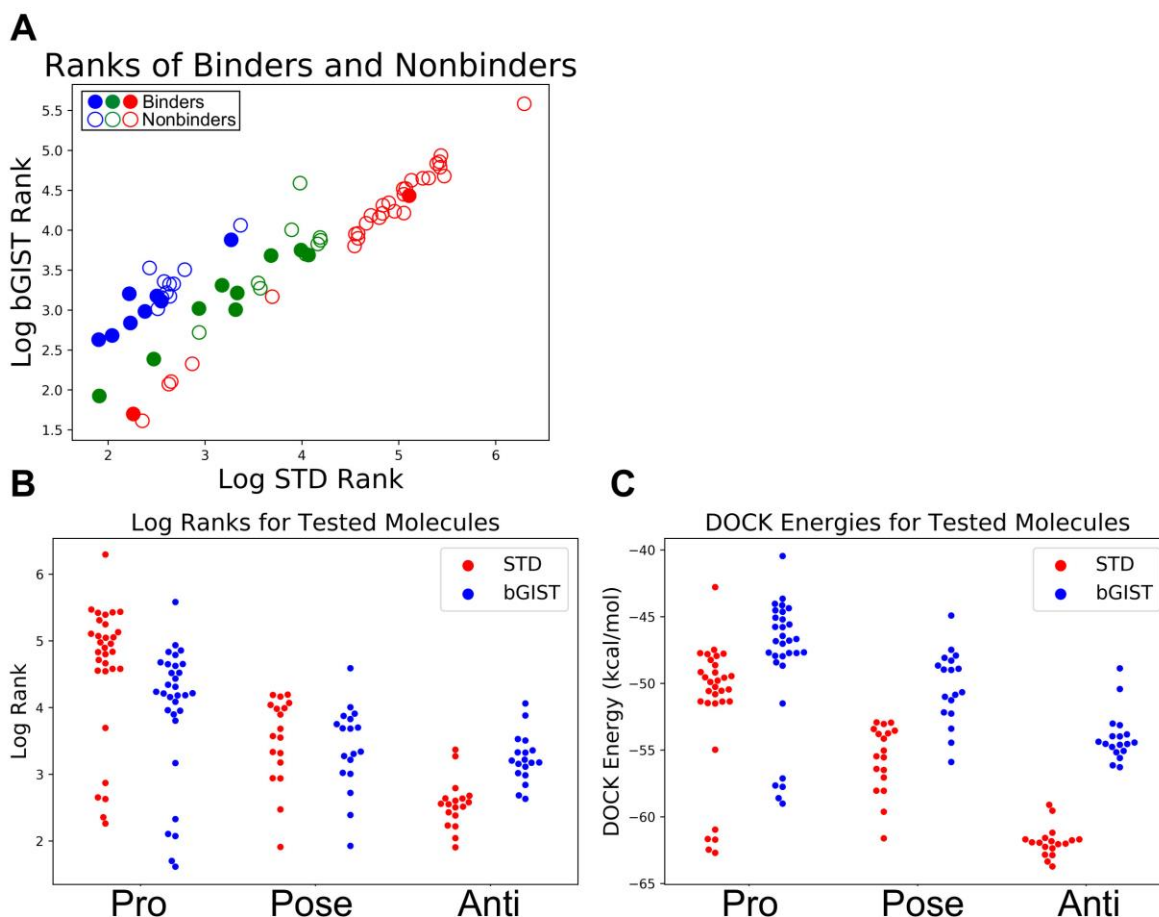


**Figure 2.3. Comparison of large-scale docking molecule ranks.**

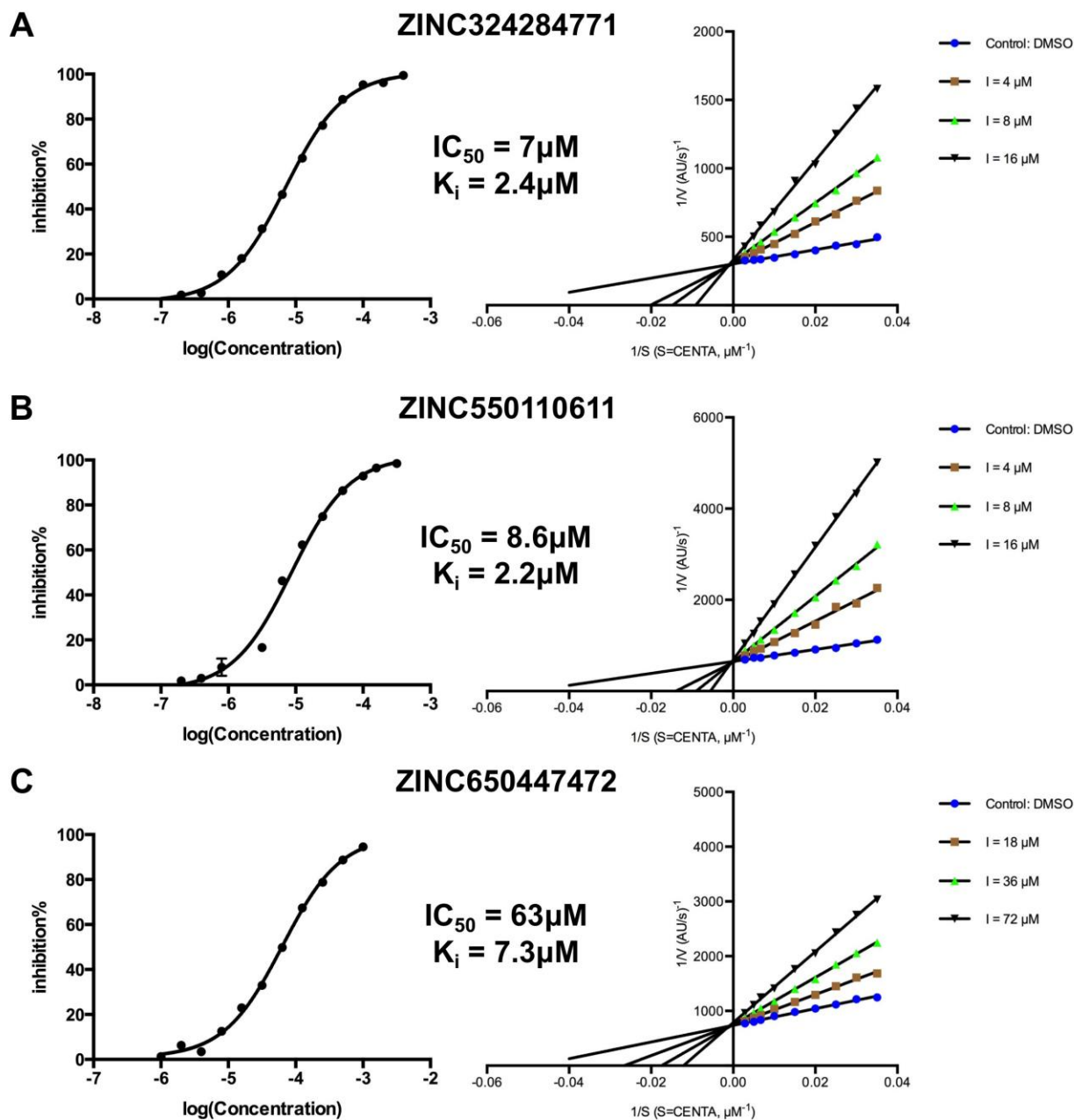
Heat plot showing the correlation of molecule ranks within the best scoring 1,000 (A) and 1,000,000 molecules from the 300 million molecule prospective screen of AmpC using standard and blurry GIST scoring functions. Venn diagrams of molecules shared within the best scoring 1,000 (C) and top 1,000,000 (D).

With these criteria, we initially bought 36 molecules comprising 12 molecules whose ranks improved with blurry GIST (pro-bGIST) as well as 18 molecules whose poses changed substantially between the two scoring functions. Thirty of these molecules were successfully synthesized (83% success rate) and after testing for binding, we identified only 1 molecule from the 12 pro-bGIST molecules and 9 of 18 pose changing molecules that substantially inhibited ( $\geq 50\%$ ) hydrolysis of CENTA by

AmpC at 300  $\mu\text{M}$  as monitored by UV-Vis spectrophotometry (Table 2.2, Figure 2.4). Of the 9 pose changing molecules, two molecules, **ZINC324284771** and **ZINC5550110611**, had  $\text{IC}_{50}\text{s}$  of 2.4 and 2.2  $\mu\text{M}$ , respectively (Figure 2.5).



**Figure 2.4. Comparison of pro-bGIST, anti-bGIST, and pose-changing molecules.** Ranks of molecules in the standard and blurry GIST scoring functions that were prioritized (red), that changed poses (green) and that were deprioritized upon addition of blurry GIST (blue). Filled circles and open circles represent tested molecules that showed  $\geq 50\%$  and  $\leq 50\%$  inhibition of AmpC at 300 $\mu\text{M}$ , respectively. B) Log ranks of tested Pro-bGIST, pose-changing, and Anti-bGIST molecules. C) DOCK Energies for tested Pro-bGIST, pose-changing, and Anti-bGIST molecules.

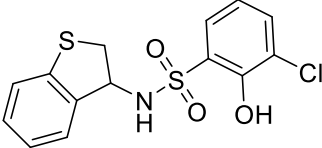
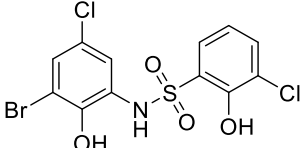
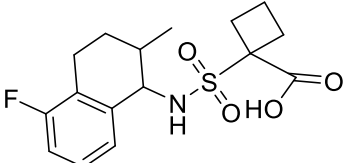
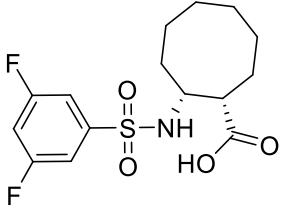
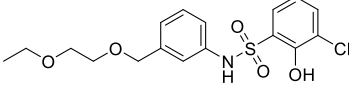
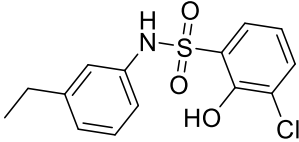
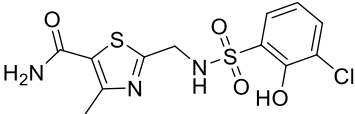
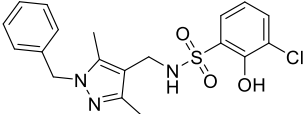
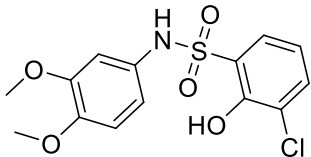
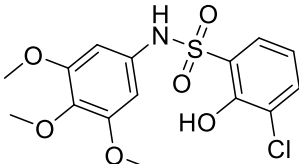
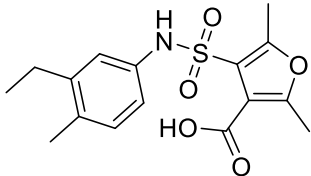
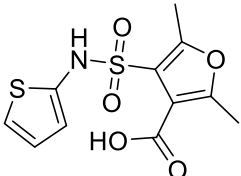


**Figure 2.5. Representative inhibition curves for AmpC inhibitors.** Inhibition curves and Lineweaver-Burk plots for ZINC324284771 (A), ZINC550110611 (B), and ZINC650447472 (C), which are pose-changing molecules that rank within the top 3,000 in both scoring functions (see **Table A.2.1**).

Regardless, the one pro-bGIST molecule, ZINC905040387, represents a new chemotype for AmpC, a cyclobutyl carboxylate whose closest known AmpC ligand is 0.29 by Extended Connectivity Fingerprint 4 (ECFP4) Tanimoto coefficient (Tc). We reasoned that the higher hit rate for pose changing molecules was due to their higher Tanimoto coefficients to known AmpC ligands compared with the pro-bGIST molecules (**Figure A.2.4**). Thus, we decided to extract anionic molecules that contained carboxylates and phenolates and resembled known AmpC molecules from ZINC15, dock these to AmpC with both scoring functions, and re-order them into the original docking hit lists to identify more rank changing molecules within this subset. After this new docking, we found that blurry GIST prioritized 1129 carboxylate- and 79 phenolate-containing molecules, compared to the 6 carboxylate- and 85 phenolate-containing molecules it deprioritized, suggesting that blurry GIST was correctly identifying molecules with strong enough electrostatic interactions with AmpC to overcome the blurry GIST desolvation enthalpies (**Figure A.2.4**). From these we ordered 19 new pro-bGIST molecules as well as 18 molecules that had better ranks without the blurry GIST term (anti-bGIST) and ensured that they overlapped in Tanimoto coefficient space to known AmpC inhibitors. Of these new molecules, only 1 of the 19 pro-bGIST molecules and 8 of the 18 anti-bGIST molecules substantially inhibited ( $\geq 50\%$ ) hydrolysis of CENTA by AmpC at 300  $\mu\text{M}$  (**Figure 2.4**). We noted that of the molecules that substantially inhibited AmpC, the majority of these being pose-changers and anti-bGIST molecules, many of these resembled known ligands, but also that they ranked highly in both scoring functions and had highly favorable DOCK energies. These higher rankings and more favorable energies are consistent with the volume occupied by the molecules

that were prioritized or deprioritized by blurry GIST (**Figure A.2.5**). Molecules that were prioritized by blurry GIST were typically restricted in the space they occupied in the active site, limiting their contact with AmpC to reduce blurry GIST penalties, while deprioritized molecules were more likely to fill the pocket, and make more van der Waals contacts and electrostatic interactions. These new binders ranked from 80 to 127,809 in the standard scoring function, and from 50 to 27,133 in the bGIST scoring function (**Table A.2.1**). Eighteen of the nineteen binding molecules were in the top 10,000 molecules in one or both scoring functions.

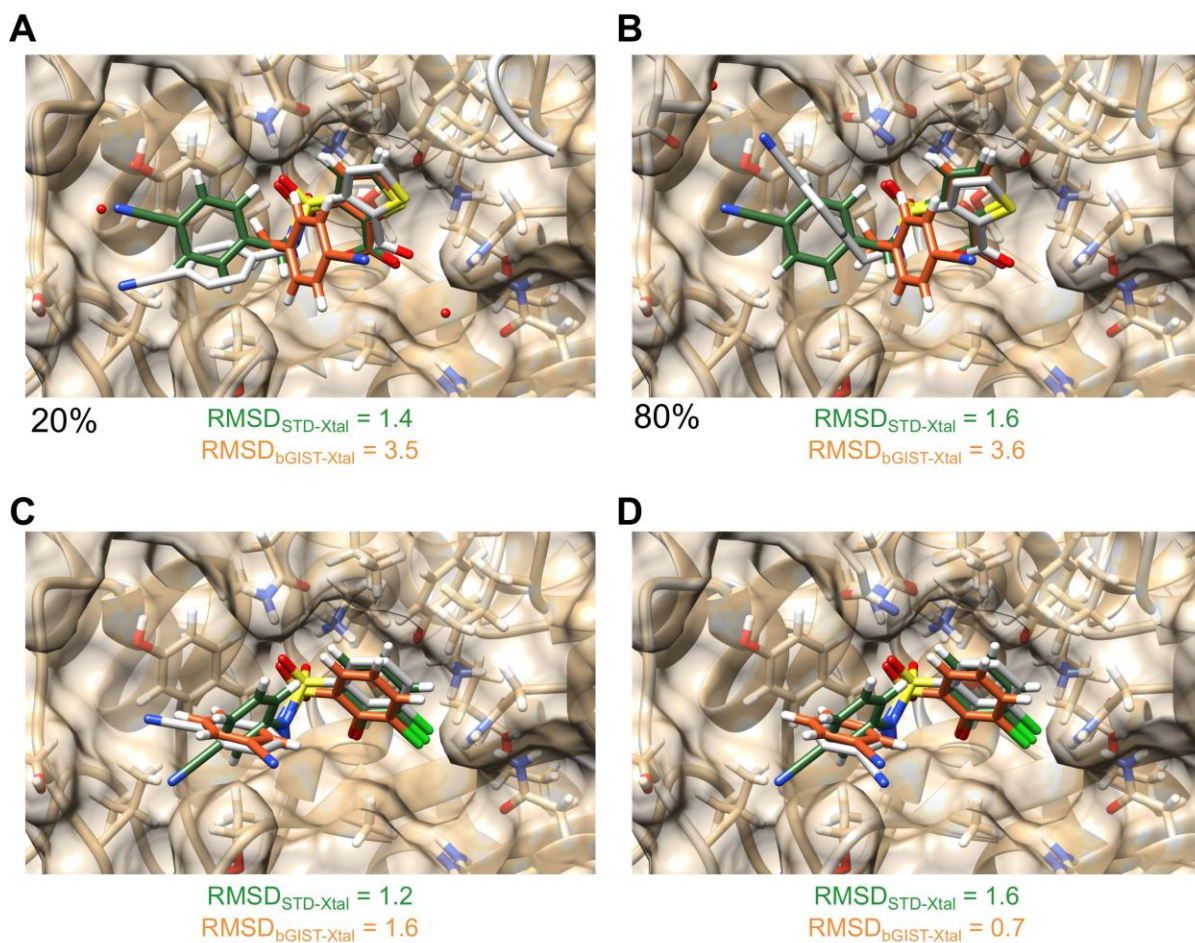
**Table 2.2. A selection of binding molecules.**

Molecule	Inhibition at 300 $\mu$ M	Rank in Standard Scoring	Rank in bGIST Scoring	Rank Log Difference	Closest Known AmpC Inhibitor (ECFP4 Tanimoto Coefficient)
<b>PRO bGIST</b>					
 Z3989663601, ZINC001474992853	74.13	182	50	0.56	 ZINC000549719284 0.43
 Z2903948616, ZINC000905040387	65.21	127809	27133	0.67	 ZINC000580868636 0.29
<b>Anti-bGIST</b>					
 Z2275041991, ZINC000450990100	87.59	165	1600	0.99	 ZINC000581714578 0.71
 Z3989661637, ZINC001561899653	80.03	170	691	0.61	 ZINC001208058246 0.48
<b>Pose Changer</b>				<b>RMSD between STD and bGIST poses (Å)</b>	
 Z2027054051, ZINC000339202812	81.77	296	244	1.5	 ZINC000559249118 0.76
 Z1993712482, ZINC000324284771	98.61	865	1047	1.1	 CHEMBL370041 0.56



To determine whether the unforeseen data was due to the molecular dynamics parameter choices we made, we ran molecular dynamics on AmpC, followed by GIST analysis using different force fields and solvent models, rescored the blurry GIST poses, and re-sorted them based on these new blurry and displacement GIST energies (**Figure A.2.6**). We find that these same pro-bGIST molecules and anti-bGIST molecules reappear, suggesting that the molecular dynamics parameters chosen do not significantly affect the choice of molecules purchased.

We were able to crystallize two molecules that exhibited different geometries upon addition of the blurry GIST term (**Figure 2.6**). These included two nitrile moiety-containing molecules – ZINC37748240, which coordinates the oxyanion hole of AmpC through a carboxylate, and ZINC339208618, which coordinates the oxyanion hole through a phenolate. In the crystal structure of ZINC37748240, we see two poses of the ligand at 20% and 80% occupancies, and in both cases, neither scoring function predicts an identical pose. However, the pose predicted by the standard scoring function is closer to the crystallographic poses than that of the blurry GIST pose, being 1.4 Å and 1.6 Å root mean squared deviation (RMSD) away from the crystallographic pose versus blurry GIST's 3.5 Å and 3.6 Å RMSD. On the other hand, for ZINC339208618, which also contains two poses both at 50% occupancy, we find that blurry GIST predicts an almost identical pose to the crystallographic structure at 0.7 Å RMSD, as it rotates the nitrile benzene roughly 90° relative to that of the standard scoring function's orientation.

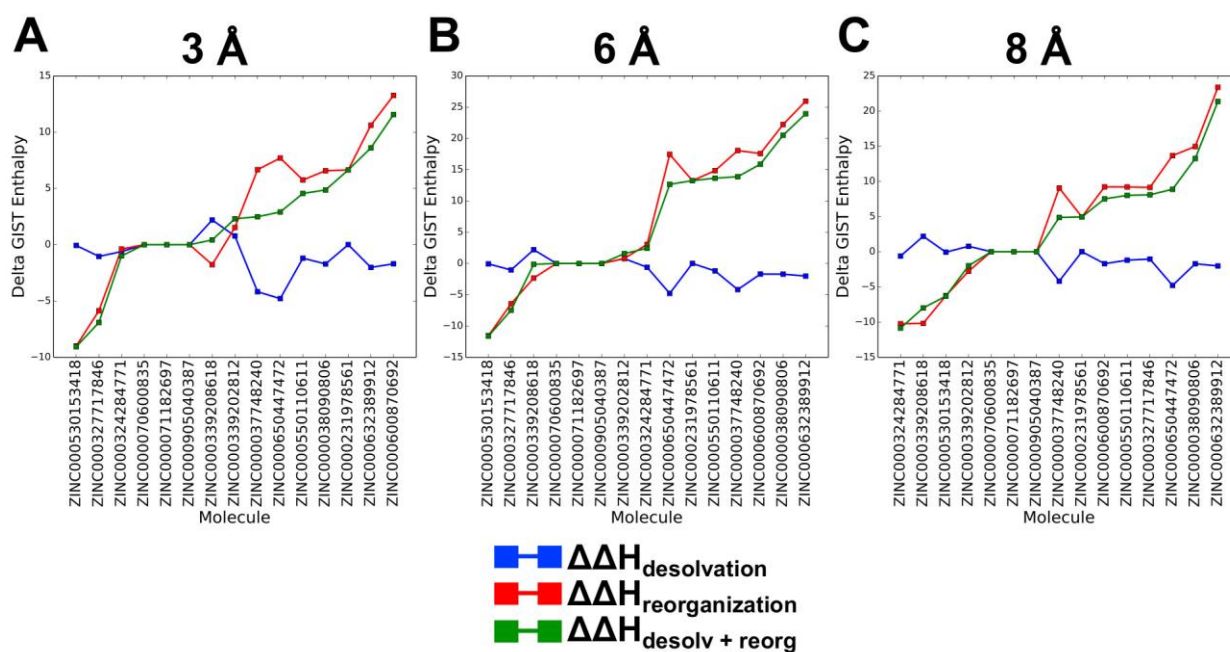


**Figure 2.6. Crystallography of pose-changing molecules.**

ZINC37748240 (1.75 Å) exhibits two conformations (A, 20%, B, 80%) and ZINC339208618 (1.7 Å) exhibits two conformations (C, 50%, D, 50%). The crystal structure poses are shown in grey, while the blurry GIST and standard poses are shown in orange and green, respectively. Root mean squared deviations to the crystal structure poses were calculated using the Hungarian algorithm incorporated into DOCK6.6.

Given the uncertain utility of the blurry GIST term in the DOCK scoring function for predicting binders and poses, we thought that this may be due to the fact that we only considered receptor desolvation of the solvent-exposed AmpC binding site. We reasoned that analyzing the water networks around the poses of the ligands might

potentially help us differentiate binders versus nonbinders. It is known that significant effects on affinity and kinetics can be due to water networks that are not involved in protein-ligand interactions<sup>47</sup>. Additionally, rearrangement of waters and the establishment of new hydrogen bonding networks around the new protein-ligand complex significantly affects the thermodynamics of binding<sup>2,4,9,29,48,49</sup>. Therefore, we ran 150ns ligand-bound molecular dynamics simulations followed by GIST calculations to understand the water energetics around the standard and blurry GIST poses of the ligands (**Figure 2.7**).



**Figure 2.7. Comparison of GIST desolvation and reorganization enthalpies**

The desolvation cost of the standard and blurry GIST poses of pose-changing molecules was computed by running GIST on a 150ns molecular dynamics simulation of the AmpC protein alone, summing up the energies of the voxels displaced using in-house Python scripts. The reorganization enthalpies were determined by running GIST on 150ns molecular dynamics simulations of standard and blurry GIST poses in the context of the AmpC protein and summing up the voxels within 3 Å (A), 6 Å (B), and 8 Å (C) from the ligand pose. We then take the difference in desolvation and reorganization energies between the standard and blurry GIST poses. Negative values indicate that the blurry GIST is more enthalpically favorable.

In this scheme, both standard and blurry GIST poses of the pose-changing molecules are simulated in the presence of the protein for 150 ns, and the GIST grids are generated of these ligand-bound MD simulations. To compute reorganization energies, we sum the enthalpies of the voxels within some distance cutoff outside of the volume of the molecule poses. We used 3, 6, and 8 Å from the ligand surfaces, representing roughly one, two, and three solvation shells from the ligand surfaces. From the receptor alone simulation, we can obtain the desolvation cost of the standard and blurry GIST poses by summing up the voxels contained within the van der Waals radii of the poses in the exact same way that was done for the original implementation of displacement GIST. Taking the difference between the reorganization and desolvation energies for the standard and blurry GIST poses provides us with the difference in solvation enthalpy between these poses. When considering reorganization enthalpies up to 8 Å from the poses in addition to desolvation enthalpies, the blurry GIST pose is favored in only four of the fifteen molecules that we considered for crystallography. Three poses are identical from the standard and blurry GIST scoring functions, thus exhibiting identical desolvation and reorganization energies, while the standard pose is favored for seven molecules. We find that for ZINC37748240, where the crystallographic poses align more closely with the standard scoring function pose, the blurry GIST pose has a less unfavorable desolvation cost, but the standard pose has a much more favorable reorganization enthalpy, such that the sum of the reorganization and desolvation enthalpies strongly favors the pose from the standard scoring function. For ZINC339208618, where the blurry GIST pose is more predictive of the crystallographic geometry, the desolvation cost is less unfavorable for the pose from the

standard scoring function, but the reorganization energies strongly favor the blurry GIST pose, suggesting that again, the reorganization energies determine the pose observed crystallographically, rather than just the desolvation cost alone.

## **2.4 Discussion.**

Four key observations emerge from this study. First, a new implementation of grid inhomogeneous solvation theory that we call blurry GIST can capture the behavior of displacement GIST while speeding up the calculation by 12-fold. The original implementation of GIST by displacement of voxels decreased docking time by 6-fold on average. Here we have incorporated a Gaussian blurring procedure to store the sum of Gaussian-weighted receptor desolvation energies in a grid prior to docking. During docking, trilinear interpolation is utilized to interpolate the receptor desolvation energies at atomic positions, leading to a negligible slowdown compared to the standard scoring function. Finally, DOCK3.7 was rewritten to score each pose of each molecule for both scoring functions, producing two ranked lists for the standard and blurry GIST scoring functions in a single docking run, thus cutting the docking time in half. Second, blurry GIST prioritizes molecules that contain chemotypes that are known to bind AmpC. These include phenolates and carboxylates that coordinate the oxyanion hole of AmpC. Given the penalizing nature of the AmpC receptor desolvation energies, molecules that do not make favorable electrostatic interactions with the protein via a negatively charged moiety are ranked lower, and only those molecules that can form these favorable electrostatic interactions can counteract the penalizing receptor desolvation energies. Reassuringly, this is what we see when manually inspecting the molecules

that rank highly, as well as those that are prioritized in the blurry GIST screen. Molecules that are in high receptor desolvation penalty areas are deprioritized if they do not have a concomitant increase in favorable electrostatic and van der Waals interactions. In standard molecular docking screens, van der Waals energies are unchecked and one may see a bias towards higher molecular weight molecules<sup>50</sup>, while blurry GIST is able to counteract this bias. Third, it seems that blurry GIST can correctly predict binding geometry over the standard scoring function, identifying the correct pose for one of the two crystal structures by less than 1 Å RMSD. Fourth, molecules that are highly ranked in both scoring functions are likely to bind. Though only 2 of the 31 molecules prioritized by blurry GIST did bind, these molecules were mainly taken from far outside the top 10,000 molecules, suggesting that AmpC has very stringent requirements for binders. Molecules must form favorable electrostatic interactions with the oxyanion hole through negatively charged moieties, but they must also form favorable van der Waals with the protein, leaving only those molecules within the highest ranked binders satisfying these criteria. This suggests that though blurry GIST did prioritize molecules that we judged visually to be potential binders, it is only those molecules within the top scoring molecules that have enough of these favorable interactions to bind. This may shape how we think about choosing molecules for purchase from AmpC, but also suggests that different proteins will have different hit rate curves<sup>35</sup>.

It is necessary to consider how the form of GIST may have affected performance. In full GIST, we were calculating full ligand displacement by summing up all voxels contained within the van der Waals radii of the ligand poses. In blurry GIST, we are

applying a Gaussian so that the extremities are weighted less heavily than the center of the atom. It is unclear whether full displacement would have performed better, but rescoring the blurry GIST poses with full displacement GIST and reranking them based on these new GIST energies suggests we would have found similar molecules, and thus similar results (**Figure A.2.6**). There are also different functional forms of inhomogeneous solvation theory<sup>12,16,31,51,52</sup>, and it is unclear which is the most accurate representation of water desolvation. Here, we only include solute-solvent and solvent-solvent enthalpy referenced to bulk solvent, potentially suggesting that entropy, which we have completely neglected and have essentially modeled waters as having no entropy change when they interact with protein relative to bulk solvent, may make a substantial contribution in this site. A possible future direction might be to test a different functional form of GIST that may integrate more successfully into the DOCK3.7 scoring function and see how it performs prospectively.

Our results here also suggest that displacement energies alone may not be able to capture the water energetics in solvent-exposed sites. Previously, we applied GIST to cytochrome *c* peroxidase, a buried model cavity with 6-8 organized water molecules that is only partially exposed to bulk solvent, finding that GIST was able to predict binders correctly, as well as correct geometry. In the solvent-exposed AmpC site with multiple water clusters and water singlets seen in crystal structures, it may be that molecular dynamics simulations and GIST are unable to capture the solvent dynamics and energetics accurately. It has been suggested that more buried sites exhibit more divergent energies because the water energetics deviate more from bulk thermodynamic properties<sup>53</sup>. It is possible that because the AmpC site is not buried and

substantially interacts with bulk solvent, the water energetics here do not deviate enough from bulk thermodynamics to achieve meaningful GIST energies. Additionally, even if meaningful GIST energies are obtained, it may be that they need to be supplemented with reorganization energies to capture the full contribution of water energetics given the substantial contact with bulk solvent.

Our system here, AmpC, is also almost completely penalizing in terms of GIST enthalpies. Cytochrome *c* peroxidase had both favorable and unfavorable water sites<sup>34</sup>, and displacement of unfavorable water sites for boosting ligand affinity has been a large focus in the literature<sup>27,54-56</sup>. It is likely that success when using inhomogeneous solvation theory-based methods is system-dependent and hydration-site-dependent. As we see here, larger molecules are deprioritized because they are penalized more by GIST, and while this can correct the high van der Waals bias in docking, it penalizes high affinity binders as these are the molecules that have enough van der Waals contacts and electrostatic interactions to bind to the AmpC active site and compete with the significant numbers of water molecules that fill the site.

Overall, our results suggest that though blurry GIST may not be able to prioritize molecules that bind, the molecules that did bind in this study are highly ranked in both scoring functions, and thus still captured by blurry GIST. Additionally, we are hopeful that blurry GIST can accurately predict binding geometry compared with the standard scoring function, which will require more crystal structures that we are currently solving.



## 2.5 Methods.

### MD simulation and GIST generation.

Chain B of AmpC  $\beta$ -lactamase (PDB: 1L2S) was processed using tLeAP part of the Amber 14 release. AmpC, as with the other 39 DUD-E systems, were placed in a box of TIP3P water such that all atoms were at least 10 Å from the boundary of the box. PMEMD.cuda was used to carry out simulations on graphics processing units (GeForce; GTX 980). The equilibration run consisted of two minimizations of up to 6,000 steps followed by six 20-ps runs at constant volume where the temperature of the simulation was raised from 0 to 298.15 K. Langevin dynamics maintained the temperature of the simulation with a collision frequency of 2.0 ps<sup>-1</sup>. A constant-pressure (NPT) run was then run to allow the volume of the box to adjust for 5 ns to maintain 1 bar of pressure. Finally, constant-volume (NVT) simulations were performed for 5 ns, under the same conditions as the subsequent production simulations. Production NVT simulations were for 50 ns. All protein heavy atoms were restrained with a 5 kcal/mol/Å<sup>2</sup> force constant and the Shake algorithm was used with a 2-fs time step. Periodic boundary conditions were applied, and the particle mesh Ewald method was used to calculate long-range electrostatics.

**GIST grids.** GIST grids were generated using the CPPTRAJ trajectory analysis program from AmberTools 14 by processing the 50-ns trajectories with a grid spacing of 0.5 Å. The grids were combined with Python scripts that are available at [https://github.com/tbalius/GIST\\_DX\\_tools](https://github.com/tbalius/GIST_DX_tools). As previously, the receptor desolvation is

estimated using GIST grids that are outputted by the CPPTRAJ trajectory analysis program. These are:

- Enthalpy between solvent (water) and solute (receptor) ( $E_{s,w}^{dens}$ )
- Enthalpy of solvent with solvent ( $E_{w,w}^{dens}$ )
- Translational entropy between water and receptor ( $TS_{s,w}^{trans}$ )
- Orientational entropy between water and receptor ( $TS_{s,w}^{orient}$ )
- Density of water around the receptor ( $g_o$ )

All grids' energies are in kcal/mol/Å<sup>3</sup>, while the density grid is unitless (density/bulk density). We found previously that the enthalpy grids ( $E_{s,w}^{dens}$ ) and ( $E_{w,w}^{dens}$ ) referenced to bulk solvent performed the best in terms of enrichment. To estimate the enthalpy difference of desolvation, we subtract the energy of water in bulk from the energy of water on the surface of the protein. For each voxel,  $i$ , the bulk solvent energy was computed as:

$$E_{w,w}^{dens\_ref}(i) = 2 \times (E_{w,w}^{dens}(i) + 0.3184 \times g_o(i))$$

Here, the constant is computed from parameters taken from the Amber14 manual, the mean energy of TIP3P solvent model,  $C_{bulk} = -9.533$  kcal/mol/water, and the number density of the TIP3P solvent model,  $C_{num\_dens} = 0.0334$  waters/Å<sup>3</sup>, where  $C_{bulk} \times C_{num\_dens} = -0.3184$  kcal/mol/Å<sup>3</sup>. The factor of two accounts for the fact that each water interacts with every other water during the simulation, but only retains half of the interaction energy to avoid double counting. Thus, by multiplying by two, we recover the full water-water interaction energy. The GIST enthalpy stored at each voxel then becomes:

$$E_{tot}^{ref2}(i) = E_{s,w}^{dens}(i) + E_{w,w}^{dens.ref}(i)$$

For the other solvent models used (TIP4PEw, TIP5P, SPCE, OPC), the formulation remains the same, but the  $C_{bulk}$  and  $C_{num\_dens}$  values change to reflect their specific values in the Amber14 manual. As previously, we truncated the GIST energies at the absolute magnitude of 3 kcal/mol/Å<sup>3</sup> as these high magnitude voxels typically diminished enrichment performance.

### Blurry GIST grids.

To speed up our DOCK calculations, we need a way to precompute displacement without double counting. In Blurry GIST scoring, we weight the grid points closer to the center of the atom higher than those points near the surface. To this end we use a Gaussian function as follows:

$$gw(d) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-d^2}{2\sigma^2}}$$

Here  $d$  a distance,  $\pi$  is the mathematical constant the quotient of circumference to diameter and  $\sigma$  is the sharpness of the peak of the function (this is the standard deviation for the normal distribution).

$$bG_{full} = \sum_{a \in Lig} \sum_{p \in gd(a)} gw(dist(p, a)) * GIST(p)$$

The blurry gist score is a double summation: we sum over all atoms in the ligand. And we have a weighted sum over the grid points displaced ( $gd(a)$ ) by each atom ( $a$ ). The weight is determined by the proximity of grid point ( $p$ ) to the center of the atom ( $a$ ) using the Gaussian function. The displacement function is dependent on the radius or the

atom. We experimented with various  $\sigma$  values, finding that the radius / 1.3 provides the best agreement with the full GIST displacement energies. By the blurry GIST definition, we do not need to worry about double counting and, so can pre-compute displacements. These GIST displacements were precalculated by placing a dummy atom with a specified radius at each grid point (a margin of grid points is excluded because the sphere goes outside the grid box). Grid points that are contained within the radius of the dummy atom are identified and are summed with a weighting factor assigned to each point based on its distance from the center of the dummy atom using the Gaussian function. The new summed value is stored at that grid point. Two of these pre-computed displacement grids are generated, one with a radius of 1.8 Å for heavy atoms and 1.0 Å for hydrogens.

$$bG_{trilinear} = \sum_{a \in Lig} trilinear(a, GIST_{precompute})$$

Blurry GIST in docking. The blurry GIST scoring function method was implemented into the DOCK 3.7 distribution. This method is much faster GIST calculation than that previously described<sup>34</sup>. In the implementation, the two precomputed blurry gist grids (heavy and hydrogen) are read into DOCK. Trilinear interpolation is used to combine the information of the 8 closest grid point to approximate the value at the center of the ligand atom. With this method there is virtually no slowdown in the calculations when compared to running DOCK3.7 without GIST.

## Monte Carlo Optimization

A Monte Carlo optimization method was implemented into the DOCK source code based on the example Python code available at:

[https://chryswoods.com/intro\\_to\\_mc/part1/metropolis.html](https://chryswoods.com/intro_to_mc/part1/metropolis.html).

In this new DOCK scheme, a translational or rotational move was randomly selected, and either a translation up to 0.2 Å or rotation up to 10 Å was applied to the best scoring pose of each molecule. The new pose was rescored for electrostatics, van der Waals, ligand desolvation, and if applicable, blurry GIST. The new pose was accepted if its energy was better than the previous pose's energy. However, if the energy was worse, the exponential of the difference between the new energy and old energy divided by the thermal energy (kT) was calculated. If this value was greater than or equal to a random value generated between 0 and 1, then the pose was accepted, and a new rotational or translational move was generated for 1000 Monte Carlo steps. The temperature was set to 1 K to limit the poses from moving too far from the DOCK-generated pose. An alternative scheme was also implemented that terminated Monte Carlo optimization if 500 Monte Carlo steps were accepted.

## Combinatorial Scoring

This scheme was implemented into the DOCK source code. Since sampling is identical between the standard and blurry GIST scoring functions when the same grids and matching spheres are used, we implemented a new scoring scheme that only performs sampling once to reduce redundant calculations. First, poses are generated, and then are scored with the blurry GIST scoring function, which includes electrostatics,

van der Waals, and ligand desolvation, which comprises the standard scoring function terms. Thus, to obtain the standard score of each of these poses, one only needs to subtract the blurry GIST energy from the total score. The best scoring poses from both standard and blurry GIST scoring functions are then minimized with their own scoring function terms using Simplex minimization. The minimized poses are then rescored with the other scoring function, and the energies of the poses are compared. If a better scoring pose for each molecule was found in the other scoring function after minimization, that pose replaced the current best scoring pose for that molecule. This ensures that the best scoring pose for both scoring functions was found, regardless of whether it was generated from standard or blurry GIST Simplex minimization (see **Figure A.2.3**)

### **Enrichment Calculations.**

Three dimensional dockable ligand and decoy files for the 40 DUD-E targets were downloaded from <http://autodude.docking.org>. The PDB structures assigned to forty DUD-E targets were retrieved and prepared in an automated fashion by in-house scripts based on the DOCK blaster pipeline<sup>65</sup> for generating docking grids. For all systems besides AmpC, the default DOCK blaster preparation was used in which the full binding site was filled with low-dielectric spheres of radius 1.9 Å for Poisson-Boltzmann calculations, thereby modeling the full binding site as low dielectric solute. The DUD-E assigned PDB ligand was used for generating 45 matching spheres, to which molecules are matched to during docking. Docking calculations were performed with DOCK3.7<sup>66</sup>. Ligand conformations were generated by OpenEye's Omega<sup>67</sup>.

Ligands were only scored if the number of ligand heavy atoms contained within the ligand ranged from 4 to 100. For each ligand hierarchy (each rigid fragment contained within the ligand), the maximum number of matches generated was set to 5000. Up to 500 Simplex minimization<sup>37</sup> steps were performed for each top scoring pose of each docked molecule, starting with initial translations of 0.2 Å and initial rotations of 5°. To judge performance, the adjusted log AUC was used, which is analogous to the area under the receiver operator characteristic curve<sup>68</sup>. The adjusted log AUC subtracts the log AUC of the random curve (14.462%) to ensure random enrichment is 0%.

### **Virtual Screen.**

Chain B of AmpC PDB code 1L2S, was used in the docking calculations. To prepare the structure for docking, atoms of the co-crystallized ligand, were used to seed the matching sphere calculation in the active site; these spheres represent favorable positions for individual ligand atoms to dock; 45 spheres were used in total. DOCK3.7 orients flexibases of pre-calculated ligand conformations into the orthosteric site by overlaying atoms of each library molecule onto these matching spheres. The receptor structure was protonated by REDUCE<sup>57</sup> and assigned AMBER united atom charges<sup>58</sup>. The magnitudes of the partial atomic charges of the residues Ser64, Ala318, and Asn152 were increased without changing the net charge of the residues, as described previously<sup>35,41,42</sup>. The volume of the low protein dielectric, which defines the boundary between solute and solvent in Poisson–Boltzmann electrostatic calculations, was extended out 1.5 Å from the protein surface using spheres calculated by SPHGEN. Scoring grids were pre-calculated using CHEMGRID for AMBER van der Waals

potential, QNIFFT for Poisson–Boltzmann-based electrostatic potentials, and SOLVMAP for ligand desolvation.

The resulting potential grids and ligand-matching parameters were evaluated for their ability to enrich known AmpC ligands over property-matched decoys. Decoys share the same physical properties as known ligands but are topologically dissimilar and are therefore unlikely to bind to AmpC. The ligands and decoys were taken from the Directory of Useful Decoys - Enhanced<sup>36</sup> benchmark, which contains 48 AmpC ligands and 2,850 property-matched decoys. Docking success was judged based on the ability to enrich the known ligands over the decoys by docking rank, using adjusted logAUC values, as is widely done in the field. We also ensured that molecules with extreme physical properties were not enriched, as can happen when only counter-screening against property-matched decoys. In particular, we wanted to ensure that anionic molecules were enriched over neutral and cationic molecules. The docking parameters were also judged on how well they reproduced the expected binding modes of the known ligands. In addition to these criteria, docking parameters that had the largest impact in terms of rank changes of molecules upon addition of blurry GIST were prioritized.

The ‘lead-like’ subset of ZINC15 (<http://zinc15.docking.org>), characterized by favourable physical properties (for example, with calculated octanol-water partition coefficients (cLopP)  $\leq 3.5$  and with molecular mass  $\leq 400$  Da), was then docked against the AmpC active site using DOCK3.7. This library contained more than 300 million molecules, most of which were make-on-demand compounds from the Enamine REAL set. Of these, more than 271 million molecules successfully docked. An average of



4,082 orientations was calculated for each, and for each orientation, an average of 563 conformations was sampled. A simplex minimizer was used for rigid-body minimization on the best-scored pose for each ligand. Overall, about 198 trillion complexes were sampled and scored. The calculation time was 161,230 core hours, or 4.49 calendar days on 1,500 cores.

The ranks for molecules in the top 1% (2.7 million molecules) for both scoring functions were retrieved, and any molecule that had a half-log order rank change (3.16-fold) was retained. This included 154,256 molecules that ranked more highly in the blurry GIST scoring function hit list and lower in the standard scoring function (pro-bGIST), and 159,071 molecules that ranked more highly in the standard scoring function hit list and lower in the blurry GIST scoring function (anti-GIST). Molecules that were identical by ECFP4-based Tanimoto coefficients to the known >200 AmpC inhibitors or molecules previously tested were removed. To identify molecules whose geometries changed between the two scoring functions, the union of the top 10,000 molecules from both the standard and blurry GIST scoring functions was collected. The root mean squared deviation (RMSD) using the Hungarian algorithm in DOCK6.6<sup>59,60</sup> was calculated on the standard and blurry GIST poses, and any molecule that had a substantial RMSD change were retained.

After filtering, the pro-bGIST, anti-bGIST, and pose changing docked poses of these molecules were filtered by the proximity of their anionic charges, if any, to the oxyanion hole in AmpC, which is coordinated by the backbone atoms of residues Ser64 and Ala318. Molecules were manually inspected for favorable geometry and

interactions. In the first round, 36 molecules were purchased, 30 of which were successfully synthesized.

In the second round of docking, we wanted to buy molecules that had a higher likelihood of inhibiting AmpC to attain a better comparison of prioritized versus deprioritized blurry GIST hit rates. Therefore, 60,216 molecules with the SMARTS patterns:

- [ND2]S(=O)(=O)c1ccsc1C(=O)[OD1] for carboxylates
- [ND2]S(=O)(=O)c1cccc([F,Cl,I,Br])c1[OD1] for phenolates

were retrieved from ZINC15 and built using the lab's ligand building pipeline. These new molecules were docked to the AmpC docking grids with the same parameters as above, except the match goal sampling value was increased to 5000 from the 1000 used for the large-scale screen. The energies of the best scoring poses from these molecules were extracted and incorporated into the original standard and blurry GIST docking hit lists. With these new docking hit lists, the poses of molecules that were prioritized or deprioritized with a half-log order rank change cut-off were collected. Molecules that were identical by ECP4 Tanimoto coefficients to known AmpC inhibitors or molecules previously bought were discarded. As before, the poses were manually inspected for favorable geometry and interactions, and 19 molecules prioritized by blurry GIST and 18 molecules deprioritized by blurry GIST were chosen for synthesis and testing.

## Ligand-Bound Molecular Dynamics and GIST

Ligand forcefield parameters were assigned with the general AMBER force field (GAFF)<sup>61</sup> using the Antechamber package in AmberTools. Antechamber assigns charges, missing bonds, angles and dihedral angles. Ligand charges were assigned using AM1-BCC<sup>62</sup>. The produced files are loaded into tleap to produce the ligand lib file of the ligand. Each system was solvated in a box TIP3P water molecules with Amber ff14SB forcefield for the protein-ligand structure. The box was created such that there is 10 Å between any atom of the protein and the edge of the box. The solvated system was minimized with an initial 1500 steps of steepest descent with all atoms except hydrogens restrained harmonically using a force constant of 100 kcal/mol Å<sup>2</sup> followed by another 1500 steps of steepest descent with all atoms except hydrogens restrained harmonically using a force constant of 5 kcal/mol Å<sup>2</sup>. This was followed by heating the system from 50 K to 298.15 K over 120 picoseconds under the conditions of constant number of particles, volume, and temperature (NVT) with all atoms restrained except hydrogens with a force constant of 5 kcal/mol Å<sup>2</sup>. An equilibration simulation was then run in constant NPT conditions for 5 ns with temperature of 298.15 K, pressure of 1 atmosphere and same atom restraints as described in the NVT equilibration step. Temperature was regulated using the Langevin thermostat with a collision frequency of 2.0 ps<sup>-1</sup>, and pressure was regulated using the Berendsen barostat<sup>44</sup> with isotropic scaling and a coupling constant of 2.0 ps<sup>-1</sup>. The snapshots of system coordinates were saved every 1 picosecond, resulting in a trajectory file with 150,000 frames. All MD simulations were performed using AMBER 18. In the production runs, all atoms except hydrogens were restrained with a force constant of 5 kcal/mol Å<sup>2</sup>. GIST maps were

produced in a 37x35x45 Å rectangular box using the GPU version of CPPTRAJ\_GIST<sup>63,64</sup> on the entire 150ns of the production run.

### **AmpC Enzymology.**

All potential inhibitors were initially dissolved in DMSO at 30mM, and more dilute stocks were prepared if necessary, maintaining the DMSO concentration at 1% v/v in 50mM sodium cacodylate buffer at pH 6.5. AmpC activity and inhibition was monitored spectrophotometrically using CENTA as a substrate<sup>38</sup>. All assays included 0.01% Triton X-100 to reduce aggregation artifacts. Active compounds were investigated more fully by IC<sub>50</sub> curves, which reflect the percentage inhibition fit to a dose-response equation in GraphPad Prism. For these compounds, K<sub>i</sub> values were calculated from Lineweaver-Burk plots.

### **AmpC Crystallography.**

The two inhibitors were cocrystallized from 1.7 M potassium phosphate with microseeding at pH values that varied from 8.7 and 8.9, as previously described<sup>35,38</sup>. Crystals were cryo-cooled in a solution that contained a reservoir solution and 25% sucrose. Reflections were measured at beamline 8.3.1 of the Advanced Light Source with a wavelength of 1.11583 nm at a temperature of 100 K. Complexes with ZINC339208618 and ZINC37748240 were measured to a resolution of 1.7 Å and 1.75 Å, respectively.

## References

1. Southall, N. T., Dill, K. A. & Haymet, A. D. J. A View of the Hydrophobic Effect. *The Journal of Physical Chemistry B* **106**, 521-533, doi:10.1021/jp015514e (2002).
2. Biela, A. *et al.* Dissecting the hydrophobic effect on the molecular level: the role of water, enthalpy, and entropy in ligand binding to thermolysin. *Angew Chem Int Ed Engl* **52**, 1822-1828, doi:10.1002/anie.201208561 (2013).
3. Snyder, P. W., Lockett, M. R., Moustakas, D. T. & Whitesides, G. M. Is it the shape of the cavity, or the shape of the water in the cavity? *The European Physical Journal Special Topics* **223**, 853-891, doi:10.1140/epjst/e2013-01818-y (2014).
4. Krimmer, S. G. *et al.* Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands. *J Med Chem* **59**, 10530-10548, doi:10.1021/acs.jmedchem.6b00998 (2016).
5. Baron, R., Setny, P. & McCammon, J. A. Water in cavity-ligand recognition. *J Am Chem Soc* **132**, 12091-12097, doi:10.1021/ja1050082 (2010).
6. Venkatakrisnan, A. J. *et al.* Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proc Natl Acad Sci U S A* **116**, 3288-3293, doi:10.1073/pnas.1809251116 (2019).
7. Hummer, G. Molecular binding: Under water's influence. *Nat Chem* **2**, 906-907, doi:10.1038/nchem.885 (2010).
8. Biela, A., Betz, M., Heine, A. & Klebe, G. Water makes the difference: rearrangement of water solvation layer triggers non-additivity of functional group contributions in protein-ligand binding. *ChemMedChem* **7**, 1423-1434, doi:10.1002/cmdc.201200206 (2012).

9. Breiten, B. *et al.* Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* **135**, 15579-15584, doi:10.1021/ja4075776 (2013).
10. Fox, J. M. *et al.* Water-Restructuring Mutations Can Reverse the Thermodynamic Signature of Ligand Binding to Human Carbonic Anhydrase. *Angew Chem Int Ed Engl* **56**, 3833-3837, doi:10.1002/anie.201609409 (2017).
11. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* **130**, 2817-2831, doi:10.1021/ja0771033 (2008).
12. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci U S A* **104**, 808-813, doi:10.1073/pnas.0610202104 (2007).
13. Li, Z. & Lazaridis, T. Computing the thermodynamic contributions of interfacial water. *Methods Mol Biol* **819**, 393-404, doi:10.1007/978-1-61779-465-0\_24 (2012).
14. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B* **102**, 3531-3541, doi:10.1021/jp9723574 (1998).
15. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* **137**, 044101, doi:10.1063/1.4733951 (2012).

16. Nguyen, C. N., Cruz, A., Gilson, M. K. & Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J Chem Theory Comput* **10**, 2769-2780, doi:10.1021/ct401110x (2014).
17. Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J Am Chem Soc* **131**, 15403-15411, doi:10.1021/ja906058w (2009).
18. Cui, G., Swails, J. M. & Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J Chem Theory Comput* **9**, 5539-5549, doi:10.1021/ct400711g (2013).
19. Bayden, A. S., Moustakas, D. T., Joseph-McCarthy, D. & Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J Chem Inf Model* **55**, 1552-1565, doi:10.1021/ci500746d (2015).
20. Pearlstein, R. A., Sherman, W. & Abel, R. Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38alpha MAP kinase inhibitors. *Proteins* **81**, 1509-1526, doi:10.1002/prot.24276 (2013).
21. Li, Z. & Lazaridis, T. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J Am Chem Soc* **125**, 6636-6637, doi:10.1021/ja0299203 (2003).
22. Bodnarchuk, M. S., Viner, R., Michel, J. & Essex, J. W. Strategies to calculate water binding free energies in protein-ligand complexes. *J Chem Inf Model* **54**, 1623-1633, doi:10.1021/ci400674k (2014).

23. Huggins, D. J., Marsh, M. & Payne, M. C. Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface. *J Chem Theory Comput* **7**, 3514-3522, doi:10.1021/ct200465z (2011).
24. Guimaraes, C. R. & Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J Chem Inf Model* **50**, 547-559, doi:10.1021/ci900497d (2010).
25. Kohlmann, A., Zhu, X. & Dalgarno, D. Application of MM-GB/SA and WaterMap to SRC Kinase Inhibitor Potency Prediction. *ACS Med Chem Lett* **3**, 94-99, doi:10.1021/ml200222u (2012).
26. Luccarelli, J., Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38alpha MAP Kinase Inhibitors. *J Chem Theory Comput* **6**, 3850-3856, doi:10.1021/ct100504h (2010).
27. Robinson, D. D., Sherman, W. & Farid, R. Understanding kinase selectivity through energetic analysis of binding site waters. *ChemMedChem* **5**, 618-627, doi:10.1002/cmdc.200900501 (2010).
28. Bucher, D., Stouten, P. & Triballeau, N. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J Chem Inf Model* **58**, 692-699, doi:10.1021/acs.jcim.7b00642 (2018).
29. Snyder, P. W. *et al.* Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci U S A* **108**, 17889-17894, doi:10.1073/pnas.1114107108 (2011).



30. Robinson, D. *et al.* Differential Water Thermodynamics Determine PI3K-Beta/Delta Selectivity for Solvent-Exposed Ligand Modifications. *J Chem Inf Model* **56**, 886-894, doi:10.1021/acs.jcim.5b00641 (2016).
31. Murphy, R. B. *et al.* WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J Med Chem* **59**, 4364-4384, doi:10.1021/acs.jmedchem.6b00131 (2016).
32. Sun, H., Zhao, L., Peng, S. & Huang, N. Incorporating replacement free energy of binding-site waters in molecular docking. *Proteins* **82**, 1765-1776, doi:10.1002/prot.24530 (2014).
33. Uehara, S. & Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **21**, doi:10.3390/molecules21111604 (2016).
34. Balias, T. E. *et al.* Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc Natl Acad Sci U S A* **114**, E6839-E6846, doi:10.1073/pnas.1703287114 (2017).
35. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229, doi:10.1038/s41586-019-0917-9 (2019).
36. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582-6594, doi:10.1021/jm300687e (2012).
37. Gschwend, D. A. & Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* **10**, 123-132, doi:10.1007/bf00402820 (1996).

38. Eidam, O. *et al.* Design, synthesis, crystal structures, and antimicrobial activity of sulfonamide boronic acids as beta-lactamase inhibitors. *J Med Chem* **53**, 7852-7863, doi:10.1021/jm101015z (2010).
39. Eidam, O. *et al.* Fragment-guided design of subnanomolar beta-lactamase inhibitors active in vivo. *Proc Natl Acad Sci U S A* **109**, 17448-17453, doi:10.1073/pnas.1208337109 (2012).
40. London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol* **10**, 1066-1072, doi:10.1038/nchembio.1666 (2014).
41. Powers, R. A., Morandi, F. & Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure* **10**, 1013-1023, doi:10.1016/s0969-2126(02)00799-2 (2002).
42. Barelier, S. *et al.* Increasing chemical space coverage by combining empirical and computational fragment screens. *ACS Chem Biol* **9**, 1528-1535, doi:10.1021/cb5001636 (2014).
43. Babaoglu, K. *et al.* Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J Med Chem* **51**, 2502-2511, doi:10.1021/jm701500e (2008).
44. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. **21**, 1087-1092, doi:10.1063/1.1699114 (1953).
45. Teotico, D. G. *et al.* Docking for fragment inhibitors of AmpC beta-lactamase. *Proc Natl Acad Sci U S A* **106**, 7455-7460, doi:10.1073/pnas.0813029106 (2009).

46. Sterling, T. & Irwin, J. J. ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model* **55**, 2324-2337, doi:10.1021/acs.jcim.5b00559 (2015).
47. Geschwindner, S. & Ulander, J. The current impact of water thermodynamics for small-molecule drug discovery. *Expert Opin Drug Discov* **14**, 1221-1225, doi:10.1080/17460441.2019.1664468 (2019).
48. Biela, A. *et al.* Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. *J Med Chem* **55**, 6094-6110, doi:10.1021/jm300337q (2012).
49. Krimmer, S. G., Betz, M., Heine, A. & Klebe, G. Methyl, ethyl, propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem* **9**, 833-846, doi:10.1002/cmdc.201400013 (2014).
50. Verdonk, M. L. *et al.* Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci* **44**, 793-806, doi:10.1021/ci034289q (2004).
51. Wahl, J. & Smiesko, M. Thermodynamic Insight into the Effects of Water Displacement and Rearrangement upon Ligand Modifications using Molecular Dynamics Simulations. *ChemMedChem* **13**, 1325-1335, doi:10.1002/cmdc.201800093 (2018).
52. Hufner-Wulsdorf, T. & Klebe, G. Protein-Ligand Complex Solvation Thermodynamics: Development, Parameterization, and Testing of GIST-Based Solvent Functionals. *J Chem Inf Model* **60**, 1409-1423, doi:10.1021/acs.jcim.9b01109 (2020).

53. Beuming, T. *et al.* Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* **80**, 871-883, doi:10.1002/prot.23244 (2012).
54. Beuming, T., Farid, R. & Sherman, W. High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci* **18**, 1609-1619, doi:10.1002/pro.177 (2009).
55. Laha, J. K. *et al.* Structure-activity relationship study of 2,4-diaminothiazoles as Cdk5/p25 kinase inhibitors. *Bioorg Med Chem Lett* **21**, 2098-2101, doi:10.1016/j.bmcl.2011.01.140 (2011).
56. Knegt, R. M. & Robinson, D. D. A Role for Hydration in Interleukin-2 Inducible T Cell Kinase (Itk) Selectivity. *Mol Inform* **30**, 950-959, doi:10.1002/minf.201100086 (2011).
57. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735-1747, doi:10.1006/jmbi.1998.2401 (1999).
58. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106**, 765-784, doi:10.1021/ja00315a051 (1984).
59. Brozell, S. R. *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* **26**, 749-773, doi:10.1007/s10822-012-9565-y (2012).

60. 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* **54**, 518-529, doi:10.1021/ci400534h (2014).
61. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).
62. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623-1641, doi:10.1002/jcc.10128 (2002).
63. Ramsey, S. *et al.* Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J Comput Chem* **37**, 2029-2037, doi:10.1002/jcc.24417 (2016).
64. Kraml, J., Kamenik, A. S., Waibl, F., Schauerl, M. & Liedl, K. R. Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces. *J Chem Theory Comput* **15**, 5872-5882, doi:10.1021/acs.jctc.9b00742 (2019).
65. Irwin, J. J. *et al.* Automated docking screens: a feasibility study. *J Med Chem* **52**, 5712-5720, doi:10.1021/jm9006966 (2009).
66. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992, doi:10.1371/journal.pone.0075992 (2013).
67. Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality

structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **50**, 572-584, doi:10.1021/ci100031x (2010).

68. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).

## Gloss to Chapter 3

This chapter came out of the work on blurry GIST and my efforts to incorporate it more readily into the DOCK scoring function. Since increasing the weight of GIST or blurry GIST decreased enrichment performance on average and I thought that the ligand desolvation term in its current weighting may be entangled with receptor desolvation as highlighted in the Introduction, I thought another approach might to weight the three other scoring function terms differently. If we could find a similar or better performing set of weights compared to the standard scoring function, we might be able to incorporate blurry GIST more readily so that it would be more impactful, while preserving its performance.

This involved altering the weights of the scoring function terms so that there were sixteen different weighting combinations, and then running each of these combinations on 41 DUD-E systems to determine enrichment performance. What we found was that down-weighting the ligand desolvation by half while keeping the other terms the same provided the best improvement in average enrichment over the standard scoring function. However, upon closer inspection, we found that one reason for this was because of charge mismatches in the molecule database files, with ligands having more extreme charges, thus artificially improving in rank over the less extreme charged decoy molecules. This was most notable in AmpC and PUR2. This inspired me to write a new pipeline for generating decoys from input ligands, in a similar fashion as Michael Mysinger's decoy generation pipeline, but one that was compatible with ZINC15. In my pipeline, one could input their own property ranges for decoys to share with their

ligands, even finding decoys that were identical in charge, which I thought at the time, might be less susceptible to scoring function weighting changes. This new pipeline and its resulting decoys were named “DUDE-Z”. The decoys themselves are not important, but rather the ability to customize decoys for ligands readily.

Additionally, the inability to differentiate good performance with just log AUC values provided motivation for a new set of control tests before virtual screening so that we would not be deceived in future screens. We name these control tests “Goldilocks”, “Extrema”, and “bootstrapping”. In Goldilocks, a small subset of ZINC15 with a wide array of properties was retrieved that matches the database at large in terms of charge, molecular weight, and clogP. When screening this subset to one’s protein preparation, the goal is to identify any flaws in parameterization before a large-scale screen that may be hidden when only screening against a set of ligands and property-matched decoys. In Extrema, one generates a set of molecules in a specific molecular weight and cLogP range, typically that of your ligands, with an equivalent number of -2, -1, 0, +1, and +2 charged molecules. The goal of this test is to identify which charges rank the best against one’s protein preparation. This can motivate one to continue to the screen with their current setup or conclude that one’s preparation is incorrect and is simply prioritizing charge instead of differentiating between ligands and decoys. In bootstrapping, one generates 50 different enrichments by choosing ligands and decoys at random from one’s docking hit list with replacement, re-ranking them, and re-calculating the log AUC. This provides one with a measure of how much variation there is in their enrichment performance, as well as to compare multiple setups against one another to determine if one performs better.



These new control tests were applied to 41 DUD-E systems and the D4 dopamine and MT1 melatonin receptors to identify parameterization errors and liabilities in each system's setups, and to compare the standard scoring function to the reweighted scoring function with down-weighted ligand desolvation. The data show that down-weighting the ligand desolvation term results in charge priority issues, and although this optimized scoring function performs better in enrichment, the differences in enrichment become insignificant when using bootstrapping. The ease of use of these new tools has made them accessible and heavily used by all members of the lab before running large virtual screens.

## **Chapter 3: Property-unmatched decoys in docking benchmarks**

Reed M. Stein<sup>1</sup>, Ying Yang<sup>1</sup>, Trent E. Balius<sup>2</sup>, Matt J. O'Meara<sup>3</sup>, Jiankun Lyu<sup>1</sup>, Jennifer Young<sup>1</sup>, Khanh Tang<sup>1</sup>, Brian K. Shoichet<sup>1\*</sup> & John J. Irwin<sup>1\*</sup>

1. Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158

2. Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc. PO Box B, Frederick, MD 21702

3. Department of Computational Medicine and Bioinformatics, University of Michigan

\*Corresponding authors: Brian Shoichet: [bshoichet@gmail.com](mailto:bshoichet@gmail.com); John Irwin:

[jir322@gmail.com](mailto:jir322@gmail.com)

### 3.1 Abstract

Enrichment of ligands versus property-matched decoys is widely used to test and optimize docking library screens. However, the unconstrained optimization of enrichment alone can mislead, leading to false confidence in prospective performance. This can arise by over-optimizing for enrichment against property-matched decoys, without considering the full spectrum of molecules to be found in a true large library screen. Adding decoys representing charge extrema helps mitigate over-optimizing for electrostatic interactions. Adding decoys that represent the overall characteristics of the library-to-be-docked allows one to sample molecules not represented by ligands and property-matched decoys, but that one will encounter in a prospective screen. We also explore the variability one can encounter in enrichment calculations, and how that can temper one's confidence in small enrichment differences. All such controls are ultimately sanity checks, and the investigator must remain vigilant to avoid being misled by artifacts. The new tools are freely available at <http://tldr.docking.org>.

## 3.2 Introduction

In large library docking screens, the goal is to discover new, typically unprecedented chemotypes for a target based on molecular fit. Calculation speed has been crucial since the field's inception<sup>1-8</sup>, and to ensure it several biophysical terms are approximated or ignored entirely. While this led to programs that can screen libraries now approaching<sup>9</sup> or exceeding<sup>10</sup> a billion molecules, discovering novel ligands for multiple targets<sup>9,11-21</sup>, the emphasis on throughput has forced docking into compromises that make predicting absolute binding energies, or even rank ordering compounds, implausible. While molecular docking screens are thus pragmatic, and while docking remains among the methods most subjected to experimental testing in computational biophysics, it is also among the biophysical methods that have most surrendered "ground truth".

Accordingly, to evaluate new docking methods, or to evaluate how well docking might perform prospectively on a new target, benchmarking studies are often performed. For a new docking method, these benchmarks evaluate the key outcomes expected of a library screen: can the method reproduce the binding orientations of known ligands for a range of targets, can it enrich known ligands from among a set of decoys over a range of disparate targets? For a particular target campaign, when an established method is being used, such benchmarks are also crucial, here focusing on the poses and the enrichment of ligands, and when available known non-binders, for that target. Optimizing sampling and the weighting of energy terms—ideally constrained by physical reasonableness—can improve performance of these retrospective controls.

Admittedly, favorable performance on retrospective benchmarks does not predict prospective success in predicting new chemotypes—the goal of library docking—but without them the likelihood of success is reduced, as is our ability to understand failure. In this sense, running detailed benchmarks on a new method or a new target fulfills the same role as controls in the experimental biological sciences, which often also lack “ground truth”, and so must rigorously control all new experiments.

Among the most widely-used benchmarks in library docking is the enrichment of annotated ligands versus property-matched “decoy” molecules<sup>22-24</sup>. A decoy molecule is one that is expected not to bind to a protein of interest; enrichment measures docking’s ability to highly rank (enrich) the annotated ligands vs. such decoys. The idea of using decoys in benchmarks follows from analogous use in protein structure-prediction<sup>25-27</sup>, and initially drew on random molecules<sup>28-30</sup>. As is true for folding decoys, it was found that it was important that decoy molecules physically resemble the known ligands, otherwise the docking program might be optimized to simply recognize gross physical differences, such as molecular weight, hydrophobicity, or charge<sup>31</sup>. Property-matched decoys match ligands by physical properties but are otherwise topologically unrelated and so presumed not to bind. Enrichment of ligands against property-matched decoys, in sensible geometries, thus offers some assurance that the docking program is recognizing ligands by their detailed interactions, and not just gross physical differences. Several benchmarking sets of ligands and property-matched decoys have been introduced<sup>32-37</sup>, including the DUD and DUD-E sets<sup>22,23</sup>, which are widely used to test new methods, while the method of matching ligands to decoys, on

which DUD-E is based, is widely called upon to construct bespoke benchmarks as controls for individual target campaigns.

The DUD-E benchmark covers 102 disparate proteins, 66,695 ligands, and 1.4 million property-matched decoys (about 50 decoys per ligand for any given target—some ligands share decoys with those from another target). Notwithstanding the importance of using property-matched decoys<sup>31</sup>, benchmarking on them alone exposes one to subtle but crucial biases, which can mislead optimization of both methods and parameterization for a particular campaign. A key challenge is that property-matched decoys do not represent the full spectrum of molecules that will be encountered in docking a diverse library of  $10^9$  molecules. For instance, they will not expose one to extreme physical differences, nor will they necessarily represent even the typical molecular properties of a large library. Here we investigate the pathologies that can emerge from optimizing from even a large and diverse set of targets, ligands and property-matched decoys, and investigate additional properties that can control for these pathologies, providing a fuller set of benchmarks to complement property-matched decoy sets, which do remain crucial.

### 3.3 Methods

**DUD-E.** Three dimensional dockable ligand and decoy files for the 41 DUD-E targets were downloaded from <http://autodude.docking.org>. For D4 dopamine and melatonin MT1 receptors, DUD-E decoys were generated from <http://dude.docking.org/generate>, and built using an in-house ligand building pipeline.

**Binders & Nonbinders.** Three dimensional dockable files for binders and nonbinders for D4 dopamine and MT1 melatonin receptors were downloaded from ZINC15. This included 81 binders and 468 nonbinders, and 105 binders and 65 nonbinders for D4 and MT1, respectively. Enrichment calculations were performed for all 16 scoring function coefficient combinations (see Docking Calculations).

**DUDE-Z.** Several DUD-E systems had large numbers of ligands and decoys, so to reduce the number of ligands for more rapid docking calculations, targets that had over 100 ligands had their ligands sorted by molecular weight and were clustered by an ECFP4 Tanimoto coefficient of 0.7. Ligands were sorted by molecular weight as 3D molecules are more likely to be in lead-like space, and so to ensure that ligands could find 3D property-matched decoys, the smallest ligand was chosen in each cluster. For targets that had less than 100 ligands, all ligands were retained for generating property-matched decoys.

As in DUD-E, decoys were matched to ligands based on molecular weight, water-octanol partition coefficient (cLogP), number of rotatable bonds, number of hydrogen

bond donors and acceptors, and net charge. We generated all protonation states for each ligand using ChemAxon's Jchem<sup>38</sup> at physiological pH and computed molecular properties using RDKit. For each protomer, the optimal goal was to find 50 property-matched decoys, but we also accepted as few as 20 decoys if the number of decoys in ZINC15 were limited in this property space. To identify matching decoys, the ZINC15 website was cURLed for up to 10,000 3D molecules matching the ligand protomer for the molecular properties listed above. Once thousands of decoys for a target were retrieved, ECFP4 Tanimoto calculations were performed using in-house programs (located at <https://github.com/docking-org/ChemInfTools>) between all ligands and all potential decoys for that target. Any decoy that had greater than 0.35 ECFP4 Tanimoto coefficient to any ligand was discarded. Next, the decoys were sorted by molecular weight, and decoys were clustered by an ECFP4 Tanimoto coefficient of 0.8, with the smallest decoy being retained from each cluster. This ensured that property-matched decoys would not contain duplicates, and that the decoys would contain relatively different scaffolds. The remaining decoys were sorted by ECFP4 Tanimoto coefficients to all ligands and were assigned such that the ligand with the least number of decoys assigned would be assigned the decoy in an iterative procedure. If more than 50 decoys could be assigned to all ligands, the remaining decoys were kept as replacements. If fewer than 50 decoys could be assigned to all ligands, the highest number of decoys that could be assigned to the ligand protomers was computed. If it was difficult to find 3D decoys for a target, an alternative approach that queries ZINC15 for molecular SMILES was used. The procedure was largely the same, except that up to 750 potential decoys were retrieved for each ligand protomer based on molecular weight and cLogP



of the decoy SMILES. Then an additional step was performed in which ChemAxon's Jchem was used to generate protonation states for these decoys' SMILES, followed by calculation of the remaining molecular properties by RDKit to determine whether they matched the ligands in property space.

**Extrema.** To generate extrema sets for all 43 targets, the molecular weight and cLogP values of the DUD-E ligand set were calculated using RDKit, and the interquartile range was calculated. For each charge, we retrieved a minimum of 1000 "in-stock" or "make-on-demand" molecules, built at physiological pH of 7.4, from ZINC15 in readily dockable format in this molecular weight and cLogP property space. Most of these molecules fall within charge ranges from -2 to +2, but there exist molecules with outlier charges as well. These dockable molecules were docked to their protein targets, and enrichment calculations were performed (see Docking Calculations).

**Goldilocks.** For generating the single Goldilocks decoy set, which is used for all 43 targets, the same procedure as with Extrema was used. However, instead of matching the decoys to an input ligand set, "in-stock" 3D-built molecules for each charge ranging from -2 to +2 within the property space ( $300 \text{ Da} \leq \text{MW} \leq 350 \text{ Da}$ ,  $2 \leq \text{cLogP} \leq 3$ ) were retrieved. For each charge, 3D-built molecules were retained until they reached half of the total number of 3D molecules with that charge, and within that molecular weight and cLogP property space (on December 10<sup>th</sup>, 2019). These dockable molecules were docked to their protein targets, and enrichment calculations were performed (see Docking Calculations).

**Docking Calculations.** The PDB structures assigned to forty DUD-E targets were retrieved and prepared in an automated fashion by in-house scripts based on the DOCK Blaster pipeline<sup>39</sup> (blastermaster.py in the DOCK3.7 distribution) for generating docking grids. The docking preparations for AmpC<sup>9,40,41</sup>, DRD4<sup>9,13</sup> (PDB:5WIU), and MT1<sup>42</sup> (PDB: 6ME3) were manually prepared. Thin sphere layers were utilized for AmpC, DRD4, and MT1 to extend the dielectric boundary from the solute surface for Poisson-Boltzmann calculations<sup>43</sup> with radii of 2.0 Å, 1.0 Å, and 1.9 Å, respectively. For all other systems, the default DOCK Blaster preparation was used in which the full binding site was filled with low-dielectric spheres of radius 1.9 Å for Poisson-Boltzmann calculations, thereby modeling the full binding site as low dielectric solute. The magnitudes of the partial charges of five AmpC residues and two MT1 residues were increased without changing the net residue charges<sup>41</sup>. For all DUD-E targets, their DUD-E assigned PDB ligand was used for generating 45 matching spheres, to which molecules are matched to during docking. For DRD4 and MT1, matching spheres were generated based on the atomic coordinates of nemonapride, and 2-phenylmelatonin, respectively. Docking calculations were performed with DOCK3.7.2<sup>44</sup>. Ligand conformations were generated by OpenEye's Omega<sup>45</sup>. Ligands were only scored if the number of ligand heavy atoms contained within the ligand ranged from 4 to 100. For each ligand hierarchy (each rigid fragment contained within the ligand), the maximum number of matches generated was set to 5000. For AmpC and DRD4, the large-scale docking setup was used, in which the target number of ligand hierarchy matches was set to 1000, and up to 500 Simplex minimization<sup>46</sup> steps were performed for each top scoring pose of each docked molecule, starting with initial translations of 0.2 Å and initial rotations of 5°. For MT1, the

target number of ligand hierarchy matches was set to 5000, and up to 500 Simplex minimization steps were performed for each top scoring pose of each docked molecule. All other DUD-E systems did not use Simplex minimization. To judge performance, the adjusted log AUC was used, which is analogous to the area under the receiver operator characteristic curve<sup>43</sup>. The adjusted log AUC subtracts the log AUC of the random curve (14.462%) to ensure random enrichment is 0%. For the DUD-E benchmarking calculations, the DUD-E ligands for each target are used as the ligand set for enrichment calculations. For all plots for DUDE-Z, Extrema, and Goldilocks, the reduced ligand set after clustering by ECFP4 Tanimoto coefficient of 0.7 is used for enrichment calculations.

To prepare different scoring function coefficient combinations, the “electrostatic\_scale” and “ligand\_desolv\_scale” parameters of the INDOCK files for each target were modified to be 0.3, 0.5, 0.7, or 1.0, generating 16 different combinations of DOCK scoring weights. The van der Waals scoring function coefficient was maintained at 1.0 for all docking calculations. All other parameters in the INDOCK file, docking grids, and matching spheres were kept identical.

**Bootstrapping.** For each bootstrap replicate (50 total for each system), ligands and decoys were chosen at random with replacement until the same sample size as the original enrichment set was reached. Each new hitlist was then sorted by the original docking energy, and a new adjusted log AUC is calculated. Z-tests were performed to test the significance of the difference between the means of two bootstrapped distributions. With the p value smaller than 0.05, the null hypothesis of equal mean and

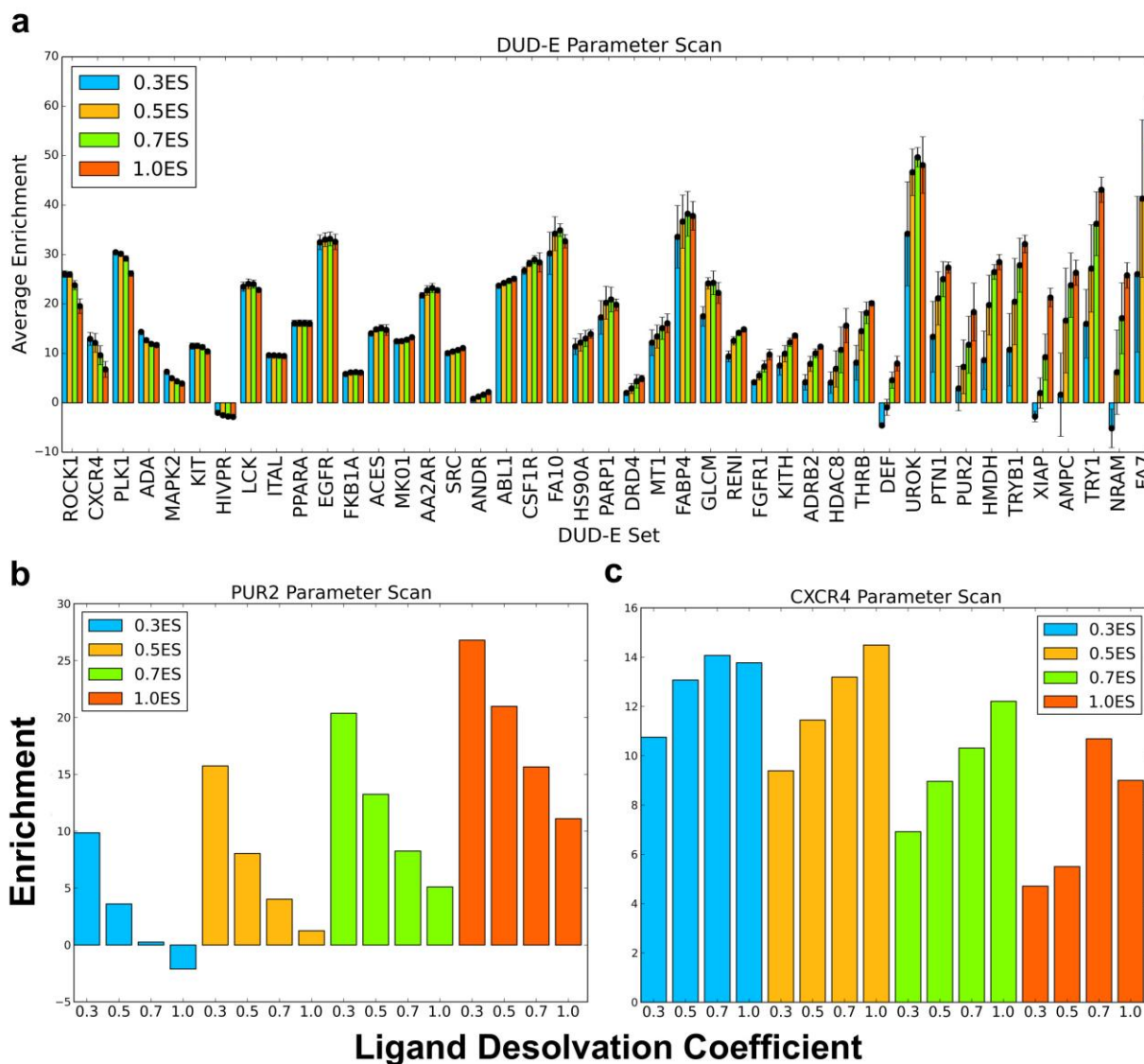
distribution is rejected. The Z-test is chosen since the number of bootstrap replicates is larger than 30 and the bootstrapped distribution follows the normal distribution.

### 3.4 Results

#### DOCK scoring function optimization using property-matched decoys

We were confronted with the liabilities of relying on property matched decoys in an investigation of different weighting terms in the DOCK3.7 scoring function<sup>43,44</sup>. We initially tried to use enrichment to guide the optimization of the scoring function by varying the coefficients of the electrostatics and ligand desolvation contributions to the total docking score. We scanned across electrostatics and ligand desolvation weighting for 41 DUD-E targets, and for the MT<sub>1</sub> melatonin receptor (MT1) and D4 dopamine receptor (DRD4), which have the advantage of hundreds of experimentally tested docking predictions<sup>9,42</sup> (**Figure 3.1**). To measure enrichment, we used a log-weighted area under the curve approach, subtracting from this enrichment expected at random (adjusted Log AUC<sup>43</sup>, **Figure 3.1**, **Table 3.1**). This approach equally weights enrichment in the top 0.1 to 1% of the library with that within the top 1 to 10% and the top 10% to 100% of the library, thus up-weighting early enrichment. Sampling sixteen combinations of weights (four electrostatics, four ligand desolvation with constant van der Waals) revealed that enrichment correlated with the electrostatics and ligand desolvation terms (**Figure 3.1**, **Table 3.1**, but see *Sensitivity Analysis*, below, for the significance of these differences). In most DUD-E targets, increasing the electrostatic coefficient increased enrichment. This included systems such as GAR transformylase (PUR2), which had its best enrichments with weights of 1.0 for electrostatics and 0.3 for

ligand desolvation (**Figure 3.1**). These same coefficients, however, negatively impacted other systems, such as C-X-C chemokine receptor type 4 (CXCR4), where the same weights that were optimal for AmpC led to worse performance. Instead, CXCR4 had its highest enrichment with weights of 0.5 on the electrostatics and of 1.0 on the ligand desolvation terms (**Figure 3.1**).



**Figure 3.1. Ligand desolvation and electrostatics weights alter enrichment.** a) For each electrostatic coefficient (0.3, 0.5, 0.7, 1.0), the average adjusted log AUC value and standard error for the four ligand desolvation coefficients (0.3, 0.5, 0.7, 1.0) is

plotted. Individual enrichment plots for each electrostatic and ligand desolvation coefficient combination for PUR2 (b), and CXCR4 (c). Enrichments for PUR2 diminish as the ligand desolvation coefficient increases, while enrichments for CXCR4 improve as the ligand desolvation coefficient increases.

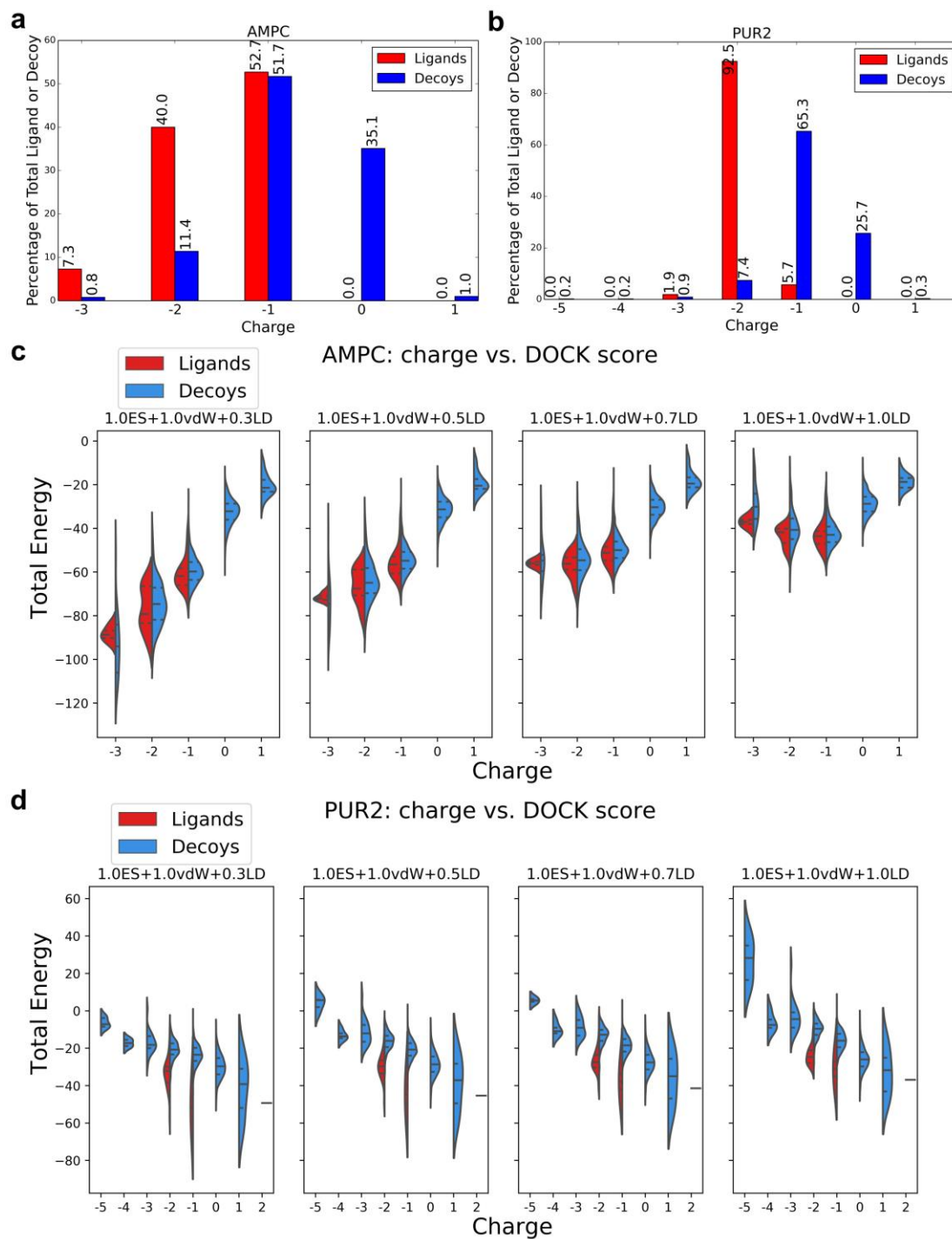
**Table 3.1.** Enrichments for DOCK3.7 Scoring Coefficients over 43 Targets

	<b>0.3ES</b>	<b>0.5ES</b>	<b>0.7ES</b>	<b>1.0ES</b>
<b>0.3LD</b>	16.3 (11, 7, 25)	13.88 (8, 10, 25)	11.84 (6, 8, 29)	9.95 (6, 8, 29)
<b>0.5LD</b>	19.23 (17, 10, 16)	17.94 (12, 12, 19)	15.82 (8, 14, 21)	12.71 (4, 14, 25)
<b>0.7LD</b>	20.09 (18, 14, 11)	19.76 (18, 18, 7)	18.65 (10, 22, 11)	15.94 (4, 17, 22)
<b>1.0LD</b>	19.84 (16, 14, 13)	20.2 (18, 17, 8)	20.01 (17, 21, 5)	19.05

Values outside the parentheses are the average adjusted log AUC enrichment values, while those within the parentheses refer to those targets that improved by 1 adjusted log AUC value, stayed within  $\pm 1$  log AUC, and diminished by 1 adjusted log AUC value vs. the standard scoring function (1.0ES+1.0vdW+1.0LD).

Closer inspection revealed that the enrichment differences, and the sensitivity to scoring coefficients, were often explained by different formal charge distributions between ligands and decoys. For instance, for AmpC, larger weighting of electrostatic interactions improved enrichments because AmpC's ligands are all anionic, whereas 35% of AmpC's DUD-E decoys are neutral (**Figure 3.2**). Thus, as the weight on the ligand desolvation term, which scales with net charge, decreases, AmpC's anionic ligands are penalized less (**Figure 3.2**). When unconstrained, as with an electrostatics weighting of 1.0 and ligand desolvation weighting of 0.5, the "optimized" scoring function, i.e. the coefficients that maximize enrichment, prioritizes charge over other molecular properties versus the unweighted, standard scoring function. Similarly, most PUR2 ligands are dianions, while its decoys are mainly mono-anionic or neutral (**Figure**

**3.2)** and docking with reduced ligand desolvation coefficients favor the ligands over the decoys (**Figure 3.2**). Even if all our molecular properties, besides charge, are well-matched in the DUD-E benchmarking sets, altering the scoring function weights of electrostatics and ligand desolvation allows DOCK to simply recognize gross physical differences between ligands and decoys, rather than detailed molecular interactions, reflecting an imbalance in the DUD-E ligand and decoy properties.



**Figure 3.2. Proportion of charged molecules in DUD-E sets affects enrichment.** Percentage of ligands or decoys in the DUD-E set with a given charge for AmpC  $\beta$ -lactamase (AmpC, a) and GAR transformylase (PUR2, b). Comparison of DOCK energy and molecule charge for AmpC  $\beta$ -lactamase (AmpC, c) and GAR transformylase (PUR2, d) for the electrostatic coefficient of 1.0 and the four ligand desolvation weights



(0.3, 0.5, 0.7, 1.0). Central dotted lines of DOCK energies represent the medians, upper dotted lines represent the third quartiles, and lower dotted lines represent the first quartiles for both scoring functions. The lowest points represent the minimum DOCK energies and the highest values represent the maximum DOCK energies. The AmpC ligands in DUD-E are predominantly anionic (a), and while this is also true of the decoys, the latter harbor a higher ratio of neutral molecules. Increasing the ligand desolvation coefficient ranks neutral molecules higher (as sorted by total DOCK energy), favoring decoys, and enrichment decreases (c). Conversely, increasing the electrostatic coefficient favors the anionic ligands, increasing the enrichment. The large majority of PUR2 ligands are di-anionic while the decoys are monoanionic (b), providing an advantage to the ligands at lower ligand desolvation coefficients (as sorted by total DOCK energy) (d), as they can form more favorable electrostatic interactions with the protein without a large ligand desolvation cost.

### **New Property-Matched Decoy Method**

The original DUD-E benchmarking set<sup>23</sup> was built to correct the charge imbalance in the original DUD set<sup>22</sup> by including net charge during property matching. However, there remains a disconnect between the charges contained within the 2D SMILES, and the charges present in 3D dockable molecules from DUD-E. For example, calculating the formal charges of the AmpC ligand and decoys SMILES contained within the DUD-E benchmarking set suggest that 60% and 38% of ligands are neutral and monoanionic, respectively, while 43% and 56% of decoys are di- and mono-anionic, respectively, compared with the actual charge representation in the dockable set (**Figure 3.2**). During molecular building, the charge populations change based on which protomers are predicted to exist at physiological pH, producing charge imbalances that were not present in the SMILES representation.

To address this, we created a new decoy preparation pipeline that better charge-matched ligands to decoys (freely available at <http://tldr.docking.org>), such that ligand

and decoy protomers are only considered in their dockable, 3D representation so that there is no likelihood of charge imbalances occurring. Up to 50 decoys are generated for each ligand taking into account charge, molecular weight, calculated LogP, number of rotatable bonds, number of hydrogen bond acceptors and donors, while ensuring that these decoys are structurally dissimilar to each other and to the ligands to which they are matched (**Table 3.2**). By default, and always for proteins with more than 100 ligands, the ligands are first clustered by an ECFP4 Tc of 0.7 to reduce the dominance of narrow congeneric series. The ligand with the smallest molecular weight from each cluster is chosen for property-matching. These changes improve the DUD-E design, without changing its underlying logic.

**Table 3.2:** Ligand and Decoy Properties for 43 Protein Targets

	<b>DUD-E</b>	<b>DUDE-Z</b>	<b>Extrema</b>	<b>Goldilocks</b>
<b># Unique Ligands</b>	8267	2312	-	-
<b># Unique Decoys</b>	477924	69994	732309	1145472
<b># Unique Decoy Scaffolds</b>	162286	33292	143423	317316

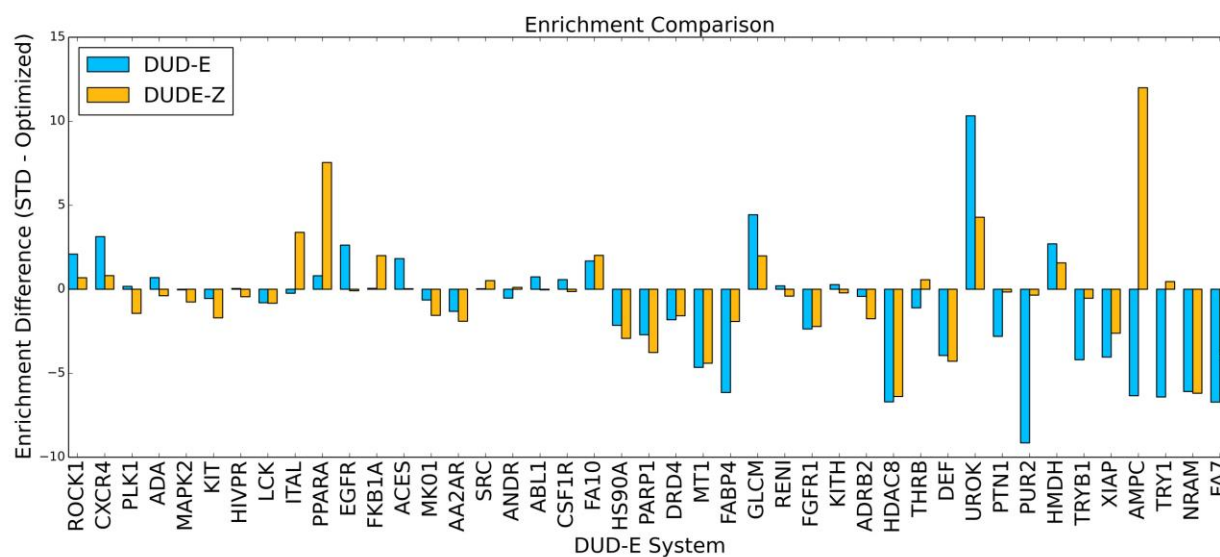
### Improved property-matched decoys reduce false enrichment

With these changes in hand, we compared the scoring function with a 0.5 weight on ligand desolvation, the “optimized” scoring function, to the standard, unweighted scoring function to determine whether the improved enrichments stood up to better charge-matching between ligands and decoys. Competition with the better charge-matched decoys reduced the enrichment differences between the standard and the “optimized” 0.5 ligand desolvation scoring functions from  $>1$  with the original DUD-E set, to 0.35, supporting the hypothesis that more closely property-matched decoys would be less susceptible to imbalances in electrostatics and ligand desolvation energies (**Figure 3.3**, and see *Sensitivity Analysis*, below, for the significance of such differences). For instance, AmpC, whose enrichment was better with the optimized scoring function by more than 6 log adjusted AUC, with the new property-matched decoy background now much favors the standard scoring function, attaining an enrichment of 20.92 over the “optimized” scoring function’s 8.93. Similarly, the DUD-E enrichment difference for PUR2 was also greater than 6 log adjusted AUC, but the difference becomes 0.35 in the new decoy set. Similar behavior where complete charge-matching reduces preference for the optimized scoring function is seen in multiple systems including fatty acid binding protein 4 (FABP4), protein-tyrosine phosphatase 1 (PTN1), tryptase beta-1 (TRYB1), and trypsin I (TRY1). The opposite also occurs, where preference for the standard scoring function is diminished in the presence of better charge-matched decoys such as in Rho-associated protein kinase 1 (ROCK1), C-X-C chemokine receptor type 4 (CXCR4), and epidermal growth factor receptor (EGFR). Overall, the average adjusted log AUC values for the 42 targets dropped from 19.05 and 20.2 for the standard and

“optimized” scoring functions, respectively, with the original DUD-E benchmarking sets, to 14.82 and 15.17 with the new, better-matched decoy sets (**Table 3.3**). This enrichment drop reflects the better choice of decoy molecules in the new benchmarks, making the challenge harder, appropriately, for the docking program.

**Table 3.3:** Average Enrichment log AUC values for Different Decoy Sets

	DUD-E	DUDE-Z	Extrema		Goldilocks	
			DUD-E Ligands	DUDE-Z Ligands	DUD-E Ligands	DUDE-Z Ligands
<b>Optimized (1.0ES+1.0vdW+0.5LD)</b>	20.2	15.17	25.80	15.97	41.84	28.33
<b>Standard (1.0ES+1.0vdW+1.0LD)</b>	19.05	14.82	25.85	15.72	41.31	27.74
<b>Difference</b>	-1.15	-0.35	0.05	-0.25	-0.53	-0.59

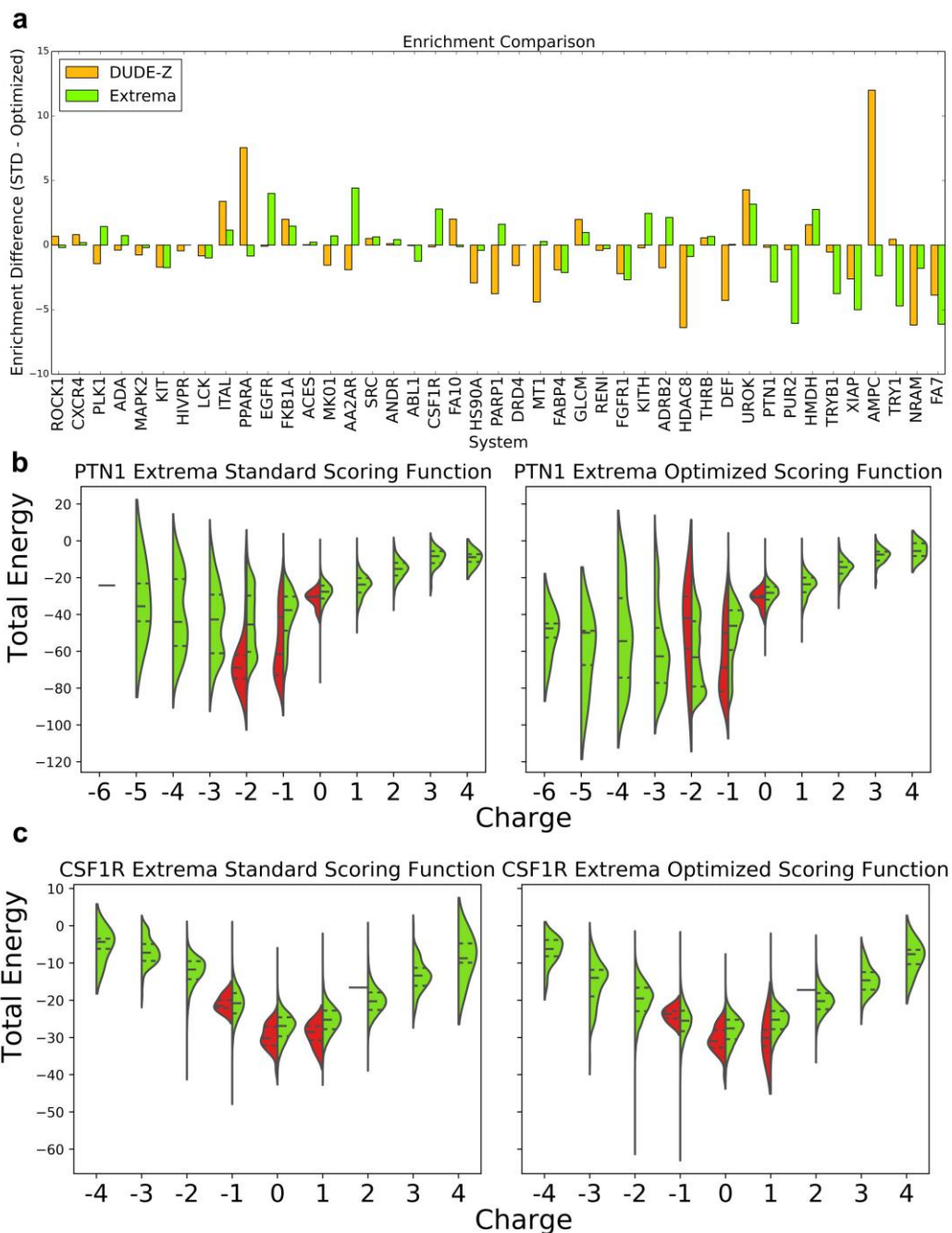


**Figure 3.3. Enrichment comparison between DUD-E and DUDE-Z.** a) Enrichment differences between the standard, unweighted scoring function and the optimized scoring function (1.0ES + 1.0vdW + 0.5LD), comparing the original DUD-E decoys (blue bars) and decoys prepared with the new DUDE-Z pipeline (orange bars), in which decoys are better charge-matched. Apparent advantages for the weighted scoring function dissipate on better charge matching.

## Beyond property-matched decoys: charge extrema

Given the sensitivity to even small differences in charge matching between ligands and decoys, we thought it worthwhile to investigate how sensitive the docking was not only to property matching, but to extremes intentionally outside the property range of the ligands. We reasoned that docking parameters might be unintentionally optimized to weight particular energetic terms at the expense over others. Such blind spots might only be illuminated when comparing the performance of physically extreme molecules. Based on our experience with the impact of electrostatic and desolvation weighting above, we focused on ligands representing charge extremes, probing for over-weighted electrostatic interactions, or underweighted desolvation penalties, in our scoring function. These charge-extrema sets were populated with decoys that have similar physical properties (molecular weight, cLogP) to the ligands queried, but include *all* charges from -2 to +2, taken from “in-stock” and “make-on-demand” libraries in ZINC15<sup>47</sup>. If many molecules bearing a net charge of -2 score better than AmpC’s mono-anions, for instance, this would indicate a bias in the scoring that would have been concealed by the charge-matched decoys. We generated sets of property-matched charge-extreme decoys for 43 targets. These charge outlier decoys ( $\leq -2$  and  $\geq +2$ ) comprised on average 37% (272K of 732K molecules) of benchmarks, ranging from 15% (tryptase beta-1, TRYB1) to 57% (neuraminidase, NRAM). For a well-balanced scoring function, which properly captures molecular interactions, including charge extrema should *improve* ligand enrichment, since decoys bearing unreasonable charges should be readily recognized, which is indeed what we see, though performance improves only slightly (**Figure 3.4, Table 3.3**, and see *Sensitivity Analysis*,

below for the significance of such differences), with systems with charged ligands being affected significantly. For example, GAR transformylase (PUR2, **Figure 3.4**) recognizes tri- and di-anionic ligands. When screened against a large extrema set with down-weighted desolvation, cations begin to dominate, behavior that the standard scoring function is, at least, partially, able to combat (**Figure 3.4**). Similar behavior is seen with protein-tyrosine phosphatase 1b (PTN1), which predominantly binds mono- and di-anions in the standard scoring function but begins to prioritize tri- and tetra-anions when the optimized scoring function is utilized. As with GAR transformylase, the increased desolvation cost in the standard scoring function actually diminishes performance relative to the “optimized” scoring function as it penalizes both extreme-charged ligands and decoys. On the other hand, epidermal growth factor receptor (EGFR) and macrophage colony stimulating factor (CSF1R, **Figure 3.4**), which perform better with the standard scoring function over the optimized scoring function with extrema, both recognize neutral ligands. When these two targets are screened with charge extrema, the standard scoring function is more equipped to penalize inappropriate charges over the optimized scoring function, which in the presence of charge extrema is flooded with anions and cations. Each of these cases can be explained by the underweighting of the ligand desolvation penalty in a scoring function optimized against the DUD-E set that **i.** had a discrepancy between ligand and decoy charges and **ii.** was not challenged with charged extrema, as we show here.



**Figure 3.4. Enrichments and charge priority of DUDE-Z and Extrema. a)** Enrichment differences between the standard scoring function and the weighted scoring function using the new DUDE-Z decoy pipeline and the charge extrema decoys. **b,c)** through **e)**. Comparing DOCK energy and molecule charge of the standard and optimized scoring functions using DUDE-Z ligands and using charge extrema decoys for **b)** protein-tyrosine phosphatase 1 (PTN1) and **c)** macrophage colony stimulating factor receptor (CSF1R). Central dotted lines of DOCK energies represent the medians,

upper dotted lines represent the third quartiles, and lower dotted lines represent the first quartiles. The lowest points represent the minimum DOCK energies and the highest values represent the maximum DOCK energies for both scoring functions. As ligand desolvation is downweighted in the optimized scoring function, more extreme charges score better, which is advantageous for targets that have extreme charged ligands like PUR2 and PTN1. However, this becomes problematic and decreases enrichment for systems whose ligands are less extreme like EGFR and CSF1R.

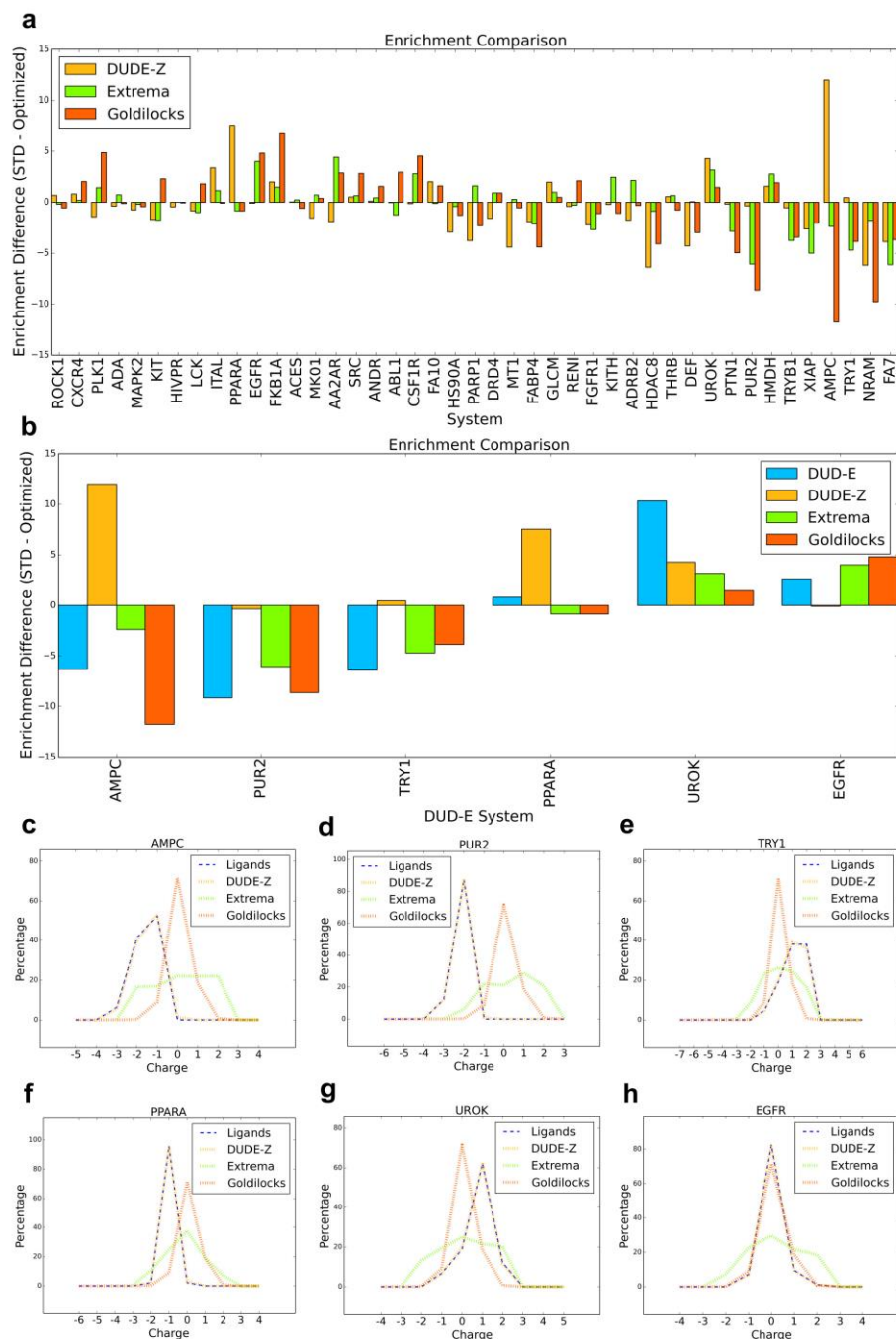
If charge extrema can reveal cryptic pathologies in docking scoring, so too can testing against molecules that are intentionally *un*matched from the physical properties of the ligands, but instead reflect the molecules of the overall library itself. Since each receptor will have its own ligand preferences, certainly with the biases from the medicinal chemistry literature, for any given receptor, the average library molecule may well-represent a physical property outside those of the receptor's ligands, exposing the docking screen to new, previously unsampled physical properties. Thus, we investigated control calculations with a set of 1.1 million ZINC molecules. These comprised over 300,000 Bemis-Murcko scaffolds<sup>48</sup> representing the middle of the range of physical parameters of the library; not too big, not too small, not too polar, and not too greasy (Goldilocks). Docking these to the 43 targets resulted in log adjusted AUC values of 27.84 and 28.33 for the standard and "optimized" scoring functions, respectively (**Table 3.3**). These are higher than the enrichments with the property-matched sets, as expected owing to its non-property-matched nature; the differences between the two scoring functions against the Goldilocks set are small (see *Sensitivity Analysis* below).

Even against a background of high enrichment, there are targets for which performance varies between the two scoring functions. Here we focus on illustrative



targets where the differences are substantial and significant (see *Sensitivity Analysis*, below). In AmpC  $\beta$ -lactamase, tests against the DUDE-Z set suggest that the standard, unweighted scoring function led to better enrichments than the putatively optimized one where ligand desolvation was down-weighted by 0.5 (**Figure 3.3**), in contrast to the DUD-E benchmark test that had led to this new weighting. Against the Goldilocks benchmark, however, the situation reverts, with the optimized scoring function performing better than the standard scoring function, with an enrichment difference over 11 in adjusted log AUC (**Figure 3.5**). This difference is only partly captured by the extrema set, where the difference is only slightly larger than 2 adjusted log AUC. Similarly, GAR transformylase (PUR2) sees the relative enrichment of the optimized scoring function rise by almost 10 units of adjusted log AUC versus the standard scoring function with the Goldilocks set vs. DUDE-Z, while with trypsin I (TRY1), ligands favor the optimized scoring function using the Goldilocks benchmark by almost 4 adjusted log AUC units versus the less than 1 unit difference using the DUDE-Z set. A few targets, such as FK506-binding protein 1A (FKB1A) and polo-like kinase 1 (PLK1) see the opposite effect—the optimized scoring function performs noticeably worse with the Goldilocks benchmark versus DUDE-Z. These differences are explained by differences in the properties of the decoys in the different benchmarks. In DUDE-Z, the decoy physical properties are tightly calibrated to those of the ligands. Conversely, Goldilocks represents the physical properties of the library to-be-docked. For targets recognizing ligands with physical properties much different than “lead-like”<sup>49</sup> molecules, which dominate the Goldilocks benchmark and the library it represents, such as AmpC, GAR transformylase (PUR2), and trypsin I (TRY1), the DUDE-Z set will be a more stringent

test (**Figure 3.5**). However, scoring term weights that optimize performance against it will not always translate to a lead-like benchmark like Goldilocks. For these systems, the key differences are in the distribution of charge states of the ligands and the decoys: in DUDE-Z, these are well matched, while in Goldilocks, and the ultra-large library that it represents, mono-, di-, and tri-anions, as well as di-cations, are far less common than among the known inhibitors of these targets (**Figure 3.5**), providing opportunities for these ligands to exploit the optimized scoring function with its down-weighted ligand desolvation term and score well. For systems that bind molecules within lead-like space, such as peroxisome proliferator-activated receptor alpha (PPARA), urokinase-type plasminogen activator (UROK), and epidermal growth factor receptor (EGFR), the enrichment differences between the standard and optimized scoring functions diminish, and even begin to favor the standard scoring function (**Figure 3.5**), as outlier charges are unable to exploit liabilities within the optimized scoring function.



**Figure 3.5: Enrichments and charge priority of DUDE-Z, Extrema, and Goldilocks.**

a) Enrichment differences between the standard scoring function and optimized scoring function comparing the new DUDE-Z benchmarks, charge extrema decoys, and the Goldilocks benchmarks, with a focus on the enrichment changes in specific targets (b). Comparison of net charge of ligands and benchmark decoys for AmpC  $\beta$ -lactamase (AmpC, c), GAR transformylase (PUR2, d), trypsin I (TRY1, e), peroxisome proliferator-activated receptor alpha (PPARA, f), urokinase-type plasminogen activator (UROK, g),

and epidermal growth factor receptor (EGFR, **h**). For systems whose ligands have more extreme charges, there is typically small overlap in ligand charges and decoy charges, providing an advantage to the extreme charged ligands with the optimized scoring function. However, in systems where the ligand charges overlap more significantly with the decoy charges, the standard scoring function begins to perform better as there are no extreme charged ligands to exploit the lower desolvation cost and rank more favorably.

Up until now, we have seen results shift as we change the benchmark from DUDE to the optimized DUDE-Z to Extrema to Goldilocks. A natural reaction might be to despair of benchmarking entirely. Our own view is that each of these benchmarks is useful, and together can inure developers and users from false conclusions around scoring function and docking parameter optimization. The different lessons that each benchmark teaches reflect weaknesses of enrichment as a metric; it nevertheless remains a crucial criterion for docking performance. These are points to which we will return.

### **Sensitivity Analysis & Statistical Significance**

Area Under the Curve (AUC) and its variants are widely used as a single value measure of docking performance<sup>43,44,50-54</sup>. In comparing an innovation with the current best practice, it is common to see improvements in enrichment across a benchmarking set. It is important to understand when such improvements are significant beyond the variation one might see with small changes to docking parameters. To assess confidence intervals on enrichment plots, we turned to an empirical bootstrapping approach. In this method, we calculate enrichments multiple times for any given benchmark, each time picking a random subset of the ligands and decoys in the set, retaining the same sample size as the original set. For many of the DUDE-Z targets, this is readily done, as only a subset of the possible ligands is typically represented, and

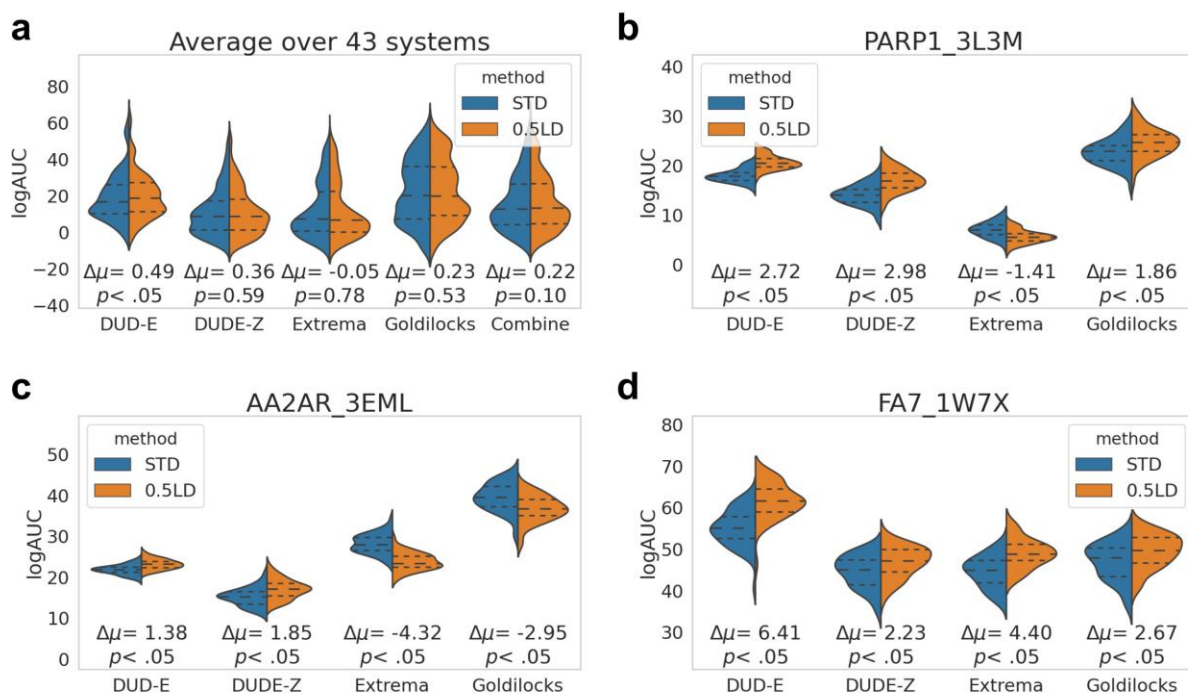
many more property-matched decoys are typically available from ZINC. With the new benchmark, whose ligands closely resemble the canonical ones, and whose decoys reflect the same property matching, a new enrichment is calculated.

Repeated for 50 random subsets of ligands and decoys for each target, this approach allows one to calculate confidence intervals of enrichment (adjusted log AUC). We did so for the same 43 targets, recording the variance of the enrichments. Based on these bootstrapping calculations, we find that the average 95% and 75% confidence interval over the 43 systems is about 9.4 and 5.8 adjusted log AUC units, respectively.

Naturally, individual systems varied in their confidence levels: from a relatively tight distribution for Androgen Receptor (ANDR, 95% CI of 3.0), to a much wider distribution for fatty acid binding protein-4 (FABP4, 95% CI of 15.6) (**Figure A.3.1**). Bootstrapping can also be used to compare the performance of two docking methods or two scoring functions. The Z-test and corresponding p-values are used here, since the number of bootstrap replicates is over 30, and the bootstrapped distribution follows the normal distribution.

Figure 2.6 shows the bootstrapped distribution comparison between the standard (STD) and “optimized” (0.5LD) scoring functions with DUD-E, DUDE-Z, Extrema, and Goldilocks as decoy sets on 41 DUD-E targets, as well as the melatonin MT<sub>1</sub> receptor and the dopamine D4 receptor where we have not only experimentally measured docking true but also docking false positives (**Fig. A.3.2**). Innovations that we might have otherwise considered successful are often found to be statistically

indistinguishable, or to be significant against one background but not another. Screening poly-ADP-ribose polymerase 1 (PARP1) with DUD-E, DUDE-Z, and Goldilocks decoy sets shows significant improvement with the optimized scoring function over the standard scoring function, whereas performance is significantly worse with Extrema (**Figure 3.6**). In adenosine 2A receptor (AA2AR, **Figure 3.6**), ligands in the presence of DUD-E and DUDE-Z decoy sets significantly favor the optimized scoring function, but flip to favoring the standard scoring function in the presence of Extrema and Goldilocks sets, versus in Coagulation Factor VII (FA7, **Figure 3.6**), ligands always significantly favor the optimized scoring function regardless of the decoy background (see **Fig. A.3.3** for difference distributions and **Fig. A.3.4** for bootstrapping plots of all 43 systems). However, we note that only when screened with the DUD-E decoys are the enrichment differences in these scoring functions significantly different (**Figure 3.6**), showing for all other decoys sets insignificant differences. When all decoy sets are combined, the bootstrapping enrichment differences remain insignificant.



**Figure 3.6. Bootstrapping enrichment differences using different decoy backgrounds.** Applying bootstrapping to the different decoy backgrounds demonstrates that while there may be statistically significant differences in terms of performance between the scoring functions for particular systems, if all the bootstrapping enrichments are combined for all decoy sets over all 43 systems, there is no statistically significant difference between the standard and optimized scoring functions, demonstrating that one can be deceived by significant differences between the two scoring functions when only considering one decoy background. Average bootstrapping statistics on the enrichments for DUD-E, DUDE-Z, Extrema, Goldilocks, and all Decoy sets (Combined) for all 43 systems (a). Individual bootstrapping statistics (50 for each) on the enrichments (adjusted log AUC values) for DUD-E, DUDE-Z, Extrema, and Goldilocks decoy backgrounds for poly-ADP-ribose polymerase I (PARP1, b), adenosine 2A receptor (AA2AR, c), and coagulation factor VII (FA7, d). From the 50 bootstrapped adjusted log AUC values generated, central dotted lines represent the medians, upper dotted lines represent the third quartiles, and lower dotted lines represent the first quartiles. The lowest points represent the minimum adjusted log AUC values and the highest points represent the maximum adjusted log AUC values generated from bootstrapping.

### 3.5 Discussion

Four themes emerge from this work. First, for all their strengths, property-matched decoys alone can mislead in evaluating docking performance. Scoring functions can exploit physical property differences between ligands and decoys even in relatively well-balanced sets, as we see comparing the original DUD-E and the refined DUDE-Z sets. Decoys that are intentionally *non*-property matched, such as the Extrema set that explores ligands with high molecular charges, and the Goldilocks set, whose decoys can be far different from the known ligands, but which represent the properties of the ultra-large database to be docked, reveal liabilities that are hidden by the property-matched sets. Second, enrichment, which is perhaps the key criterion for library docking assessment, remains a weak metric, ungrounded in physical theory or observables. Third, our understanding of this metric can be strengthened with confidence intervals, which can be readily estimated. These confidence margins are often surprisingly large, and apparently different enrichments are often statistically indistinguishable. Finally, we make the new tools developed here, including generation of better property-matched decoys (DUDE-Z), charge Extrema, Goldilocks, and bootstrapping adjusted log AUC ranges, available and free to use for the community.

Property-matched decoys remain crucial for docking evaluation<sup>22,23,31</sup>, reducing the ability of scoring functions to exploit gross physical property differences between ligands and the random molecules that had initially been used in the field<sup>28</sup>. But property-matching has its own liabilities, revealed by other backgrounds. For instance, property matching decoys to the GAR transformylase, AmpC  $\beta$ -lactamase, or trypsin I receptor



ligands will result in decoys that have charge ranges tightly distributed around -2, -1, and +1 to +2 formal charges, respectively. A scoring function that overweights electrostatic interaction energies, or underweights desolvation energies, may not be revealed by such property matched decoys. This is what we observed with what appeared to be an “optimized” function that down-weighted ligand desolvation, improving average enrichment over 43 systems. This apparent improvement was eliminated not only by better charge matching in the optimized DUDE-Z set, its basis in an over-weighted electrostatic interactions was illuminated by a charge Extrema set (Figure 4). Similarly, benchmarks that are well-matched around ligands with unusual physical properties—in this study, highly charged ligands—will not reveal liabilities that a background representing the properties of the overall library can illuminate. This is what we observe for the Goldilocks benchmark (**Figure 3.5**).

Enrichment of ligands over property matched decoys<sup>23,50,51,55-59</sup> is widely used for parameter optimization and scoring function development<sup>43,60-62</sup>. Because enrichment is ungrounded in physical theory, it is sensitive both to changes in the decoy background, which are usually only reasonable guesses, and to the ligands, which represent experimental observables, flawed though these too can be. We do not wish to undercut enrichment as a metric of docking—weak as it is, it remains crucial to progress in the field. What this study teaches is that our confidence in enrichment can be much strengthened by using multiple decoy backgrounds. Correspondingly, the significance of enrichment differences with different docking parameterization, and with different scoring functions, should be controlled for. One way to do so is via the bootstrapping

method we outline here (**Figure 3.6**), which can insulate one from false conclusions about differences that fall within the variation expected from small changes in the ligands and decoys used (scripts to implement this are available at <http://tldr.docking.org>).

Confronted with ever more decoy benchmarks, and the time it takes to run a full set of controls, it is natural to wonder if there is no end to the cottage industry of new benchmarks. One can imagine spending too much time on these sanity checks, and too little on the actual prediction of new chemical matter with prospective docking. Nevertheless, the time and expense of sourcing and physically testing new chemical matter, and for eliminating experimental artifacts<sup>47,63,64</sup> still far exceeds the cost of running these computational controls. Property-matched benchmarks are rarely composed of more than a few thousand molecules for a given target, and even the Goldilocks set comprises less than 2 million molecules, less than 1% the size of the ultra-large libraries now being prosecuted<sup>9,10,42</sup>. To make these controls accessible to the community, we provide the optimized DUDE-Z benchmarks at <http://dudez.docking.org>. We also provide a web service that allows investigators to create bespoke Extrema and Goldilocks sets, and enables bootstrapping tests for statistical significance—freely available at <http://tldr.docking.org>.

### 3.6 Acknowledgements

Supported by US National Institutes of Health grants GM71896 (to JJI) and by R35GM122481 (to BKS). We are grateful to OpenEye Scientific Software for an academic license for Omega, OEChem and other tools and to ChemAxon for an academic license for JChem, Marvin and other software. We thank the providers of public databases and free software from which ZINC has benefitted: RDKit, DrugBank, HMDB, ChEBI, ChEMBL. We thank members of the Shoichet Lab for testing the software and for timely feedback and thank Roger Sayle and John Mayfield at Nextmove Software for access to Arthor and SmallWorld, and for discussions.

### 3.7 Abbreviations

AA2AR, Adenosine A2a receptor; ABL1, Tyrosine-protein kinase ABL; ACES, Acetylcholinesterase; ADA, Adenosine deaminase; ADRB2, Beta-2 adrenergic receptor; AMPC, Beta-lactamase; ANDR, Androgen Receptor; CSF1R, Macrophage colony stimulating factor receptor; CXCR4, C-X-C chemokine receptor type 4; DEF, Peptide deformylase; DRD4, D4 Dopamine receptor; EGFR, Epidermal growth factor receptor erbB1; FA10, Coagulation factor X; FA7, Coagulation factor VII; FABP4, Fatty acid binding protein adipocyte; FGFR1, Fibroblast growth factor receptor 1; FKB1A, FK506-binding protein 1A; GLCM, Beta-glucocerebrosidase; HDAC8, Histone deacetylase 8; HIVPR, Human immunodeficiency virus type 1 protease; HMDH, HMG-CoA reductase; HS90A, Heat shock protein HSP 90-alpha; ITAL, Leukocyte adhesion glycoprotein LFA-1 alpha; KIT, Stem cell growth factor receptor; KITH, Thymidine kinase; LCK, Tyrosine-protein kinase LCK; MAPK2, MAP kinase-activated protein kinase 2; MK01, MAP

kinase ERK2; MT1, Melatonin MT1 receptor; NRAM, Neuraminidase; PARP1, Poly [ADP-ribose] polymerase-1; PLK1, Serine/threonine-protein kinase PLK1; PPARA, Peroxisome proliferator-activated receptor alpha; PTN1, Protein-tyrosine phosphatase 1B; PUR2, GAR transformylase; REN1, Renin; ROCK1, Rho-associated protein kinase 1; SRC, Tyrosine-protein kinase SRC; THRB, Thrombin; TRY1, Trypsin I; TRYB1, Trypsin beta-1; UROK, Urokinase-type plasminogen activator; XIAP, Inhibitor of apoptosis protein 3

## References

1. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. Automated docking with grid-based energy evaluation. **13**, 505-524, doi:10.1002/jcc.540130412 (1992).
2. Shoichet, B. K. & Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng* **6**, 723-732, doi:10.1093/protein/6.7.723 (1993).
3. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **9**, 1-5, doi:10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6 (1996).
4. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739-1749, doi:10.1021/jm0306430 (2004).
5. Lemmen, C. & Lengauer, T. Time-efficient flexible superposition of medium-sized molecules. *J Comput Aided Mol Des* **11**, 357-368, doi:10.1023/a:1007959729800 (1997).
6. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **11**, 425-445, doi:10.1023/a:1007996124545 (1997).
7. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261**, 470-489, doi:10.1006/jmbi.1996.0477 (1996).

8. Welch, W., Ruppert, J. & Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol* **3**, 449-462, doi:10.1016/s1074-5521(96)90093-9 (1996).
9. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229, doi:10.1038/s41586-019-0917-9 (2019).
10. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, doi:10.1038/s41586-020-2117-z (2020).
11. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185-190, doi:10.1038/nature19112 (2016).
12. Lansu, K. *et al.* In silico design of novel probes for the atypical opioid receptor MRGPRX2. *Nat Chem Biol* **13**, 529-536, doi:10.1038/nchembio.2334 (2017).
13. Wang, S. *et al.* D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* **358**, 381-386, doi:10.1126/science.aan5468 (2017).
14. Korczynska, M. *et al.* Structure-based discovery of selective positive allosteric modulators of antagonists for the M2 muscarinic acetylcholine receptor. *Proc Natl Acad Sci U S A* **115**, E2419-E2428, doi:10.1073/pnas.1718037115 (2018).
15. Huang, X. P. *et al.* Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **527**, 477-483, doi:10.1038/nature15699 (2015).
16. Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* **59**, 4103-4120, doi:10.1021/acs.jmedchem.5b02008 (2016).

17. Ballante, F. *et al.* Docking Finds GPCR Ligands in Dark Chemical Matter. *Journal of Medicinal Chemistry* **63**, 613-620, doi:10.1021/acs.jmedchem.9b01560 (2020).
18. Patel, N. *et al.* Structure-based discovery of potent and selective melatonin receptor agonists. *eLife* **9**, e53779, doi:10.7554/eLife.53779 (2020).
19. Kiss, R. *et al.* Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. *J Med Chem* **51**, 3145-3153, doi:10.1021/jm7014777 (2008).
20. Mannel, B. *et al.* Structure-Guided Screening for Functionally Selective D2 Dopamine Receptor Ligands from a Virtual Chemical Library. *ACS Chem Biol* **12**, 2652-2661, doi:10.1021/acscchembio.7b00493 (2017).
21. Scharf, M. M., Bunemann, M., Baker, J. G. & Kolb, P. Comparative Docking to Distinct G Protein-Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Mol Pharmacol* **96**, 851-861, doi:10.1124/mol.119.117515 (2019).
22. Huang, N., Shoichet, B. K. & Irwin, J. J. Benchmarking sets for molecular docking. *J Med Chem* **49**, 6789-6801, doi:10.1021/jm0608356 (2006).
23. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582-6594, doi:10.1021/jm300687e (2012).
24. Reau, M., Langenfeld, F., Zagury, J. F., Lagarde, N. & Montes, M. Decoys Selection in Benchmarking Datasets: Overview and Perspectives. *Front Pharmacol* **9**, 11, doi:10.3389/fphar.2018.00011 (2018).

25. Novotny, J., Bruccoleri, R. & Karplus, M. An analysis of incorrectly folded protein models. Implications for structure predictions. *J Mol Biol* **177**, 787-818, doi:10.1016/0022-2836(84)90049-4 (1984).
26. Park, B. & Levitt, M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys. *J Mol Biol* **258**, 367-392, doi:10.1006/jmbi.1996.0256 (1996).
27. Samudrala, R. & Levitt, M. Decoys 'R' Us: a database of incorrect conformations to improve protein structure prediction. *Protein Sci* **9**, 1399-1401, doi:10.1110/ps.9.7.1399 (2000).
28. Pham, T. A. & Jain, A. N. Parameter estimation for scoring protein-ligand interactions using negative training data. *J Med Chem* **49**, 5856-5868, doi:10.1021/jm050040j (2006).
29. Bissantz, C., Folkers, G. & Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **43**, 4759-4767, doi:10.1021/jm001044l (2000).
30. Kellenberger, E., Rodrigo, J., Muller, P. & Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**, 225-242, doi:10.1002/prot.20149 (2004).
31. Verdonk, M. L. *et al.* Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci* **44**, 793-806, doi:10.1021/ci034289q (2004).
32. Gatica, E. A. & Cavasotto, C. N. Ligand and decoy sets for docking to G protein-coupled receptors. *J Chem Inf Model* **52**, 1-6, doi:10.1021/ci200412p (2012).



33. Weiss, D. R., Bortolato, A., Tehan, B. & Mason, J. S. GPCR-Bench: A Benchmarking Set and Practitioners' Guide for G Protein-Coupled Receptor Docking. *J Chem Inf Model* **56**, 642-651, doi:10.1021/acs.jcim.5b00660 (2016).
34. Wallach, I. & Lilien, R. Virtual decoy sets for molecular docking benchmarks. *J Chem Inf Model* **51**, 196-202, doi:10.1021/ci100374f (2011).
35. Wang, L., Pang, X., Li, Y., Zhang, Z. & Tan, W. RADER: a RAPid DEcoy Retriever to facilitate decoy based assessment of virtual screening. *Bioinformatics* **33**, 1235-1237, doi:10.1093/bioinformatics/btw783 (2017).
36. Cleves, A. E. & Jain, A. N. Structure- and Ligand-Based Virtual Screening on DUD-E(+): Performance Dependence on Approximations to the Binding Pocket. *J Chem Inf Model*, doi:10.1021/acs.jcim.0c00115 (2020).
37. Bauer, M. R., Ibrahim, T. M., Vogel, S. M. & Boeckler, F. M. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0--a public library of challenging docking benchmark sets. *J Chem Inf Model* **53**, 1447-1462, doi:10.1021/ci400115b (2013).
38. Csizmadia, F. JChem: Java applets and modules supporting chemical database handling from web browsers. *J Chem Inf Comput Sci* **40**, 323-324, doi:10.1021/ci9902696 (2000).
39. Irwin, J. J. *et al.* Automated docking screens: a feasibility study. *J Med Chem* **52**, 5712-5720, doi:10.1021/jm9006966 (2009).
40. Powers, R. A., Morandi, F. & Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure* **10**, 1013-1023, doi:10.1016/s0969-2126(02)00799-2 (2002).

41. Eidam, O. *et al.* Fragment-guided design of subnanomolar beta-lactamase inhibitors active in vivo. *Proc Natl Acad Sci U S A* **109**, 17448-17453, doi:10.1073/pnas.1208337109 (2012).
42. Stein, R. M. *et al.* Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609-614, doi:10.1038/s41586-020-2027-0 (2020).
43. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
44. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992, doi:10.1371/journal.pone.0075992 (2013).
45. Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **50**, 572-584, doi:10.1021/ci100031x (2010).
46. Gschwend, D. A. & Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* **10**, 123-132, doi:10.1007/bf00402820 (1996).
47. Sterling, T. & Irwin, J. J. ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model* **55**, 2324-2337, doi:10.1021/acs.jcim.5b00559 (2015).
48. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* **39**, 2887-2893, doi:10.1021/jm9602928 (1996).

49. Oprea, T. I., Davis, A. M., Teague, S. J. & Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* **41**, 1308-1315, doi:10.1021/ci010366a (2001).
50. Neves, M. A., Totrov, M. & Abagyan, R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des* **26**, 675-686, doi:10.1007/s10822-012-9547-0 (2012).
51. Repasky, M. P. *et al.* Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J Comput Aided Mol Des* **26**, 787-799, doi:10.1007/s10822-012-9575-9 (2012).
52. Perryman, A. L., Santiago, D. N., Forli, S., Martins, D. S. & Olson, A. J. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein-ligand binding challenge. *J Comput Aided Mol Des* **28**, 429-441, doi:10.1007/s10822-014-9709-3 (2014).
53. Latti, S., Niinivehmas, S. & Pentikainen, O. T. Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *J Cheminform* **8**, 45, doi:10.1186/s13321-016-0158-y (2016).
54. Swift, R. V., Jusoh, S. A., Offutt, T. L., Li, E. S. & Amaro, R. E. Knowledge-Based Methods To Train and Optimize Virtual Screening Ensembles. *J Chem Inf Model* **56**, 830-842, doi:10.1021/acs.jcim.5b00684 (2016).
55. Zhou, Z., Felts, A. K., Friesner, R. A. & Levy, R. M. Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a

- diverse set of pharmaceutically relevant targets. *J Chem Inf Model* **47**, 1599-1608, doi:10.1021/ci7000346 (2007).
56. Brozell, S. R. *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* **26**, 749-773, doi:10.1007/s10822-012-9565-y (2012).
57. McGann, M. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des* **26**, 897-906, doi:10.1007/s10822-012-9584-8 (2012).
58. Spitzer, R. & Jain, A. N. Surflex-Dock: Docking benchmarks and real-world application. *J Comput Aided Mol Des* **26**, 687-699, doi:10.1007/s10822-011-9533-y (2012).
59. Ashtawy, H. M. & Mahapatra, N. R. Task-Specific Scoring Functions for Predicting Ligand Binding Poses and Affinity and for Screening Enrichment. *J Chem Inf Model* **58**, 119-133, doi:10.1021/acs.jcim.7b00309 (2018).
60. Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W. & Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J Mol Biol* **322**, 339-355, doi:10.1016/s0022-2836(02)00777-5 (2002).
61. Balius, T. E. *et al.* Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc Natl Acad Sci U S A* **114**, E6839-E6846, doi:10.1073/pnas.1703287114 (2017).
62. Murphy, R. B. *et al.* WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J Med Chem* **59**, 4364-4384, doi:10.1021/acs.jmedchem.6b00131 (2016).

63. Shoichet, B. K. Screening in a spirit haunted world. *Drug Discov Today* **11**, 607-615, doi:10.1016/j.drudis.2006.05.014 (2006).
64. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* **53**, 2719-2740, doi:10.1021/jm901137j (2010).

## Gloss to Chapter 4

While I was working on incorporating blurry GIST into the DOCK scoring function, Brian and I had discussed me taking on an applications project to identify novel ligands for G protein-coupled receptors. In the beginning of my fourth year, he pitched me a project on identifying type-selective ligands for the melatonin receptors, which were recently crystallized by Vadim Cherezov's lab at USC, but not yet published. This was part of the Illuminating the Druggable Genome (IDG) project, whose goal was to find ligands for "orphan receptors", proteins that had no known endogenous ligand.

There are two melatonin receptors in mammals, MT<sub>1</sub> and MT<sub>2</sub>, whose biological functions overlap, providing a need for selective molecules to disentangle their differences. Interestingly, both receptors do, in fact, have a known endogenous ligand – melatonin. However, while there were a handful of MT<sub>2</sub>-selective ligands, there were no reliable MT<sub>1</sub>-selective ligands in functional assays or *in vivo*, potentially explaining its inclusion in the IDG. Of those MT<sub>1</sub>-selective ligands, it wasn't immediately clear why they were selective. Complicating the situation further was that there were no obvious binding site differences of which we could take advantage. Of the 21 residues in the orthosteric sites, 20 of them are identical, and the difference is a valine to leucine mutation, a replacement that we were not confident the DOCK scoring function could capitalize on. With this in mind, we docked only to the MT<sub>1</sub> crystal structure and as usual, focused on chemical novelty for purchasing and testing, selecting molecules that did not look like known melatonin receptor ligands, and made different interactions with

the sites than those in the known ligands. If we couldn't find selective molecules, we wanted molecules that had interesting functional activity.

What we found was all of the above – MT<sub>1</sub>- and MT<sub>2</sub>-selectivity, inverse agonism, signaling bias, and potency at picomolar and low nanomolar concentrations. We successfully analogued one of the initial 15 docking hits, an MT<sub>2</sub>-selective inverse agonist, into both MT<sub>2</sub>-selective agonists and MT<sub>1</sub>-selective inverse agonists, and it was these MT<sub>1</sub>-selective inverse agonists that were taken *in vivo*. In a model of jet-lag, consistent with their *in vitro* characterization as inverse agonists, they behaved like the gold standard known nonselective antagonist, luzindole, at a ten-fold lower dose, increasing the number of days it takes mice to acclimate to a new light-dark schedule after a 6-hour advance in darkness. Surprisingly, in a circadian rhythm phase shift assay, the inverse agonists behave like the agonist melatonin, advancing the phase of the mouse internal circadian clock by 1.3-1.5 hours. This project was a glorious success and unveiled some amazing new biology that needs to be further elucidated. There are caveats, however. We do not know where the MT<sub>1</sub>-selectivity or the inverse agonism arises from. However, binding and functional assays with our new molecules with mutant receptors as done in the crystallography paper from the Roth and Cherezov labs, may help determine whether the docked poses are correct, explain which residues are involved, and whether on- and off-rates of the ligands determine the activity, which seems to explain some of the differences in affinity between the two receptors. Regardless, we were able to find completely novel molecules, never synthesized before to our knowledge, that have interesting biological effects, and that can be used to tease apart the pharmacological and functional differences of these two receptors.

## Chapter 4: Virtual discovery of MT receptor ligands to modulate circadian rhythms

Reed M. Stein<sup>1†</sup>, Hye Jin Kang<sup>†2</sup>, John D. McCorvy<sup>2,9†</sup>, Grant C. Glatfelter<sup>3,10†</sup>, Anthony J. Jones<sup>3</sup>, Tao Che<sup>2</sup>, Samuel Slocum<sup>2</sup>, Xi-Ping Huang<sup>2</sup>, Olena Savych<sup>4</sup>, Yurii S. Moroz<sup>5,6</sup>, Benjamin Stauch<sup>7,8</sup>, Linda C. Johansson<sup>7,8</sup>, Vadim Cherezov<sup>7,8</sup>, Terry Kenakin<sup>2</sup>, John J. Irwin<sup>1</sup>, Brian K. Shoichet<sup>1\*</sup>, Bryan L. Roth<sup>2\*</sup>, Margarita L. Dubocovich<sup>3\*</sup>

1. Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA

2. Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina 27599-7365, USA

3. Department of Pharmacology and Toxicology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo (SUNY), Buffalo, New York 14203, USA

4. Enamine Ltd., Chervonotkatska Street78, Kyiv 02094, Ukraine

5. National Taras Shevchenko University of Kyiv, Volodymyrska Street 60, Kyiv, 01601, Ukraine;

6. Chemspace, 7 Deer Park Drive, Suite M-5, Monmouth Junction, NJ 08852, USA

7. Bridge Institute, USC Michelson Center for Convergent Biosciences, University of Southern California, Los Angeles, California 90089, USA

8. Department of Chemistry, University of Southern California, Los Angeles, California 90089

9. Current Address: Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA



10. Current Address: Designer Drug Research Unit, National Institute on Drug Abuse Intramural Research Program, Baltimore, Maryland 21224, USA

†These authors contributed equally.

\* Corresponding authors. Email: [bshoichet@gmail.com](mailto:bshoichet@gmail.com); [bryan\\_roth@med.unc.edu](mailto:bryan_roth@med.unc.edu); [mdubo@buffalo.edu](mailto:mdubo@buffalo.edu)

The text of this chapter is adapted from:

Stein, R.M. \*; Kang, HJ\*; McCorvy, JD\*; Glatfelter, GC\*; Jones, AJ; Che, T; Slocum S; Huang, XP; Savych, O; Moroz, YS; Stauch, B; Johansson, LC; Cherezov, V; Irwin JJ; Shoichet BK; Roth, BL; Dubocovich, ML. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609-614, doi:10.1038/s41586-020-2027-0 (2020).

\*these authors contributed equally

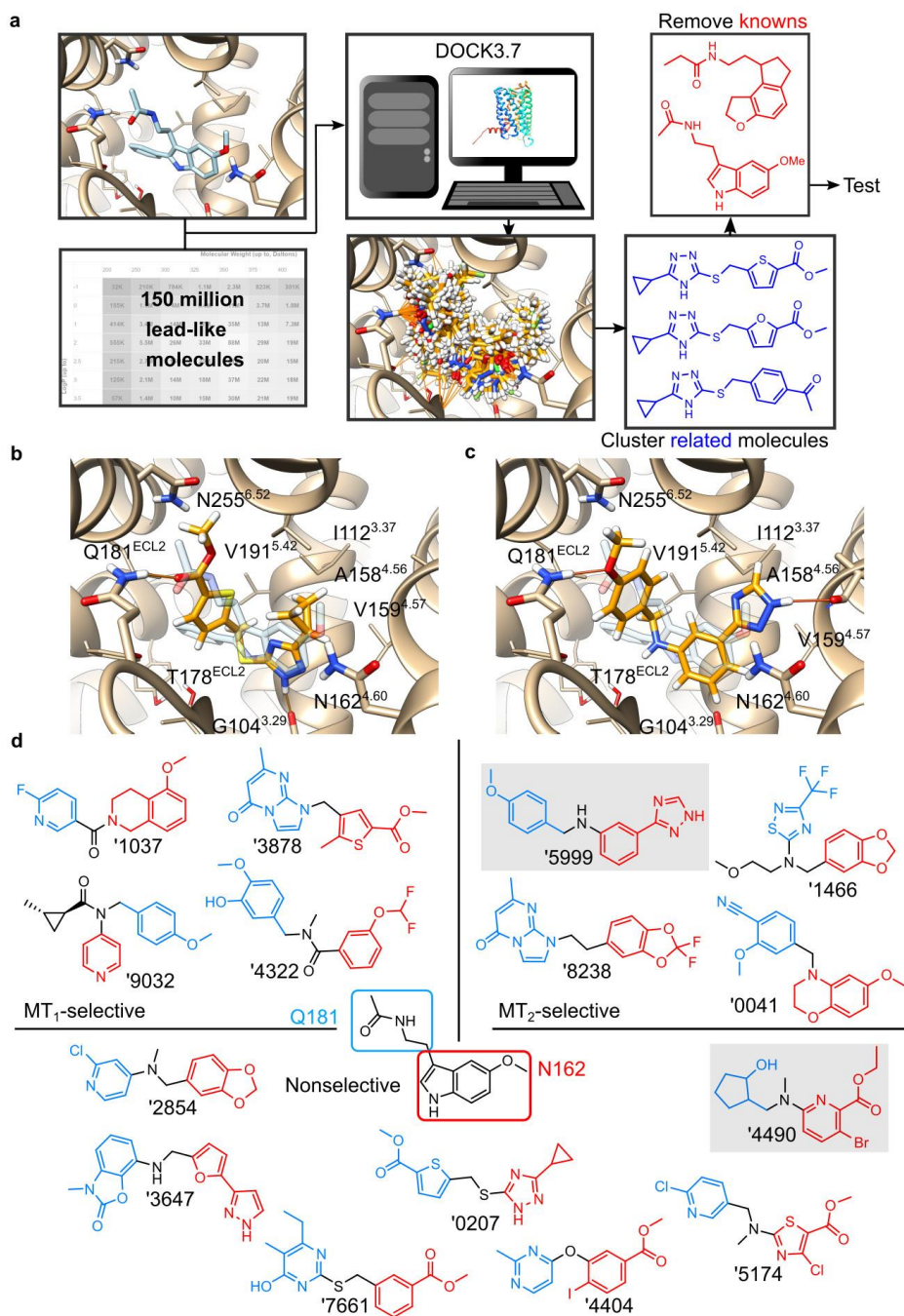
## 4.1 Summary Paragraph

The neuromodulator melatonin synchronizes circadian rhythms and related physiological functions *via* actions at two G protein-coupled receptors: MT<sub>1</sub> and MT<sub>2</sub>. Circadian release of high nighttime levels of melatonin from the pineal gland activates melatonin receptors in the suprachiasmatic nucleus of the hypothalamus, synchronizing physiology and behavior to the light-dark cycle<sup>1-4</sup>. The two receptors are established drug targets for aligning circadian phase in disorders of sleep<sup>5,6</sup> and depression<sup>7, 1-4,8,9</sup>. Despite their importance, few if any *in vivo* active MT<sub>1</sub> selective ligands have been reported<sup>2,8,10-12</sup>, hampering both the understanding of circadian biology and the development of targeted therapeutics. We docked over 150 million virtual molecules against an MT<sub>1</sub> crystal structure, prioritizing structural fit and chemical novelty. Thirty-eight high-ranking molecules were synthesized and tested, revealing ligands in the 470 pM to 6 μM range. Structure-based optimization led to two selective MT<sub>1</sub> inverse agonists, topologically unrelated to previously explored chemotypes, that were tested in mouse models of circadian behavior. Unexpectedly, the MT<sub>1</sub>-selective inverse agonists advanced the phase of the mouse circadian clock by 1.3-1.5 hrs when given at subjective dusk, an agonist-like effect eliminated in MT<sub>1</sub>- but not in MT<sub>2</sub>-knockout mice. This study illustrates opportunities for modulating melatonin receptor biology via MT<sub>1</sub>-selective ligands, and for the discovery of new, *in vivo*-active chemotypes from structure-based screens of diverse, ultra-large libraries.

## 4.2 Results

**Ultra-large library docking for new melatonin receptor ligands.** The recent determination of the MT<sub>1</sub> and MT<sub>2</sub> receptor crystal structures<sup>13,14</sup> afforded us the opportunity to seek new chemotypes with new functions, including MT<sub>1</sub>-selective ligands, by computational docking of an ultra-large make-on-demand library<sup>15</sup>, seeking molecules that complemented the main ligand binding (orthosteric) site of the receptor. Given the similar MT<sub>1</sub> and MT<sub>2</sub> sites, where 20 of 21 residues are identical, and the challenges of docking for selectivity<sup>16</sup>, we sought to prioritize new, high-ranking chemotypes from the docking screen, unrelated to known melatonin receptor ligands, expecting these to differentially interact with the two melatonin receptor types<sup>17-19</sup>.

We docked over 150 million “lead-like” molecules, characterized by favorable physical properties, from ZINC (<http://zinc15.docking.org>)<sup>15,20</sup>. These largely make-on-demand molecules have not been previously synthesized, but are usually accessible by two component reactions. Use of complex building blocks in these reactions biases toward diverse, structurally interesting molecules<sup>15,20</sup>. Each library molecule was sampled in an average of over 1.6 million poses (orientations x conformations) in the MT<sub>1</sub> orthosteric site<sup>13</sup> by DOCK3.7<sup>21</sup>, more than 72 trillion complexes for the library overall, scoring each for physical complementarity to the receptor site<sup>21</sup>. Seeking diversity, the top 300,000 scoring molecules were clustered by topological similarity, resulting in 65,323 clusters, and those that were similar to known MT<sub>1</sub> and MT<sub>2</sub> ligands from ChEMBL23<sup>22</sup> were eliminated (see Methods) (**Fig. 4.1, Table A.4.1**).



**Figure 4.1. Large library docking finds novel, potent melatonin receptor ligands.**  
**a**, Docking for new melatonin receptor chemotypes from the make-on-demand library.  
**b**, Docked pose of **'0207**, an hMT<sub>1</sub>/hMT<sub>2</sub> non-selective agonist with low nanomolar activity. **c**, Docked pose of **'5999**, an MT<sub>2</sub>-selective inverse agonist. In **b-c**, the crystallographic geometry of 2-phenylmelatonin is shown in transparent blue, for context. **d**, The initial 15 docking actives are shown, highlighting groups that correspond to melatonin's acetamide side chain (blue) and its 5-methoxy-indole (red) in their docked poses and receptor interactions. Shaded molecules are inverse agonists.

The best scoring molecules from each of the top 10,000 clusters were inspected for engagement with residues that recognize ligands in the MT<sub>1</sub> crystal structure<sup>13,14</sup>, and for new polar partners in the MT<sub>1</sub> site. In the docked complexes, these included hydrogen bonds with Q181<sup>ECL2</sup>, N162<sup>4.60</sup>, T178<sup>ECL2</sup>, N255<sup>6.52</sup>, and with the backbone atoms of A158<sup>4.56</sup>, G104<sup>3.29</sup>, and F179<sup>ECL2</sup>. Conformationally strained molecules and those with unsatisfied hydrogen bond donors were deprioritized<sup>23</sup>. Within the best-scoring clusters, all members were inspected and the one that best fit these criteria was prioritized. Ultimately, 40 molecules with ranks ranging from 16 to 246,721, or the top 0.00001% to top 0.1% of the over 150 million docked, were selected for *de novo* synthesis and testing. Of the 38 molecules successfully synthesized (a 95% fulfillment rate), 15 had activity at either or both of the human MT<sub>1</sub> and MT<sub>2</sub> receptors in functional assays (**Table A.4.1, Fig. 4.1**), a hit rate of 39% (number-active/number-physically-tested).

#### **In vitro pharmacology reveals new chemotypes with multiple functions.**

These active molecules included both agonists and inverse agonists, consistent with the emphasis on chemotype novelty (**Table A.4.1, Fig. 4.1**). This novelty is supported quantitatively by their low topological similarity to known melatonin receptor ligands<sup>24</sup>, and visually by comparison of the new ligands to their closest analogs among the knowns (**Table A.4.1**). The different chemotypes often engaged the same residues that recognize 2-phenylmelatonin in the crystal structures. Examples include the hydrogen-bond interactions with N162<sup>4.60</sup> made by the methoxy group of 2-phenylmelatonin, but in

the docked models by esters (**ZINC92585174**), pyridines (**ZINC151209032**), and benzodioxoles (**ZINC301472854**). Similarly, while 2-phenylmelatonin stacks an indole with F179<sup>ECL2</sup>, the docked ligands stack benzoxazines (**ZINC482850041**), thiophenes (**ZINC419113878**), and furans (**ZINC433313647**). While 2-phenylmelatonin hydrogen bonds with Q181<sup>ECL2</sup> via its acetamide, the docked ligands use esters or even pyridines (**Fig. 1**). The new ligands also dock to interact with new residues, including hydrogen bonds with T178<sup>ECL2</sup>, N255<sup>6.52</sup>, A158<sup>4.56</sup>, G104<sup>3.29</sup>, and F179<sup>ECL2</sup> (**Fig. 4.1b,c, Fig. A.4.3**).

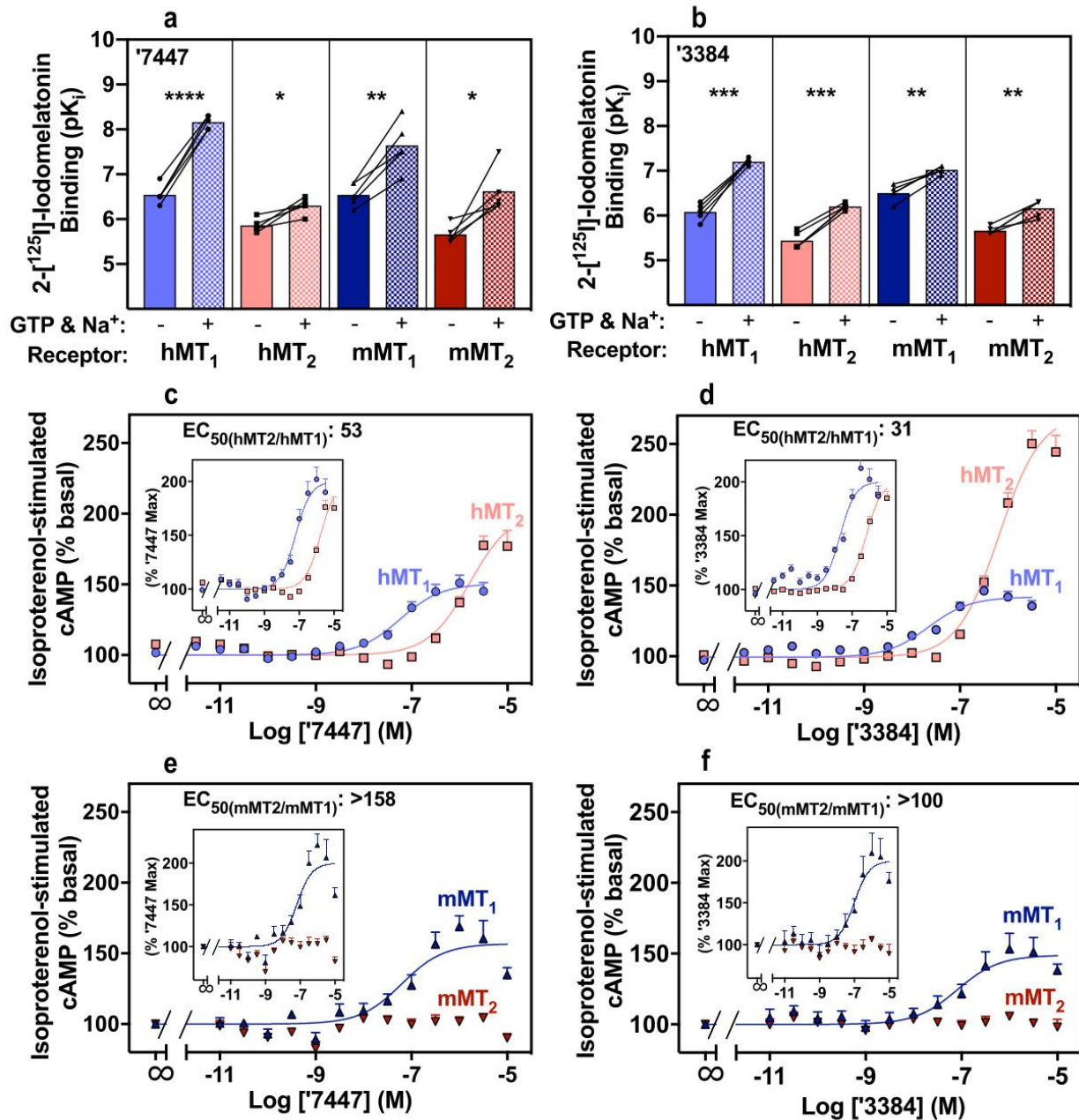
Consistent with docking against an agonist-bound MT<sub>1</sub> structure, four of the new ligands were MT<sub>1</sub>-selective agonists (**Fig. A.4.1a,b**), with EC<sub>50</sub> values in the 2 to 6 μM range, and without detectable MT<sub>2</sub> activity up to 30 μM: **'3878**, **'9032**, **ZINC353044322**, and **ZINC182731037**. Strikingly, **ZINC159050207**, although non-selective between the receptor types, is a 1 nM MT<sub>1</sub> agonist, among the most potent molecules found directly from a docking screen<sup>25-30</sup> (**Table A.4.1, Fig. 4.1b, Fig A.4.1c,d**). Admittedly, many ligands were just as active at the MT<sub>2</sub> receptor, or even selective for it (**Table A.4.1, Fig A.4.1**). Thus, whereas the initial docking against the MT<sub>1</sub> structure found new, potent chemotypes, and some of these were type selective, they were just as likely to prefer the MT<sub>2</sub> type as the MT<sub>1</sub> type. This attests to both the strengths and weaknesses of chemotype novelty as a strategy for compound prioritization, and to the need for further optimization.

We sought to improve twelve of these chemotype families, selecting analogs from the make-on-demand library. Several thousand such were docked into the MT<sub>1</sub> site (**Table A.4.2**) (see Methods). Of the 131 synthesized and tested, 94 analogs had activity at either or both MT<sub>1</sub> or MT<sub>2</sub> melatonin receptors at concentrations  $\leq 10 \mu\text{M}$  (**Table A.4.2, Fig. A.4.2**); of the twelve chemotype families, five saw improved potency. While this structure-based analoging could often find more potent ligands, their efficacy, selectivity, and bias were sensitive to small structural changes (**Fig. A.4.3**).

We were particularly interested in type-selective ligands with *in vivo* efficacy, as these are unreported in the field. We investigated two MT<sub>1</sub>-selective inverse agonists, **ZINC555417447** and **ZINC157673384**, and a selective MT<sub>2</sub> agonist, **ZINC128734226** (from here on referred to as **UCSF7447**, **UCSF3384** and **UCSF4226**, respectively), for their affinities (**Fig. 4.2, Fig. A.4.11**), *in vitro* signaling, pharmacokinetics (**Table A.4.3**), selectivity on mouse as well as the human receptors (hMT<sub>1</sub> and hMT<sub>2</sub>) (**Fig. 4.2, Figs. A.4.10 and A.4.11**), and for their efficacies in mouse models of circadian behavior (**Fig. 4.3, Figs. A.4.4-5, Fig. A.4.7**). As expected, **UCSF7447** and **UCSF3384** competed for 2-[<sup>125</sup>I]-iodomelatonin binding with higher affinity for the hMT<sub>1</sub> receptors. K<sub>i</sub> values in the absence of GTP, 304 nM and 938 nM, respectively, were improved by uncoupling G protein from the receptor by GTP addition, with K<sub>i</sub> values improving to 7.5 nM and 63 nM, respectively, supporting their status as inverse agonists (**Fig. 4.2a-b, Fig. A.4.6 and Fig. A.4.10**). Both **UCSF7447** and **UCSF3384** increased basal cAMP, also as expected for inverse agonists, with EC<sub>50</sub> values of 41 and 21 nM at hMT<sub>1</sub>, selectivity for hMT<sub>1</sub> over hMT<sub>2</sub> of 53- and 31-fold, and hMT<sub>1</sub> inverse agonist efficacies of 62% and 47%,

respectively (**Fig. 4.2c-d, Fig. A.4.6**). The third molecule, **UCSF4226** was an hMT<sub>2</sub>-selective agonist with an MT<sub>2</sub>/MT<sub>1</sub> selectivity of 54 in 2-[<sup>125</sup>I]-iodomelatonin binding assays and a selectivity of 91 in BRET assays; in isoproterenol-stimulated cAMP inhibition, the agonist had an EC<sub>50</sub> of 7.1 nM at hMT<sub>2</sub>, a value closely matched by an EC<sub>50</sub> of 6.3 nM in BRET assays (**Fig. A.4.11**). Upon intravenous administration in mice, the three molecules were CNS permeable, with brain/plasma ratios ranging from 1.4 to 3.0. Plasma half-lives ranged from 0.27 to 0.32 hours (**Table A.4.3**), similar to melatonin<sup>2</sup>. Against mouse MT<sub>1</sub> and MT<sub>2</sub> receptors (mMT<sub>1</sub>, mMT<sub>2</sub>) *in vitro*, the selectivity of the two inverse agonists improved over the human receptors being over 158 and over 100 times more selective for the mMT<sub>1</sub> receptor to increase basal cAMP with no activity observed against the mMT<sub>2</sub> receptor up to 10 μM for either compound (**Fig. 4.2e-f; Fig. A.4.10**). Conversely, while the agonist **UCSF4226** lost little activity on the mouse receptor, its selectivity for the mMT<sub>2</sub> receptor was much diminished (**Fig. A.4.11**). Accordingly, we moved forward to mouse *in vivo* experiments with the two selective MT<sub>1</sub> inverse agonists.





**Figure 4.2. Affinity, efficacy, and potency of MT<sub>1</sub>-selective inverse agonists**  
 (a,b) Affinity (pK<sub>i</sub>) of inverse agonists '7447 (a) and '3384 (b) by 2-[<sup>125</sup>I]-iodomelatonin competition for hMT<sub>1</sub>, hMT<sub>2</sub>, mMT<sub>1</sub>, and mMT<sub>2</sub> receptors stably expressed in CHO cells. Binding was measured in the absence and presence of 100 μM GTP, 1 mM EDTA.Na<sub>2</sub>, and 150 mM NaCl. GTP uncouples G proteins from melatonin receptors promoting inactive conformations<sup>31</sup> and higher affinity for inverse agonists; thus, the solid bars show higher affinity than the paired checker bars. Connected symbols represent pK<sub>i</sub>

values of individual determinations run in parallel.  $K_i$  values were derived from competition binding curves (see **Fig. A.4.10**). Bars represent the averages of five independent determinations. Statistical significance between  $pK_i$  averages were calculated by two-tailed paired student *t test* (*t*, *df* and *P* values under described under **Data Analysis in Methods**).

\**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001 when compared with corresponding  $pK_i$  averages values derived in the absence of GTP.

(c - f) Concentration-response curves on hMT<sub>1</sub>, hMT<sub>2</sub>, mMT<sub>1</sub>, and mMT<sub>2</sub> receptors transiently-expressed in HEK cells, monitoring isoproterenol-stimulated cAMP production with **'7447 c**: hMT<sub>1</sub> pEC<sub>50</sub>: 7.39 ± 0.10, E<sub>max</sub>: -62 ± 13%, n = 8; hMT<sub>2</sub> pEC<sub>50</sub>: 5.66 ± 0.10, E<sub>max</sub>: -84 ± 9%, n = 8, and **e**: mMT<sub>1</sub> pEC<sub>50</sub>: 7.20 ± 0.17, E<sub>max</sub>: -56 ± 5 %, n = 5; mMT<sub>2</sub> pEC<sub>50</sub>: n/d, n=5, E<sub>max</sub>: n/d, n = 5) and **d**: **'3384**: hMT<sub>1</sub>pEC<sub>50</sub>: 7.68 ± 0.09, E<sub>max</sub>: -47 ± 12%, n = 13; hMT<sub>2</sub> pEC<sub>50</sub>: 6.18 ± 0.04, E<sub>max</sub>: -153 ± 14 %, n = 12; and **f**: mMT<sub>1</sub> pEC<sub>50</sub>: 7.00 ± 0.22, E<sub>max</sub>: -49 ± 3 %, n = 5; and mMT<sub>2</sub> pEC<sub>50</sub>: n/d, E<sub>max</sub>: n/d, n = 5) treatment. Data for **'7447** and **'3384** was normalized to isoproterenol-stimulated basal activity. Inset graphs represent data normalized to maximal ligand effect. Data represent mean ± s.e.m. from the indicated number (*n*) of biologically independent experiments run in triplicate.

**UCSF7447 ('7447); UCSF3384 ('3384)**

### **In vivo pharmacology reveals new MT<sub>1</sub>-selective activities.**

We first examined the *in vivo* activity of the two MT<sub>1</sub>-selective inverse agonists in a mouse model of re-entrainment. In this “east-bound jet-lag” model, mice are subjected to an abrupt six-hour advance of the light-dark cycle and treated at the new dark onset for three consecutive days to assess re-entrainment rate. At 30 μg/mouse, the agonist melatonin accelerates re-entrainment to the new cycle, consistent with its use in the treatment of east-bound human jet-lag (**Fig. 4.3b**). Conversely, the prototypical non-selective antagonist/inverse agonist luzindole, administered at 300 μg/mouse, decelerates re-entrainment, measured by the number of days to adapt to the new dark onset, as expected for an inverse agonist<sup>43,32,33,34</sup>. The selective MT<sub>1</sub> inverse agonists **UCSF7447** and **UCSF3384**, dosed 30 μg/mouse (about 1 mg/Kg), also decelerated re-

entrainment (**Fig. 4.3a,b, Fig. A.4.4c,d,l**), phenocopying luzindole (encouragingly, at a 10-fold lower dose).

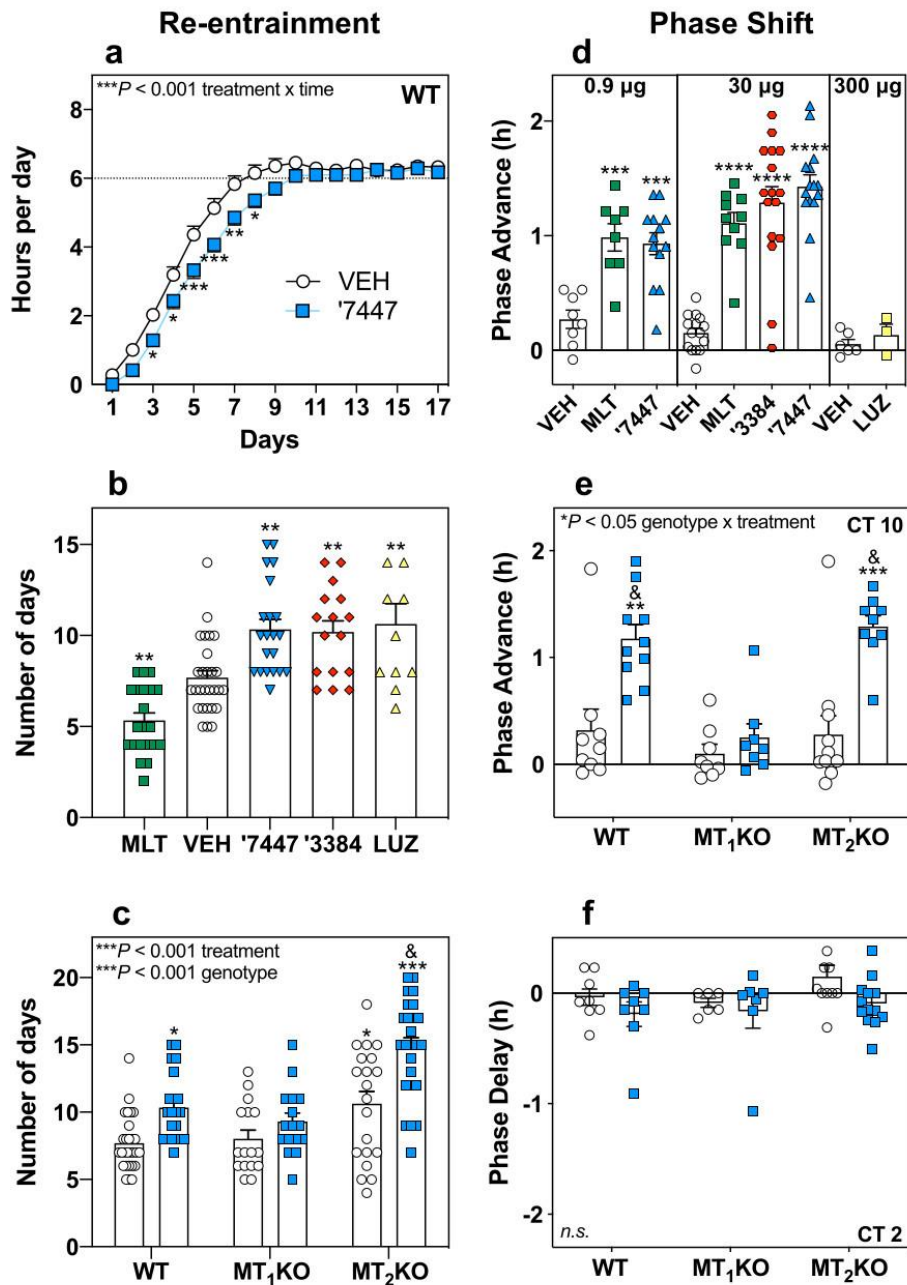
Superficially, the shared effect of decelerating re-entrainment by **UCSF7447**, **UCSF3384** (**Fig. 4.3a-c, Fig. A.4.4c,d,l**) and luzindole<sup>34</sup> might seem expected, as they all share the same function as melatonin receptor antagonists/inverse-agonists. However, luzindole is MT<sub>1</sub>/MT<sub>2</sub> non-selective, unlike **UCSF7447** and **UCSF3384**. Their phenocopying of luzindole suggests that deceleration of re-entrainment by all three molecules—slowing “jet-lag” accommodation—is mediated via the MT<sub>1</sub> receptor alone. Supporting this, the effect of **UCSF7447** was eliminated in an MT<sub>1</sub>KO mouse (**Fig. 4.3c, Fig. A.4.4h,i,m**), but not in an MT<sub>2</sub>KO mouse, where its effect was actually increased, adding to the deceleration afforded by deletion of the MT<sub>2</sub> receptor alone (**Fig. 4.3c, Fig. A.4.4j,k,n**).

The effect of the MT<sub>1</sub>-selective inverse agonists on circadian phase was even more unexpected. Here, we measured their effects on circadian phase by monitoring the running wheel activity onset of freely running mice in constant darkness<sup>35-37</sup> and administering them at subjective dusk (circadian time 10, CT 10). Both inverse agonists phase-advanced circadian wheel running rhythm onset, an effect characteristic of melatonin, the endogenous, non-selective agonist, and of non-selective agonist drugs like ramelteon<sup>38</sup> and agomelatine<sup>9,39</sup> (**Fig. 4.3d-f, Fig. A.4.5b-d,g,h**). Whereas MT<sub>1</sub>-selective inverse agonists have few if any precedents *in vivo*, we would have ordinarily expected the opposite effect of the agonist<sup>40,41</sup>, delaying rather than advancing circadian

phase. Instead, **UCSF7447** advanced the onset of activity by approximately 1 hour at 0.9  $\mu\text{g}/\text{mouse}$  (about 0.03 mg/Kg), an effect similar to that of melatonin at its  $\text{ED}_{50}$  (0.72  $\mu\text{g}/\text{mouse}$ )<sup>35</sup> (**Fig. 4.3d, Fig. A.4.5g,h**). At a higher dose (30  $\mu\text{g}/\text{mouse}$ , about 1 mg/Kg), both **UCSF7447** and **UCSF3384** advanced the onset of running wheel activity with an amplitude similar to melatonin<sup>35</sup> at this circadian time (CT 10). Intriguingly, whereas melatonin and ramelteon advance phase when dosed at dusk (CT 10), and delay phase when given at dawn (CT 2)<sup>36-38,42</sup>, **UCSF7447** did not affect phase at dawn (**Fig. 4.3f, Figure A.4.5r-w**), only working at dusk (**Fig. A.4.7a-c**).

The phenocopying of the non-selective agonist melatonin by the  $\text{MT}_1$ -selective inverse agonists, in shifting circadian phase, motivated us to investigate mechanism of action and the role of off-targets. Accordingly, both molecules, as well as the  $\text{hMT}_2$ -selective agonist **UCSF4226**, were tested against a panel of common off-targets (**Fig. A.4.8**). By radioligand competition, no activity was seen up to a concentration of 10  $\mu\text{M}$  for the new ligands. Against a panel of 318 GPCRs, activity was observed for only seven receptors when screened at a single concentration, none of which replicated in full concentration-response (**Fig. A.4.9**). Consistent with activity via the  $\text{MT}_1$  receptor, the advance in the onset of running wheel activity at dusk (CT 10) by **UCSF7447** was eliminated in  $\text{MT}_1\text{KO}$  mice but not in  $\text{MT}_2\text{KO}$  mice (**Fig. 4.3e, Fig A.4.5l-q**). These observations suggest that the  $\text{MT}_1$ -selective inverse agonists **UCSF7447** and **UCSF3384** are not only potent, with effects on phase shift for UCSF7447 at 0.9  $\mu\text{g}/\text{mouse}$  (about 0.03 mg/Kg) (**Fig 4.3d**) and efficacies resembling the long-established reagent luzindole in the jet-lag model at 10-fold lower doses, but that their unexpected

activity in circadian phase is via the MT<sub>1</sub> receptor. We note that the lack of precedence for this behavior reflects a lack of MT<sub>1</sub> selective inverse agonists to probe for it, something addressed by this study.



**Figure 4.3. MT<sub>1</sub>-selective inverse agonists behave as agonists and inverse agonists**

**a - b**, Inverse agonists '3384 and '7447 decelerate re-entrainment rate [a, VEH vs '7447 (30  $\mu$ g/mouse); mixed-effect two-way repeated measures ANOVA (treatment x time interaction:  $F_{16,735} = 3.39$   $P = 8.20 \times 10^{-6}$ ], and increase number of days to re-entrainment after 6 h advance of dark onset in the "east-bound jet-lag" paradigm [b, VEH vs. MLT, '3384, and '7447 (30  $\mu$ g/mouse) or LUZ (300  $\mu$ g/mouse); one-way

ANOVA ( $F_{4,92} = 16.97$   $P = 1.86 \times 10^{-10}$ ). **c**, Inverse agonist **'7447** targets  $MT_1$  receptors to increase number of days to re-entrainment [VEH (white) vs. **'7447** (blue; 30  $\mu\text{g}/\text{mouse}$ ); two-way ANOVA (treatment:  $F_{1,120} = 24.82$   $P = 2.14 \times 10^{-6}$ , genotype:  $F_{2,120} = 23.44$   $P = 2.55 \times 10^{-9}$ ]. **d**, Inverse agonists **'3384** and **'7447** phase advance circadian wheel activity onset in constant dark at CT 10 (dusk), resembling agonist melatonin [**left**: VEH vs. MLT or **'7447** (0.9  $\mu\text{g}/\text{mouse}$ ); one-way ANOVA ( $F_{2,26} = 13.60$   $P = 9.08 \times 10^{-5}$ ); **center**: VEH vs. MLT, **'3384** or **'7447** (30  $\mu\text{g}/\text{mouse}$ ); one-way ANOVA ( $F_{3,52} = 32.05$   $P = 7.15 \times 10^{-12}$ ); **right**: VEH vs LUZ (300  $\mu\text{g}/\text{mouse}$ ); two-tailed unpaired students  $t$  test ( $t = 0.92$   $df = 7$   $P = 0.39$ ]. **e**, The phase advance of wheel activity onset by **'7447** is mediated via the  $MT_1$  receptor at CT 10 (dusk) [VEH (white) vs. **'7447** (blue; 30  $\mu\text{g}/\text{mouse}$ ); two-way ANOVA (treatment x genotype interaction:  $F_{2,49} = 4.46$   $P = 0.0166$ ]. **f**, Inverse agonist **'7447**, unlike melatonin, did not phase delay in constant dark at CT 2 (dawn) [VEH (white) vs. **'7447** (blue; 30  $\mu\text{g}/\text{mouse}$ ); two-way ANOVA (treatment x genotype interaction:  $F_{2,49} = 0.384$   $P = 0.684$ ]. Panel **f** has 1 value not shown due to scale, but is included in the analysis (value = 0.91 h). Data shown represent mean  $\pm$  s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  for comparisons to WT VEH. & $P < 0.001$  for comparisons to  $MT_2$ KO VEH. Post-test analysis used Sidak's (**a**), Tukey's (**c**, **e**, **f**), or Dunnet's (**b** & **d**; all  $P < 0.05$ ). Details for all statistical analyses and reporting of  $n$  values for each condition (depicted as scatter dot plots where appropriate) are found in Methods (Statistics & Reproducibility). Vehicle (VEH), melatonin (MLT), luzindole (LUZ), **UCSF7447** (**'7447**), **UCSF3384** (**'3384**). All treatments were given via s.c. injection.

### 4.3 Discussion

From a large library docking screen emerged multiple new chemotypes for melatonin receptors (**Fig. 4.1**), with new signaling and new pharmacology. Three features of this study merit emphasis. **First**, docking a library of over 150 million diverse, make-on-demand molecules found ligands topologically unrelated to known melatonin receptor ligands, with picomolar and nanomolar activity on the melatonin receptors. **Second**, the chemical novelty of these molecules translated functionally, conferring melatonin receptor type selectivity. Whereas the deceleration of re-entrainment (jet-lag model) by the new inverse agonists resembled that of the classic non-selective antagonist/inverse agonist luzindole, their high selectivity for the  $MT_1$  receptor, and the chemical-genetic epistasis in the  $MT_1$ KO mouse, convincingly implicates the  $MT_1$  receptor in this response. Unexpectedly, the new inverse agonists

conferred an agonist-like effect in circadian phase shift experiments when administered at dusk, perhaps suggesting previously unknown signaling control for the MT<sub>1</sub> receptor in the SCN, which has known time of day dependent receptor mediated signaling pathways<sup>43</sup>. **Third**, these are the first MT<sub>1</sub>-selective inverse agonists active *in vivo*, with efficacy at doses as low as 0.9 µg/mouse in circadian phase shift. Their efficacy in modulating time-dependent circadian entrainment supports their potential as leads towards therapeutics in conditions and diseases affected by alterations in phase<sup>5-7,44</sup>.

Certain caveats bear airing. While we sought MT<sub>1</sub>-selective ligands, we found ligands for both melatonin receptor types, reflecting their conserved orthosteric sites. Indeed, rather than adopting a structure-based strategy for type selectivity, we simply focused on chemical novelty among the high-ranking docked molecules<sup>15,17</sup>. While the 39% docking hit rate was high, and the hits were potent, this likely reflects a site that is unusually well-suited to ligand binding: it is small, solvent-occluded, and largely hydrophobic. These high hit rates and potencies may not always translate to other targets<sup>45,46</sup>.

The key observations of this work should nevertheless be clear. From a structure-based screen of a diverse, 150 million compound virtual library sprang 15 new chemical scaffolds, topologically unrelated to known melatonin receptor ligands and synthesized *de novo* for this project. From their chemical novelty emerged new activities, including inverse agonists and ligands with melatonin receptor type-selectivity. The potency, brain exposure, and selectivity of these new ligands enable one to begin to disentangle the physiological role of the MT<sub>1</sub> receptor. Accordingly, we are making the MT<sub>1</sub>-selective inverse agonist **UCSF7447**, and the hMT<sub>2</sub> selective agonist **UCSF4226**, openly available to the community, as probe pairs coupled with a close



analog that has no measurable activity on the melatonin receptors (**Table A.4.4**). We note that only a small fraction of even the highest-ranking chemotypes from the docking were tested here; it is likely that hundreds-of-thousands of melatonin receptor ligands, representing tens-of-thousands of new chemotypes<sup>15</sup>, remain to be discovered from the make-on-demand library, which continues to grow (<http://zinc15.docking.org>). This study suggests that not only potent ligands may be revealed by docking such a library, but also that the new chemotypes explored can illuminate new *in vivo* pharmacology.

## References

1. Zisapel, N. New perspectives on the role of melatonin in human sleep, circadian rhythms and their regulation. *Br J Pharmacol* **175**, 3190-3199, doi:10.1111/bph.14116 (2018).
2. Dubocovich, M. L. *et al.* International Union of Basic and Clinical Pharmacology. LXXV. Nomenclature, classification, and pharmacology of G protein-coupled melatonin receptors. *Pharmacol Rev* **62**, 343-380, doi:10.1124/pr.110.002832 (2010).
3. Liu, J. *et al.* MT1 and MT2 Melatonin Receptors: A Therapeutic Perspective. *Annu Rev Pharmacol Toxicol* **56**, 361-383, doi:10.1146/annurev-pharmtox-010814-124742 (2016).
4. Dubocovich, M. L. Melatonin receptors: role on sleep and circadian rhythm regulation. *Sleep Med* **8 Suppl 3**, 34-42, doi:10.1016/j.sleep.2007.10.007 (2007).
5. Munday, K., Benloucif, S., Harsanyi, K., Dubocovich, M. L. & Zee, P. C. Phase-dependent treatment of delayed sleep phase syndrome with melatonin. *Sleep* **28**, 1271-1278, doi:10.1093/sleep/28.10.1271 (2005).
6. Rajaratnam, S. M. *et al.* Melatonin agonist tasimelteon (VEC-162) for transient insomnia after sleep-time shift: two randomised controlled multicentre trials. *Lancet* **373**, 482-491, doi:10.1016/S0140-6736(08)61812-7 (2009).
7. Lewy, A. J. *et al.* The phase shift hypothesis for the circadian component of winter depression. *Dialogues Clin Neurosci* **9**, 291-300 (2007).
8. Jockers, R. *et al.* Update on melatonin receptors: IUPHAR Review 20. *Br J Pharmacol* **173**, 2702-2725, doi:10.1111/bph.13536 (2016).

9. de Bodinat, C. *et al.* Agomelatine, the first melatonergic antidepressant: discovery, characterization and development. *Nat Rev Drug Discov* **9**, 628-642, doi:10.1038/nrd3140 (2010).
10. Descamps-Francois, C. *et al.* Design and synthesis of naphthalenic dimers as selective MT1 melatonergic ligands. *J Med Chem* **46**, 1127-1129, doi:10.1021/jm0255872 (2003).
11. Spadoni, G. *et al.* Bivalent ligand approach on N-{2-[(3-methoxyphenyl)methylamino]ethyl}acetamide: synthesis, binding affinity and intrinsic activity for MT(1) and MT(2) melatonin receptors. *Bioorg Med Chem* **19**, 4910-4916, doi:10.1016/j.bmc.2011.06.063 (2011).
12. Zlotos, D. P., Riad, N. M., Osman, M. B., Dodda, B. R. & Witt-Enderby, P. A. Novel difluoroacetamide analogues of agomelatine and melatonin: probing the melatonin receptors for MT1 selectivity. *MedChemComm* **6**, 1340-1344, doi:10.1039/C5MD00190K (2015).
13. Stauch, B. *et al.* Structural basis of ligand recognition at the human MT1 melatonin receptor. *Nature* **569**, 284-288, doi:10.1038/s41586-019-1141-3 (2019).
14. Johansson, L. C. *et al.* XFEL structures of the human MT2 melatonin receptor reveal the basis of subtype selectivity. *Nature* **569**, 289-292, doi:10.1038/s41586-019-1144-0 (2019).
15. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229, doi:10.1038/s41586-019-0917-9 (2019).

16. Weiss, D. R. *et al.* Selectivity Challenges in Docking Screens for GPCR Targets and Antitargets. *J Med Chem* **61**, 6830-6845, doi:10.1021/acs.jmedchem.8b00718 (2018).
17. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185-190, doi:10.1038/nature19112 (2016).
18. Huang, X. P. *et al.* Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **527**, 477-483, doi:10.1038/nature15699 (2015).
19. Lansu, K. *et al.* In silico design of novel probes for the atypical opioid receptor MRGPRX2. *Nat Chem Biol* **13**, 529-536, doi:10.1038/nchembio.2334 (2017).
20. Sterling, T. & Irwin, J. J. ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model* **55**, 2324-2337, doi:10.1021/acs.jcim.5b00559 (2015).
21. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992, doi:10.1371/journal.pone.0075992 (2013).
22. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res* **42**, D1083-1090, doi:10.1093/nar/gkt1031 (2014).
23. Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* **59**, 4103-4120, doi:10.1021/acs.jmedchem.5b02008 (2016).
24. Muchmore, S. W. *et al.* Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J Chem Inf Model* **48**, 941-948, doi:10.1021/ci7004498 (2008).

25. Katritch, V. *et al.* Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J Med Chem* **53**, 1799-1809, doi:10.1021/jm901647p (2010).
26. de Graaf, C. *et al.* Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J Med Chem* **54**, 8195-8206, doi:10.1021/jm2011589 (2011).
27. Mannel, B. *et al.* Structure-Guided Screening for Functionally Selective D2 Dopamine Receptor Ligands from a Virtual Chemical Library. *ACS Chem Biol* **12**, 2652-2661, doi:10.1021/acscchembio.7b00493 (2017).
28. Kiss, R. *et al.* Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. *J Med Chem* **51**, 3145-3153, doi:10.1021/jm7014777 (2008).
29. Congreve, M. *et al.* Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J Med Chem* **55**, 1898-1903, doi:10.1021/jm201376w (2012).
30. Langmead, C. J. *et al.* Identification of novel adenosine A(2A) receptor antagonists by virtual screening. *J Med Chem* **55**, 1904-1909, doi:10.1021/jm201455y (2012).
31. Lefkowitz, R. J., Mullikin, D. & Caron, M. G. Regulation of beta-adrenergic receptors by guanyl-5'-yl imidodiphosphate and other purine nucleotides. *J Biol Chem* **251**, 4686-4692 (1976).
32. Adamah-Biassi, E. B., Stepien, I., Hudson, R. L. & Dubocovich, M. L. Effects of the Melatonin Receptor Antagonist (MT2)/Inverse Agonist (MT1) Luzindole on Re-entrainment of Wheel Running Activity and Spontaneous Homecage Behaviors in

- C3H/HeN Mice. *The FASEB Journal* **26**, 1042.1045-1042.1045, doi:10.1096/fasebj.26.1\_supplement.1042.5 (2012).
33. Dubocovich, M. L. Luzindole (N-0774): a novel melatonin receptor antagonist. *J Pharmacol Exp Ther* **246**, 902-910 (1988).
34. Browning, C., Beresford, I., Fraser, N. & Giles, H. Pharmacological characterization of human recombinant melatonin mt(1) and MT(2) receptors. *Br J Pharmacol* **129**, 877-886, doi:10.1038/sj.bjp.0703130 (2000).
35. Dubocovich, M. L., Yun, K., Al-Ghoul, W. M., Benloucif, S. & Masana, M. I. Selective MT2 melatonin receptor antagonists block melatonin-mediated phase advances of circadian rhythms. *FASEB J* **12**, 1211-1220, doi:10.1096/fasebj.12.12.1211 (1998).
36. Benloucif, S. & Dubocovich, M. L. Melatonin and light induce phase shifts of circadian activity rhythms in the C3H/HeN mouse. *J Biol Rhythms* **11**, 113-125, doi:10.1177/074873049601100204 (1996).
37. Burgess, H. J., Revell, V. L., Molina, T. A. & Eastman, C. I. Human phase response curves to three days of daily melatonin: 0.5 mg versus 3.0 mg. *J Clin Endocrinol Metab* **95**, 3325-3331, doi:10.1210/jc.2009-2590 (2010).
38. Rawashdeh, O., Hudson, R. L., Stepien, I. & Dubocovich, M. L. Circadian periods of sensitivity for ramelteon on the onset of running-wheel activity and the peak of suprachiasmatic nucleus neuronal firing rhythms in C3H/HeN mice. *Chronobiol Int* **28**, 31-38, doi:10.3109/07420528.2010.532894 (2011).

39. Van Reeth, O. *et al.* Comparative effects of a melatonin agonist on the circadian system in mice and Syrian hamsters. *Brain Res* **762**, 185-194, doi:10.1016/s0006-8993(97)00382-x (1997).
40. Ersahin, C., Masana, M. I. & Dubocovich, M. L. Constitutively active melatonin MT(1) receptors in male rat caudal arteries. *Eur J Pharmacol* **439**, 171-172 (2002).
41. Soares, J. M., Jr., Masana, M. I., Ersahin, C. & Dubocovich, M. L. Functional melatonin receptors in rat ovaries at various stages of the estrous cycle. *J Pharmacol Exp Ther* **306**, 694-702, doi:10.1124/jpet.103.049916 (2003).
42. Lewy, A. J. *et al.* The human phase response curve (PRC) to melatonin is about 12 hours out of phase with the PRC to light. *Chronobiol Int* **15**, 71-83 (1998).
43. Gillette, M. U. & Mitchell, J. W. Signaling in the suprachiasmatic nucleus: selectively responsive and integrative. *Cell Tissue Res* **309**, 99-107, doi:10.1007/s00441-002-0576-1 (2002).
44. Reid, K. J. *et al.* Familial advanced sleep phase syndrome. *Arch Neurol* **58**, 1089-1094, doi:10.1001/archneur.58.7.1089 (2001).
45. Kufareva, I., Gustavsson, M., Zheng, Y., Stephens, B. S. & Handel, T. M. What Do Structures Tell Us About Chemokine Receptor Function and Antagonism? *Annu Rev Biophys* **46**, 175-198, doi:10.1146/annurev-biophys-051013-022942 (2017).
46. Cooke, R. M., Brown, A. J., Marshall, F. H. & Mason, J. S. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug Discov Today* **20**, 1355-1364, doi:10.1016/j.drudis.2015.08.003 (2015).

47. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735-1747, doi:10.1006/jmbi.1998.2401 (1999).
48. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106**, 765-784, doi:10.1021/ja00315a051 (1984).
49. Carlsson, J. *et al.* Structure-based discovery of A2A adenosine receptor ligands. *J Med Chem* **53**, 3748-3755, doi:10.1021/jm100240h (2010).
50. Gallagher, K. & Sharp, K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys J* **75**, 769-776, doi:10.1016/S0006-3495(98)77566-6 (1998).
51. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
52. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res* **44**, D1054-1068, doi:10.1093/nar/gkv1037 (2016).
53. Tolmachev, A. *et al.* Expanding Synthesizable Space of Disubstituted 1, 2, 4-Oxadiazoles. *ACS combinatorial science* **18**, 616-624 (2016).
54. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat Struct Mol Biol* **22**, 362-369, doi:10.1038/nsmb.3014 (2015).



55. Kenakin, T., Watson, C., Muniz-Medina, V., Christopoulos, A. & Novick, S. A simple method for quantifying functional selectivity and agonist bias. *ACS Chem Neurosci* **3**, 193-203, doi:10.1021/cn200111m (2012).
56. Kenakin, T. Biased Receptor Signaling in Drug Discovery. *Pharmacol Rev* **71**, 267-315, doi:10.1124/pr.118.016790 (2019).
57. Longo, P. A., Kavran, J. M., Kim, M. S. & Leahy, D. J. Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol* **529**, 227-240, doi:10.1016/B978-0-12-418687-3.00018-5 (2013).
58. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215-220, doi:10.1038/nature11691 (2012).
59. Popovska-Gorevski, M., Dubocovich, M. L. & Rajnarayanan, R. V. Carbamate Insecticides Target Human Melatonin Receptors. *Chem Res Toxicol* **30**, 574-582, doi:10.1021/acs.chemrestox.6b00301 (2017).
60. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K<sub>1</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem Pharmacol* **22**, 3099-3108 (1973).
61. Sumaya, I. C., Masana, M. I. & Dubocovich, M. L. The antidepressant-like effect of the melatonin receptor ligand luzindole in mice during forced swimming requires expression of MT<sub>2</sub> but not MT<sub>1</sub> melatonin receptors. *J Pineal Res* **39**, 170-177, doi:10.1111/j.1600-079X.2005.00233.x (2005).
62. Dubocovich, M. L., Hudson, R. L., Sumaya, I. C., Masana, M. I. & Manna, E. Effect of MT<sub>1</sub> melatonin receptor deletion on melatonin-mediated phase shift of

circadian rhythms in the C57BL/6 mouse. *J Pineal Res* **39**, 113-120,  
doi:10.1111/j.1600-079X.2005.00230.x (2005).

## 4.5 Methods

### Molecular docking

The MT<sub>1</sub> receptor bearing nine thermostabilizing point mutations, as determined crystallographically<sup>13</sup>, was used in the docking calculations. To prepare the structure for docking, atoms of the co-crystallized ligand, 2-phenylmelatonin, were used to seed the matching sphere calculation in the orthosteric site; these spheres represent favorable positions for individual ligand atoms to dock; overall 45 spheres were used. DOCK3.7 orients flexibases of pre-calculated ligand conformations into the orthosteric site by overlaying atoms of each library molecule onto these matching spheres. The receptor structure was protonated by REDUCE<sup>47</sup> and assigned AMBER united atom charges<sup>48</sup>. For residues N162<sup>4,60</sup> and Q181<sup>ECL2</sup>, the partial atomic charges of the side chain amide was increased without changing residue net charge, as previously<sup>49</sup>. The volume of the low protein dielectric, which defines the boundary between solute and solvent in Poisson-Boltzmann electrostatic calculations, was extended out 1.9 Å from the protein surface using spheres calculated by SPHGEN. Scoring grids were pre-calculated by CHEMGRID for AMBER van der Waals potential, QNIFFT<sup>50</sup> for Poisson-Boltzmann-based electrostatic potentials, and SOLVMAP<sup>51</sup> for ligand desolvation.

The resulting potential grids and ligand matching parameters were evaluated for their ability to enrich known MT<sub>1</sub> ligands over property-matched decoys. Decoys share

the same physical properties as known ligands but are topologically dissimilar and so unlikely to bind. Thirty-one known MT<sub>1</sub> melatonin receptor ligands, both agonists and antagonists, were extracted from the IUPHAR database<sup>52</sup>, and 1550 property-matched decoys were generated using the DUD-E pipeline. Docking success was judged on the ability to enrich the known ligands over the decoys by docking rank, using adjusted logAUC; this is widely done in the field. We also ensured that molecules with extreme physical properties were not enriched, as can happen when only counter-screening against property-matched decoys. In particular, we wanted to ensure that neutral molecules were enriched over charged ones. The docking parameters were also judged on how well they reproduced the known ligands' expected binding modes and their ability to hydrogen-bond with N162<sup>4.60</sup> and Q181<sup>ECL2</sup>.

The “lead-like” subset of ZINC15 (<http://zinc15.docking.org>), characterized by favorable physical properties (e.g., with calculated octanol-water partition coefficients (cLopP)  $\leq 3.5$ , and with molecular weights  $\leq 350$ ), was then docked against the MT<sub>1</sub> orthosteric site, using DOCK3.7<sup>21</sup>. This library contained over 150 million molecules, mostly make-on-demand from the Enamine REAL set<sup>15</sup>. Of these, over 135 million molecules successfully docked, with over 36 million receiving a favorable score ( $< 0$  kcal/mol). An average of 3,445 orientations were calculated for each, and for each orientation, an average of 485 conformations were sampled. A simplex minimizer was used for rigid-body minimization on the best-scored pose for each ligand. Overall, about 72 trillion complexes were sampled and scored. The calculation time was 45,020 core hours, or 1.25 calendar days on 1,500 cores.

To reduce redundancy of the best-ranking docked molecules, the top 300,000 ranked molecules were clustered by ECFP4-based Tanimoto coefficient (Tc) of 0.5, and the best-scoring member was used to represent the cluster. The resulting 65,323 clusters were filtered for novelty by calculating ECFP4-based Tcs against >1,100 MT<sub>1</sub> and MT<sub>2</sub> receptor ligands from the ChEMBL23<sup>22</sup> database. Molecules with Tc  $\geq$  0.38 to known MT<sub>1</sub>/MT<sub>2</sub> ligands were not further pursued.

After filtering for novelty, the docked poses of the best-scoring members of each cluster were filtered by the proximity of their polar moieties to N162<sup>4.60</sup> or Q181<sup>ECL2</sup>, and manually inspected for favorable geometry and interactions. Of the best-scoring molecules so prioritized, all members of its cluster within the top 300,000 molecules were also inspected, and sometimes one of these was chosen if they exhibited more favorable poses or chemical properties. Ultimately, forty compounds were chosen for testing, thirty-eight of which were successfully synthesized. To our knowledge, none of these compounds has been previously available and we are unaware of reports of them being previously synthesized.

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

Compounds were synthesized using 72,000 qualified in stock building blocks and 130 well-characterized, two component reactions at Enamine. Historically, molecules have been synthesized in three to four weeks with an 85% fulfillment rate; in this project delivery time was six weeks, but with a 95% fulfillment rate for the 40 molecules

prioritized from the initial docking screen. Each reaction is tested for conditions including temperatures, completion time, and mixing<sup>53</sup>. Typically, compounds are made in parallel by combining reagents and solvents in a single vial in the appropriate conditions to allow the reaction to proceed to completion. The product-containing vial is filtered by centrifugation into a second vial to remove precipitate and the solvent is evaporated under reduced pressure; the product is then purified by HPLC. Identity and purity are assessed by LC/MS and, as appropriate, <sup>1</sup>H NMR. All compounds were shipped 90% pure or better, and the main three compounds **UCSF7447**, **UCSF3384** and **UCSF4226** were independently confirmed to be  $\geq 95\%$  pure by LC/MS in secondary confirmation analyses at a second lab (**Fig. A.3.12**).

### **Structure-based ligand optimization**

After experimental testing (below), 12 of the 15 active ligands from docking were prioritized for optimization, representing a range of activities and type selectivity (**Table A.4.2**). Several thousand analogs of these ligands, each bearing the same scaffold as the parent molecule and with  $T_c < 0.38$  to annotated melatonin receptor ligands, were selected from the ZINC database and docked to the MT<sub>1</sub> binding site, again using DOCK3.7. The resulting docked poses were manually evaluated for interactions with N162<sup>4.60</sup> or Q181<sup>ECL2</sup>, and 132 analogs were selected for *de novo* synthesis at Enamine, in two iterations. Of these, 131 were successfully synthesized, a >99% fulfillment rate.

## **Cell Culture**

HEK293T cells were maintained with complete Dulbecco's modified Eagle's medium (DMEM), supplemented by 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin G and 100 µg/ml streptomycin. Cells were maintained at 37°C in the presence of 5% CO<sub>2</sub>.

## **Tango arrestin recruitment assay**

MT<sub>1</sub> and MT<sub>2</sub> Tango constructs were designed and assays were performed as previously described<sup>54</sup>. Briefly, HTLA cells stably expressing TEV protease fused β-arrestin (kindly provided by Dr. Richard Axel) and tTA dependent luciferase reporter gene were transfected with MT<sub>1</sub> or MT<sub>2</sub> Tango construct. The next day, transfected cells were seeded into poly-L-lysine coated 384-well white clear bottom cell culture plates with DMEM containing 1% dialyzed FBS at a density of 20,000 cells per well in 40 µl for another six hours. Drug solution was prepared in the same media used for cell plating at 5X final concentration and 10 µl per well was added for overnight incubation. The next day, media and drug solutions were discarded and loaded with 20 µl per well of Bright-Glo reagent (Promega). Plates were incubated for 20 mins in the dark followed by being counting using SpectraMax luminescence reader (Molecular Device). Data were analyzed using GraphPad Prism 6.0.

## **cAMP assay**

MT<sub>1</sub> and MT<sub>2</sub> receptors were tested using Promega's split luciferase based GloSensor cAMP biosensor technology. HEK293T cells were plated in 15 cm cell culture dish (at a ~50% cell confluency) with DMEM supplemented with 10% dialyzed FBS, 2 mM L-glutamine, 100 units/ml penicillin G and 100 µg/ml streptomycin for 4-6 hour. Then, cells were co-transfected with 8 µg of construct which encodes either MT<sub>1</sub> or MT<sub>2</sub> (de-Tango-ized constructs) and 8 µg of Glosensor DNA. Next day, transfected cells were seeded into poly-L-lysine coated 384-well white clear bottom cell culture plates with complete DMEM supplemented with 1% dialyzed FBS at a density of 20,000 cells per well for another 24 h. The next day, cell medium was discarded and loaded with 20 µl of assay buffer (1× HBSS, 20 mM HEPES, pH 7.4, 0.1% BSA). To measure agonist activity of MT<sub>1</sub> or MT<sub>2</sub> receptor, 10 µl of test compound solution at 3X final concentration was added for 15 minutes followed by addition of 10 µl of luciferin/isoproterenol mixture (at a final concentration of 4 mM and 200 nM respectively) for another 15 mins for luminescence quantification. Then, plates were counted using SpectraMax luminescence reader (Molecular Device). Data were analyzed using GraphPad Prism 8 (Graphpad Software Inc., San Diego, CA).

## **Log(E<sub>max</sub>/EC<sub>50</sub>) calculation and ligand bias quantification**

The  $\Delta\text{Log}(E_{\text{max}}/EC_{50})$  was calculated with melatonin as a reference agonist for G protein and  $\beta$ -arrestin pathway, and the  $\Delta\Delta\text{Log}(E_{\text{max}}/EC_{50})$  was calculated between two pathways for each ligand<sup>55</sup>, as were corresponding bias plots<sup>56</sup>. The bias factor is unitless and defined as  $10^{\Delta\Delta\text{Log}(E_{\text{max}}/EC_{50})}$ .

### **GPCR-ome counter-screen**

Screening of compounds in the PRESTO-Tango GPCR-ome was accomplished as previously described<sup>54</sup> with several modifications. First, HTLA cells were plated in DMEM with 10% FBS and 10 U/mL penicillin-streptomycin. Next, the cells were transfected using an in-plate PEI method<sup>57</sup>. PRESTO-Tango receptor DNAs were resuspended in OptiMEM and hybridized with PEI prior to dilution and distribution to 384-well plates and subsequent addition to cells. After overnight incubation, drugs were added to cells without replacement of the medium. The remaining steps of the PRESTO-Tango protocol were followed as previously described. For those six receptors where activity was reduced to less than 0.5 fold of basal (RLU) or for the one receptor where basal signaling was increased greater than 3-fold of basal, assays were repeated in full dose-response. None of the seven confirmed, and we discount the apparent activity seen in the single-point assay.

### **Inhibition screen**

Binding assays were performed by the NIMH Psychoactive Drug Screening program as detailed previously<sup>58</sup>. Detailed binding assay protocols are available on-line at: <https://pdspdb.unc.edu/pdspWeb/content/UNC-CH%20Protocol%20Book.pdf>

### **BRET recruitment assay**

To measure G protein recruitment BRET assay, HEK293T cells were co-transfected in a 1:1:1:1 ratio of Gai3-RLuc, Gβ3, GFP2-Gy9, and hMT<sub>1</sub> or hMT<sub>2</sub> (de-Tango-ized constructs) respectively. After 24 hours, transfected cells were plated in



poly-L-lysine coated 96-well white clear bottom cell culture plates with DMEM containing 1% dialyzed FBS, 100 units/ml Penicillin G, and 100 µg/ml Streptomycin at a density of 40,000 cells in 200 µL per well and incubated overnight. The following day, media was removed and cells were washed once with 100 µL of assay buffer (1X HBSS, 20 mM HEPES, pH 7.4, 0.1% BSA). Then 60 µL of assay buffer was loaded per well followed by addition of 10 µL of the RLuc substrate, Coelenterazine 400a (Nanolight) at 5 µM final concentration for 5 mins. Drug stimulation was performed with the addition of 30 µL of 3X drug dilution of melatonin or **UCSF4226** in assay buffer supplemented with 0.01% (w/v) ascorbic acid per well and incubated at RT for another 5 mins. Both luminescence (400 nm) and fluorescent GFP2 emission (515 nm) were read for the plate for 1 second per well using Mithras LB940. The ratio of GFP2/RLuc was calculated per well and analyzed using “log (agonist) vs. response” in Graphpad Prism 8 (Graphpad Software Inc., San Diego, CA).

## **Radioligand Binding**

### Reagents and Ligands

2-[<sup>125</sup>I]-Iodomelatonin (SA: 2,200 ci, 81.4TBq/mmol) was purchased from Perkin Elmer (Shelton, CT, USA). Guanosine 5'-triphosphate sodium salt hydrate (GTP), melatonin and all other chemicals and reagents were obtained from Sigma-Aldrich (St. Louis, MO, USA).

### Compound Preparation

For receptor binding studies, **UCSF7447** was dissolved in 50% DMSO/50% ethanol for 13 mM stock solution, diluted 1/10 in 100% ethanol then 1/10 again in 50% ethanol/50% Tris-HCl buffer, pH 7.4 25 deg C. Both **UCSF3384** and **UCSF4226** were dissolved in 100% ethanol for 13 mM stock solutions and then diluted 1/10 in 50% ethanol/50% Tris-HCl buffer, pH 7.4. Further dilutions were done in the same Tris-HCl buffer.

### 2-[<sup>125</sup>I]-Iodomelatonin Competition Binding

CHO cells stably expressing FLAG-tagged recombinant hMT<sub>1</sub>, hMT<sub>2</sub>, mMT<sub>1</sub>, or mMT<sub>2</sub> melatonin receptors were grown in culture as monolayers in Ham's F12 media supplemented with fetal calf serum (10%), penicillin (1%; 10,000 I.U/ml)/streptomycin (5%; 10,000 µg/ml) in CO<sub>2</sub> at 37°C as described. Cells were grown for 4 days to 90–95% confluence, then washed with PBS (potassium phosphate buffer, 10 mM, pH 7.4), detached with PBS containing 0.25 M sucrose and 1 mM EDTA, and pelleted by centrifugation (1,700 x g, 5 min) as described<sup>59</sup>. Cell pellets were suspended and homogenized in control buffer (50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>; pH 7.4 at 25°C) and washed twice by centrifugation (17,000 x g, 15 min) in control or inactive conformation buffer (50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 100 µM GTP, 1 mM EDTA.Na<sub>2</sub>, 150 mM NaCl, pH 7.4 at 25°C) as described<sup>59</sup>. 2-[<sup>125</sup>I]-Iodomelatonin binding affinity was determined on membranes from CHO-hMT<sub>1</sub> (9.6 ± 0.3 µg protein/assay; B<sub>max</sub>: 1,154 ± 38 fmol/mg protein, n = 3), CHO-hMT<sub>2</sub> (15 ± 1 µg protein/assay; B<sub>max</sub>: 352 ± 19 fmol/mg protein, n = 3), CHO-mMT<sub>1</sub> (6.0 ± 0.022 µg protein/assay (n=3); B<sub>max</sub>: 1,705 ± 337 fmol/mg protein,

n = 3) and CHO-mMT<sub>2</sub> (6.4 + 0.7 µg protein/assay (n=3); Bmax: 725 + 93 fmol/mg protein, n = 3) cells. Ligand competition (10 pM to 100 µM) for 2-[<sup>125</sup>I]-iodomelatonin (104 ± 2 pM, n = 30) binding was performed in control or inactive conformation buffer in a total volume of 0.26 mL as described<sup>59</sup>. Assays were incubated for 1 hour at 25°C. Bound radioligand was separated from free by rapid vacuum filtration using glass microfiber filters (Whatman, Krackeler Scientific, Inc., Albany NY, USA) saturated in 0.5% polyethylenimine solution. Total radioactivity bound to the filters was determined on a gamma counter<sup>59</sup>.

### Data Analysis

K<sub>i</sub> values were calculated from IC<sub>50</sub> values using GraphPad PRISM™ 8.0 according to the Cheng-Prusoff equation<sup>60</sup>:  $K_i = IC_{50}/(1 + [L]/K_D)$  where L is the concentration of radioligand, K<sub>D</sub> is the dissociation constant of 2-[<sup>125</sup>I]-iodomelatonin in control or inactive conformation buffer for the hMT<sub>1</sub> (control K<sub>D</sub> = 116 pM; Inactive K<sub>D</sub> = 280 pM) and hMT<sub>2</sub> receptors (control K<sub>D</sub> = 80 + 13 pM; GTP K<sub>D</sub> = 461 + 159 pM), and for mMT<sub>1</sub> receptors (control K<sub>D</sub> = 87 + 6 pM; GTP K<sub>D</sub> = 201 + 67 pM) (n=3). Affinity shifts induced by G protein uncoupling were measured by subtracting pK<sub>i(inactive)</sub> from pK<sub>i(Control)</sub> (ΔpK<sub>i</sub>) and normalization by melatonin ΔpK<sub>i</sub> (CHO-hMT<sub>1</sub>: 1.19; CHO-hMT<sub>2</sub>: 0.41). Affinity shifts or lack thereof with G protein uncoupling indicate apparent efficacy<sup>31</sup> as ligands are classified as agonists (ΔpK<sub>i</sub> % MLT > 20 %), antagonists (ΔpK<sub>i</sub> % MLT < 20 %, > -20 %), or inverse agonists (ΔpK<sub>i</sub> % MLT < -20 %) accordingly. Individual data points were excluded from cell based when meeting the exclusion criteria for the outliers Grubbs test.

Data shown in **Fig. 4.2a and b** were analyzed by **two-tailed paired student t test**.

## ***In-vivo* Methods**

### Animals and Housing

Male and female C3H/HeN (C3H) wild-type (WT), MT<sub>1</sub> knockout (MT<sub>1</sub>KO), and MT<sub>2</sub> knockout (MT<sub>2</sub>KO) mice (average 6.28 months) used in this study were raised in our breeding colony at University at Buffalo. C3H/HeN mice homozygous for the MT<sub>1</sub> and MT<sub>2</sub> melatonin receptor gene deletion and their WT controls were generated from breeding pairs donated by Dr. S. M. Reppert (University of Massachusetts Medical School, Worcester, MA, USA) and backcrossed with C3H/HeN mice (Harlan, now Envigo, Indianapolis, IN, USA) for at least seven generation as described in detail<sup>61</sup>. Genotype was confirmed using tail samples at the end of each experiment and was verified periodically during the tenure of the colony. The strains of mice in our breeding colony were re-derived periodically by backcrossing with WT mice to reduce genetic drift.

Mice were group housed (3 - 5 per cage) with corncob bedding in polycarbonate translucent cages (30 X 19 cm) and maintained in a 14/10 light-dark (LD) cycle (Zeitgeber time 0 or ZT 0 corresponds to lights on and ZT 14 to lights off) in temperature and humidity controlled rooms with *ad libitum* access to food and water in the Laboratory Animal Facility at the University at Buffalo. Light levels were 200 - 300 lux at the level of the cage. Treatments and animal care performed in the dark were under a dim red safelight (15 watts, Kodak 1A filter) with illuminance of less than 3 lux<sup>36</sup>. All

experimental procedures using mice were conducted in accordance with guidelines set forth by the National Institutes of Health and approved by the University at Buffalo Institutional Animal Care and Use Committee.

### **Circadian Rhythm Measurement**

Circadian rhythm phase was determined for each mouse using the onset of running wheel activity defined as CT 12 (circadian time 12: onset of wheel activity). Running wheel activity was measured continuously via magnetic microswitches detecting wheel revolutions with a computer equipped with Clocklab data collection software™ (Actimetrics: Wilmette, IL). All actigraphy data was visualized and analyzed using ClockLab™ and MATLAB™ software. All mice were individually housed in cages (33 x 15 cm) equipped with running wheels in light-tight ventilated cabinets with controlled temperature and LD cycles (Phenome Technologies: Skokie, IL). Male and female mice were housed in separate cabinets for all experiments.

### **Phase Shift**

Changes in circadian phase induced by vehicle or drugs administered at various circadian times were assessed in WT, MT<sub>1</sub>KO, and MT<sub>2</sub>KO male and female C3H/HeN mice (3 to 8 months) using methods and protocols previously described<sup>35,36</sup>. Following a period of 14 days in a LD cycle mice were placed in constant dark (DD) beginning at Zeitgeber Time (ZT) 12 (dark onset) (ZT 0 = lights on). Mice were kept in DD (2 - 3 weeks) until a stable free-running phase of running wheel activity rhythm onset was established. Circadian times of treatment were predicted from best fit lines of running

wheel activity onsets for of running either pre (7 - 14 days) and post (7 - 14 days) treatment. Treatment times were within a 2-hour window at CT 2 (CT 1 - 3), CT 6 (CT 5 - 7), or CT 10 (CT 9 - 11). Mice were treated (0.1 ml/mouse, s.c.) with vehicle (30% ethanol saline, s.c.) or drugs (melatonin, **UCSF3384**, **UCSF7447**, at 0.9 µg and 30 µg/mouse or luzindole at 300 µg/mouse in vehicle) for three consecutive days at the appropriate circadian time under dim red light. Vehicle or drug treatments were repeated for 3 consecutive days at the selected circadian time following the three-pulse treatment protocol described<sup>36</sup>. Phase shifts were quantified using the best-fit lines for onsets of activity during pre and post treatment periods. Differences are characterized as phase delays (pre-treatment ahead of post treatment best fit line onset) or phase advances (post treatment ahead of pre-treatment best fit line onset) of running wheel activity onset rhythms.

### **Re-entrainment Experiments**

Male and female C3H/HeN WT, MT<sub>1</sub>KO, and MT<sub>2</sub>KO mice (3 to 6 months) were maintained under a 12:12 LD cycle for at least 2 weeks prior experimental manipulations to allow stable entrainment to dark onset before advance of the LD cycle. Actigraphy data was recorded as described above and all experimental protocols performed as described<sup>62</sup>. On the first day of treatment, the dark onset was advanced 6 hours. This resulted in a short night and mice were treated (0.1 ml / mouse s.c.) with vehicle (30% ethanol/70%saline, s.c.) or drugs (melatonin, **UCSF3384** or **UCSF7447** at 30 µg /mouse, or luzindole 300µg /mouse, in vehicle) for three consecutive days 10 - 30 minutes prior to the new dark onset. Post treatment, mice were given 14 - 20 days to re-

entrain running wheel activity onsets to the new dark onset. Using exported running wheel activity onsets from actograms, onset hours advanced each day were determined by subtracting this value each day from the average onset of running wheel activity for 3 days prior to treatment for each mouse. Further, using the data from this calculation combined with visualization of actograms, the number of days to reach stable re-entrainment was determined for each mouse.

#### *In vivo* Compound Preparation

All compounds were administered in fixed doses of either 0.9  $\mu\text{g}$  or 30  $\mu\text{g}$  subcutaneously (s.c.) in a volume of 0.1 ml per mouse, which are equivalent to doses of 0.03 or 1 mg/Kg for a 30 g mouse, respectively. Vehicle (VEH) was 30% ethanol/70% saline for all doses. Melatonin, UCSF7447, and UCSF3384 were prepared as stock solutions of 3 mg/mL (100% ethanol) using sonication and vortexing to ensure each drug was dissolved. Subsequently, stock solutions were diluted to 0.3 mg/mL (30  $\mu\text{g}$ /0.1 mL injection) or 0.009 mg/mL (0.9  $\mu\text{g}$ /0.1 mL injection) in vehicle. Luzindole was prepared similarly except the starting stock solution was 30 mg/mL in 100% ethanol and it was administered from a solution of 3 mg/mL (300  $\mu\text{g}$ /0.1 mL injection) in vehicle. Treatment dilutions were prepared just before use under sonication with intermittent vortexing between steps and used within 5 minutes of preparation.

#### Biostatistics and Reproducibility

All statistical analyses as described in further detail for each experiment were conducted using GraphPad Prism 8™ (La Jolla, CA). For phase shift and re-

entrainment experiments we determined statistical power a-priori ( $\alpha$  error probability = 0.05) based on data for a known effect size for melatonin in these paradigms (G-power 3.0.10)<sup>35,62</sup>. Individual actograms of wheel running activity were excluded from analysis based on the exclusion criteria described below, which was completed by at least two individuals blind to treatment before data analysis was started. For *re-entrainment actograms* exclusion criteria includes: a) low running, sporadic activity, significant missing wheel activity data and/or lack of entrainment prior to treatment; b) entrainment of running activity more than 1 h before or after the “old” or “new dark” onset; c) re-entrainment to new dark onset before administration of the third injection (entrainment to injection). For *phase shift actograms* exclusion criteria includes: a) low running, sporadic activity, missing wheel activity data and/or lack of free running activity rhythms; b) tau change > 0.3 h; c) at least 2 out of 3 injections occurred outside of the target pre-determined time-range for treatment (CT 1 - 3, 5 - 7, 10 - 12). All data sets were visualized for normality using QQ plots of predicted vs. actual residuals. Actigraphy data was generated for visualization blind to treatment prior to the quantification and statistical analysis stages. Comparisons for **Fig. 4.3a**, **Fig. A.4.4l, m, n** were made by mixed effect two-way repeated measures ANOVA (treatment x time) with Sidak’s post hoc test ( $P < 0.05$ ). Number of days to re-entrainment was compared via one-way ANOVA or two-way ANOVA for **Fig. 3b, c** with a Dunnett’s or Tukey’s post hoc test ( $P < 0.05$ ) respectively. Group comparisons for phase shift in **Fig. 4.3d (left & center)** & **Fig. A.4.7a - c** were made by one-way ANOVA ( $P < 0.05$ ) comparing hours shifted of circadian running wheel activity rhythm onsets (**Fig. 3.3d left**: 3 groups - vehicle, melatonin, **UCSF7447**; **Fig. 4.3d center**: 4 groups - vehicle, melatonin,



**UCSF7447, UCSF3384; Fig. A.4.7a - c:** 4 groups - vehicle, melatonin, **UCSF7447**, luzindole) accompanied with post-hoc analyses by Dunnett's to determine individual group differences compared to vehicle ( $P < 0.05$ ). **Fig. 4.3d (right)** comparisons between vehicle and luzindole were made via a two-tailed unpaired students  $t$  test ( $P < 0.05$ ). Data in **Fig. 4.3e & f** were compared via a two-way ANOVA (3 x 2: genotype x treatment) with Tukey's post hoc analyses ( $P < 0.05$ ). Either the overall interaction or the main effects were reported and interpreted for two-way ANOVAs as appropriate for assumptions of each data set. No sex differences in treatment effects were evident in any data set when assessed via two-way ANOVA or three-way ANOVA where appropriate; therefore, data were pooled between male and female mice for analyses described. The  $n$  values represent the number of individual mice per condition or independent biological replicates in each experiment. Each data set represents 2 - 4 independent experiments. The  $n$  value for each *in vivo* experiment is listed below:

Figure 3.3a, vehicle ( $n = 28$  mice<sup>#</sup>) vs. **UCSF7447** ( $n = 21$  mice<sup>#</sup>). Figure 3.3b, vehicle ( $n = 28$ ) vs. melatonin ( $n = 21$ ), **UCSF7447** ( $n = 21$ ), **UCSF3384** ( $n = 16$ ), or luzindole ( $n = 11$ ). Figure 3.3c, WT ( $n = 28$  vehicle;  $n = 21$  **UCSF7447**), MT<sub>1</sub>KO ( $n = 16$  vehicle;  $n = 16$  **UCSF7447**), and MT<sub>2</sub>KO ( $n = 20$  vehicle;  $n = 25$  **UCSF7447**). Figure 3.3d, (left panel) - vehicle ( $n = 8$ ) vs. melatonin ( $n = 8$ ) or **UCSF7447** ( $n = 13$ ). Figure 3.3d, (center panel) - vehicle ( $n = 15$ ) vs. melatonin ( $n = 10$ ), **UCSF3384** ( $n = 16$ ), or **UCSF7447** ( $n = 15$ ). Figure 3.3d, (right panel) - vehicle ( $n = 6$ ) vs luzindole ( $n = 3$ ). Figure 3.3e, WT ( $n = 9$  vehicle;  $n = 10$  **UCSF7447**), MT<sub>1</sub>KO ( $n = 8$  vehicle;  $n = 8$  **UCSF7447**), and MT<sub>2</sub>KO ( $n = 11$  vehicle;  $n = 9$  **UCSF7447**). Figure 3.3f, WT ( $n = 8$  vehicle;  $n = 8$  **UCSF7447**), MT<sub>1</sub>KO ( $n = 6$  vehicle;  $n = 7$  **UCSF7447**), and MT<sub>2</sub>-KO ( $n = 10$  vehicle;  $n = 13$  **UCSF7447**).

Fig. A.3.4h, C3H WT - vehicle ( $n = 28$  mice<sup>#</sup>) vs. **UCSF3384** ( $n = 16$  mice<sup>#</sup>).

Fig. A.3.4i, C3H MT<sub>1</sub>KO - vehicle ( $n = 16$  mice<sup>#</sup>) vs. **UCSF7447** ( $n = 16$  mice<sup>#</sup>). Fig.

A.3.4j, C3H MT<sub>2</sub>KO - vehicle ( $n = 21$  mice<sup>#</sup>) vs. **UCSF7447** ( $n = 25$  mice).

Fig. A.3.7a, CT 2 - vehicle ( $n = 3$ ), melatonin ( $n = 3$ ), luzindole ( $n = 6$ ), or **UCSF7447** ( $n$

= 3). Fig. A.3.7b, CT 6 - vehicle ( $n = 8$ ), melatonin ( $n = 4$ ), luzindole ( $n = 9$ ), or

**UCSF7447** ( $n = 9$ ). Fig. A.3.7c, CT 10 - vehicle ( $n = 6$ ), melatonin ( $n = 8$ ), luzindole ( $n =$

3), or **UCSF7447** ( $n = 4$ )

### Pharmacokinetics

Pharmacokinetic experiments were performed by Sai Life Sciences Limited (Hyderabad, India). Plasma pharmacokinetics and brain distribution for **UCSF7447**, **UCSF3384**, and **UCSF4226** were investigated following a single intravenous dose of 2 mg/kg in nine male C57BL/6 mice. Each compound was formulated in 5% N-methyl pyrrolidone, 5% Solutol HS-15, and 90% normal saline. Blood samples (approximately 60  $\mu$ L from each of three mice) were collected under light isoflurane anesthesia from retro orbital plexus at 0.08, 0.25, 0.5, 1, 2, 4, 8, 12, and 24 hr. Immediately after collection, plasma was harvested by centrifugation and stored at -70°C until analysis. For blood collected at 0.5, 4, and 24 hr, animals were euthanized with excess CO<sub>2</sub> asphyxiation and brain samples were collected and homogenized in ice-cold phosphate buffer saline (pH-7.4). Total homogenate volume was three times the brain weight.

All samples were processed for analysis by protein precipitation using acetonitrile and analyzed with fit-for-purpose LC/MS/MS method (Lower limit of quantification = 2.01

ng/mL for plasma and 6.03 ng/g for brain for **UCSF7447**, 5.01 ng/mL for plasma and 3.00 ng/g for brain for **UCSF3384**, 1.01 ng/mL for plasma and 6.09 ng/g for brain for **UCSF4226**). The non-compartmental analysis module in Phoenix WinNonlin® (Version 7.0) was used to assess the pharmacokinetic parameters. Maximum concentration ( $C_{max}$ ) and time to reach maximum concentration ( $T_{max}$ ) were measured. The areas under the concentration time curve ( $AUC_{last}$  and  $AUC_{inf}$ ) and elimination half-life was calculated by the linear trapezoidal rule. The terminal elimination rate constant,  $k_e$ , was determined by regression analysis of the linear terminal portion of the log plasma concentration-time curve. The terminal half-life ( $T_{1/2}$ ) was estimated as  $0.693/k_e$ .

**Code Availability:** 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 at <http://blaster.docking.org/>

**Data Availability Statement:** Probe pairs (two similar ligands with and without activity) of inverse agonists selective for  $MT_1$  and agonists selective for  $hMT_2$  are available by arrangement with Sigma (Table A.3.4). The identities of the compounds docked in this study are freely available from the ZINC database, <http://zinc15.docking.org>, and active compounds may be purchased from Enamine. Figures with associated raw data include: Fig. 4.1, Tables A.4.1&2, Figs. A.4.1&2, Table A.4.1, for which further data are included in Table A.4.5 (compound purity information); Fig. A.3.3, for which bias information is included in Table A.4.6; Fig. 4.2, for which GPCRome screening, concentration-response curves, competition binding, and LC/MS

data is included in Figs. A.4.1-5; Fig. 4.3, for which further data is included in Figs. A.4.4-5; Fig. A.4.7.

#### **4.6 Acknowledgements.**

Supported by the US NIH awards U24DK1169195 (to BLR & BKS), R35GM122481 (to BKS), the NIMH Psychoactive Drug Screening Contract (to BLR), GM133836 (to JJI), ES023684 (to MLD), UL1TR001412 and KL2TR001413 (to the University at Buffalo), PhRMA Foundation Fellowship (73309 to AJJ), Jacobs School of Medicine and Biomedical Sciences unrestricted funds (to MLD), R35GM127086 (to VC), EMBO ALTF 677-2014 (to BS), HFSP long-term fellowship LT000046/2014-L (to LCJ), postdoctoral fellowship from the Swedish Research Council (to LCJ), and the National Science Foundation (NSF) BioXFEL Science and Technology Center 1231306 (to BS & VC). We would like to thank Dr. Gregory Wilding from the Biostatistics, Epidemiology and Research Design (BERD) Core of the Clinical and Translational Science Institute at the University at Buffalo, for statistical advice regarding analyses of *in-vivo* data.

#### **4.7 Author Contributions.**

BKS, BLR, and MLD conceived the study. RMS performed the docking and structure-based optimization. JDM & HJK performed the initial binding and functional assays and analysis, assisted by TC, while AJJ performed the 2-[<sup>125</sup>I]-iodomelatonin and GTP-perturbation assays. SS performed the profiling studies. GCG performed the *in vivo* mouse pharmacology experiments and all animal husbandry. YSM and OS directed the compound synthesis, purification and characterization. BS, LCJ, VC, BLR,

XPH, JDM determined and validated the structures of the MT<sub>1</sub> and MT<sub>2</sub> receptor types, and made them available before publication. JJI created the ultra-large libraries. BLR supervised the pharmacology studies; BKS supervised the docking and compound optimization; MLD supervised the binding studies and the *in vivo* mouse circadian rhythms experiments. MLD & GCG designed all in vivo experiments. RMS, BKS, MLD, GCG, JDM, HJK, and BLR wrote the paper with contributions from other authors.

**Competing Financial Interests:** B.K.S. and J.J.I. are founders of a company, BlueDolphin LLC, that works in the area of molecular docking. All other authors declare no competing interests.

## Gloss to Chapter 5

With the success of docking to the melatonin receptors, Brian pitched me and Chase Webb, a new graduate student, the challenge of finding novel ligands for the CB<sub>1</sub> cannabinoid receptor, in collaboration with the Skiniotis and Kobilka labs at Stanford and Roth lab at UNC Chapel Hill. CB<sub>1</sub> is the target of phytocannabinoids like THC, the main psychoactive ingredient in marijuana, as well as cannabidiol, endocannabinoids like the lipid-based anandamide and 2-arachidonoyl glycerol, as well as dangerous synthetic cannabinoids like “Spice”. Though there were crystal structures for the CB<sub>1</sub> receptor, the Skiniotis lab had just solved the structure in complex with the G protein by cryoEM, and the initial goal was to determine whether we could find novel ligands from a cryoEM structure. Chase and I parameterized the system and had the first round of molecules tested in the summer of 2018, with Sam Slocum and XP Huang from the Roth lab finding that none of the molecules had reproducible potency in the PRESTO-Tango assay. We were skeptical of the results, but Brian suggested that they may have reflected our focus on buying lead-like ( $MW \leq 350$ ,  $cLogP \leq 3.5$ ) molecules for this lipid receptor that typically binds large, greasy molecules. We planned for a second go at the receptor, but it would take time to build these kinds of molecules as lead-like molecules are prioritized for building in ZINC15. Thus, this project was put on hold for a few months for these large, greasy molecules to be built.

After enough large, greasy molecules were built, the cryoEM structure of CB<sub>1</sub> in complex with the G protein was officially released in *Cell*, though this structure was different from the one we initially received from the Skiniotis lab. In this published

structure, roughly 40 of the residues were incompletely modeled (“stubbed”), with several of these residues in the binding site. With Tia Tummino, a new graduate student in the lab, we decided to take advantage of the crystal structures previously published and prioritize finding analgesics as part of a new focus in the lab, as CB<sub>1</sub> may be a promising therapeutic target for pain. After parameterizing the crystal structure for docking, we focused on molecules in a higher molecular weight and cLogP property space. In this second round, XP found similar results as before, irreproducible Tango and now, GloSensor assay curves. Given that we’ve had difficulties with lipid receptors previously, we thought that these data may be due to high nonspecific binding and the “stickiness” of the receptor. We were also worried that the data from the Roth lab were problematic, with a lack of reproducibility that they usually achieve. Additionally, the control molecules exhibited large variance, and there were expression problems with the receptors. We therefore turned to the Makriyannis lab at Northeastern, experts in cannabinoid binding assays. They tested our second round of molecules and found that 8 of the 46 molecules may be high affinity molecules. This project is still ongoing, as we now have 12 more molecules that they haven’t tested from the second round. Additionally, the 8 potential hits need dose response curves, but we are planning to do analog-by-catalog, as well as determine why these molecules aren’t picked up in Tango or GloSensor assays, so we can get functional data, and relate this to pain phenotypes. We may even re-purchase molecules from the first round of docking and determine if these hit the receptor in the binding assays.

## **Chapter 5: Large-scale docking on the CB<sub>1</sub> cannabinoid receptor**

Reed M. Stein<sup>1</sup>, Tia Tummino<sup>1</sup>, Chase Webb<sup>1</sup>, Xi-Ping Huang<sup>2</sup>, Samuel Slocum<sup>2</sup>,  
Christos Iliopoulos-Tsoutsouvas<sup>3</sup>, Georgios Skiniotis<sup>4</sup>, Brian K. Kobilka<sup>4</sup>, Alexandros  
Makriyannis<sup>3</sup>, Bryan L. Roth<sup>2</sup>, Brian K. Shoichet<sup>1</sup>

1. Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA
2. Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
3. Center for Drug Discovery, Department of Pharmaceutical Sciences; Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA
4. Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA



## 5.1 Abstract

*Cannabis*, whose psychoactive constituent  $\Delta^9$ -tetrahydrocannabinol (THC) targets the CB<sub>1</sub> cannabinoid receptor, has been used recreationally and medicinally for millennia<sup>1</sup>. Activation of CB<sub>1</sub>, one of the most abundant G protein-coupled receptors in the central nervous system, by cannabinoids is implicated in analgesic<sup>2</sup>, anxiolytic<sup>3</sup>, anti-obesity<sup>4,5</sup>, and anti-nausea<sup>6</sup> effects. Regardless, the usage of cannabinoids as therapeutics has been limited by their psychotropic effects, memory and cognition impairment, motor disturbances, as well as legislative barriers<sup>7,8</sup>. Here, we performed two virtual screens with the goal of identifying agonists to treat neuropathic pain that would lack these negative side effects. We initially performed a virtual screen of more than 225 million lead-like molecules to a CB<sub>1</sub> cryoEM prioritizing those molecules that favorably complemented the orthosteric site, and that were chemically unrelated to known cannabinoids. Of these compounds, 55 molecules were synthesized and tested, revealing no molecules that were functionally active. We then turned to a CB<sub>1</sub> crystal structure, and docked over 74 million large, greasy molecules, with 58 molecules synthesized and tested. Though none were reproducibly active in functional assays, 8 of 46 tested in radioligand displacement assays exhibited high affinity. Re-testing of all 113 molecules, followed by dose-response curves are currently underway, with a goal towards structure-based optimization of these hits, and *in vivo* testing of analgesia.

## 5.2 Introduction.

The usage of cannabinoids for therapeutic applications has been riddled with controversy, as well as seemingly more effective, and less negative side-effect-inducing alternatives<sup>9,10</sup>. Widespread prohibition in the early 20<sup>th</sup> century resulted in the termination of essentially all research on cannabis as a therapeutic, and it was only the popularity of its recreational use during the 1960s that spurred a newfound interest in its research, with researchers identifying  $\Delta^9$ -tetrahydrocannabinol (THC) as the main psychoactive component of cannabis in 1964<sup>1</sup>. It wasn't until 1990 that researchers identified the receptor responsible, the CB<sub>1</sub> cannabinoid receptor<sup>11</sup>, which was followed by the characterization of the homologous CB<sub>2</sub> cannabinoid receptor<sup>12</sup>, both G protein-coupled receptors. There is significant interest in using cannabinoids as therapeutics for multiple indications such as nausea, anxiety, obesity, multiple sclerosis, seizures, and pain, and there are currently three marketed synthetic cannabinoids: two for treating chemotherapy-induced nausea and one for treating neuropathic pain and multiple sclerosis symptoms<sup>8,13</sup>. However, despite these potential avenues for treatment, the field of cannabis research is riddled with inconclusive results regarding the efficacy of cannabinoids due to variability in research methods. Similarly, cannabinoids are plagued by negative side effects, including psychoactivity, respiratory and cardiovascular disorders, addiction, psychosis, mood disorders, and suicidal ideation<sup>14-17</sup>. Researchers have proposed various strategies for reducing these negative side effects including the development of peripherally restricted CB<sub>1</sub> agonists for neuropathic pain<sup>2,18-20</sup>. Additionally, ajulemic acid, a synthetic analog of THC, activates both CB<sub>1</sub> and CB<sub>2</sub> receptors, and has been shown to be effective in reducing chronic neuropathic

pain, while showing no psychotropic effects or dependency<sup>21</sup>, suggesting that molecules that target the orthosteric site can maintain analgesic effects with no negative side effects. However, the high lipophilicity of ajulemic acid and related phytocannabinoids limits their optimization as drug candidates. Here, we attempt to identify novel cannabinoids in drug-like space that can sidestep these negative side effects and treat neuropathic pain.

### 5.3 Results

With the recent determination of crystal and cryoEM structures of both cannabinoid receptors<sup>22-26</sup>, we sought previously undescribed chemotypes with new functions by docking an ultralarge make-on-demand library<sup>27</sup> to the orthosteric site of the CB<sub>1</sub> receptor. We prioritized high-ranking chemotypes that were unrelated to known cannabinoid receptor ligands with the hope that these new chemotypes would interact differently with the CB<sub>1</sub> receptor, conferring signaling properties with new biological effects<sup>28-30</sup>.

In the first screen, we docked more than 225 million 'lead-like' molecules, which are characterized by favorable calculated octanol-water partition coefficients (cLogP ≤ 3.5) and molecular masses (MW ≤ 350) from ZINC (<http://zinc15.docking.org>). Each library molecule was sampled in an average of more than 1.4 million poses (orientations x conformations) in the CB<sub>1</sub> orthosteric site using DOCK3.7<sup>31</sup>, with a total of 123 trillion complexes being generated and scored for complementarity to the site. The top 300,000 molecules were clustered by topological similarity, resulting in 51,365 clusters,

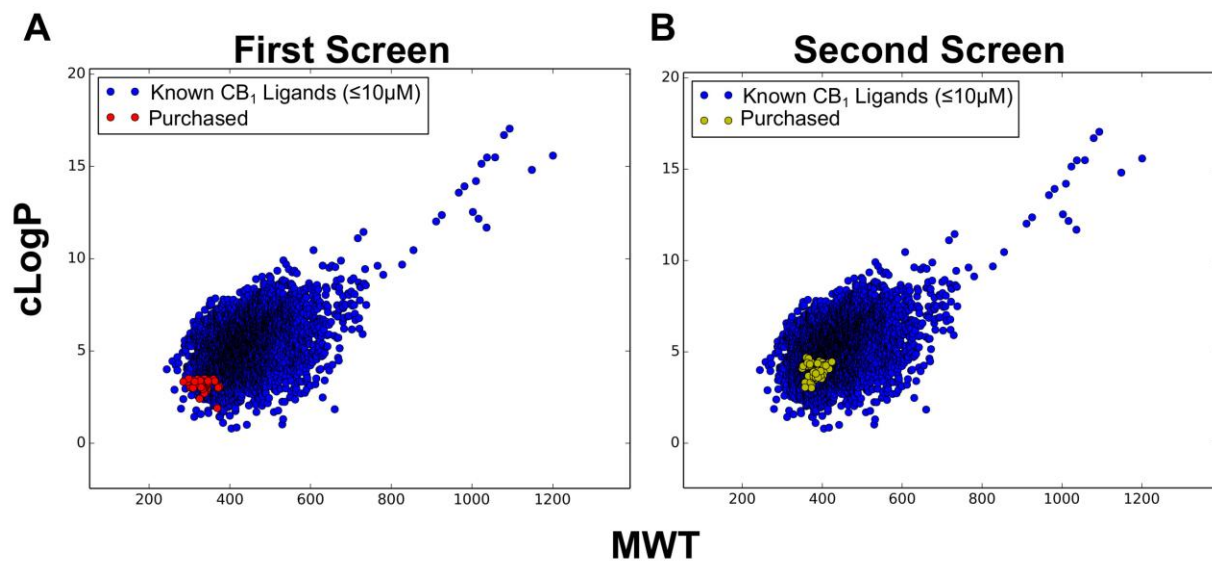
and molecules that were similar to known CB<sub>1</sub> and CB<sub>2</sub> ligands from ChEMBL24<sup>32</sup> were removed from further inspection.

The best-scoring molecules from the top 10,000 clusters were inspected for interaction with important residues in the CB<sub>1</sub> site, including hydrogen bonds with S383<sup>7,39</sup> and H178<sup>2,65</sup>, as well as other polar partners including T201<sup>3,37</sup>.

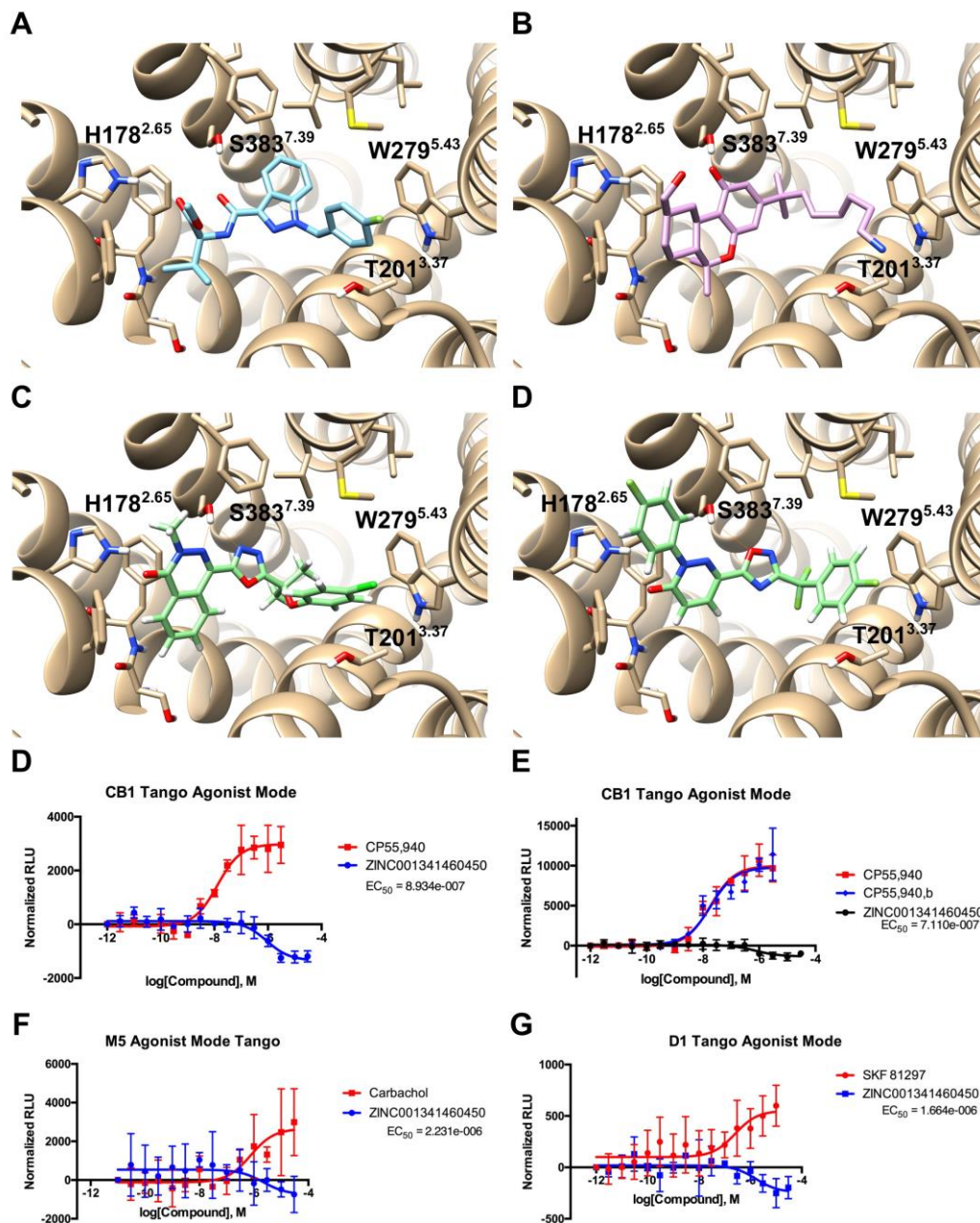
Conformationally strained molecules, as well as those with unsatisfied hydrogen-bond donors, were eliminated<sup>33</sup>. If a representative cluster member fit these criteria, all its cluster members were inspected, and the best molecule in terms of geometry and chemical properties was chosen for synthesis and testing. This resulted in 60 molecules for purchase, with 55 being synthesized for testing. Of the 55 tested, none of these molecules had activity in PRESTO-Tango functional assays, which we believed to be due to assay artifacts, but also the 'lead-like' nature of the library we docked, which lies at the periphery in property space compared to known CB<sub>1</sub> ligands (**Figure 5.1-2**).

We therefore turned to a larger, greasier subset of ZINC ranging from cLogP of 3.5 to >5, and molecular mass ranging from 350 to >500 Daltons, which comprised over 74 million molecules. Docking again to the CB<sub>1</sub> orthosteric site, and prioritizing novel chemotypes unrelated to known cannabinoid ligands, we focused on molecules that overlapped significantly with known CB<sub>1</sub> ligands in terms of physical properties like molecular weight and cLogP (**Figure 5.1**), as well as interaction properties like the number of proposed hydrogen bonds in the orthosteric site, and similar chemical moieties such as gem-dimethyl groups and halogen-containing benzene rings making stacking interactions with W279<sup>5,43</sup> (**Figure 5.2**). Of these, we purchased 60 molecules, 58 being successfully synthesized. As before, none of the molecules were reproducibly

active in functional assays, prompting us to perform radioligand displacement assays. Of 46 molecules in the second virtual screen, 8 exhibited high affinity in single-point radioligand displacement assays (**Table 5.1**). One of the most potent ligands, ZINC1341460450, demonstrated inverse agonist activity in Tango assays (**Figure 5.2**), but this activity could not be reproduced. Similarly, this molecule showed activity at unrelated targets like the muscarinic acetylcholine M5 and D1 dopamine receptors, suggesting that it may be promiscuous, or that the formulation of the compound in the functional assays affects its activity. Re-testing of all 58 compounds from the second screen and original 55 compounds in light of these new data are currently underway. In the future, we hope to determine why these compounds are not reproducible in functional assays and use structure-based optimization to improve potency and functional outputs of the 8 high affinity binders.

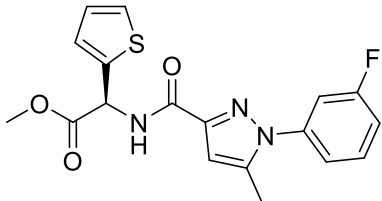
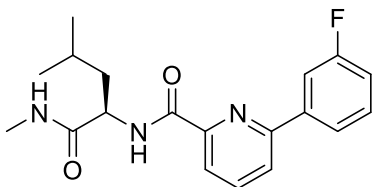
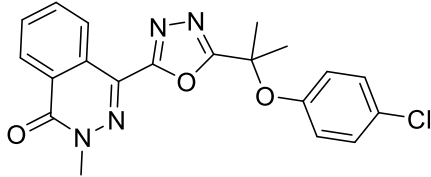
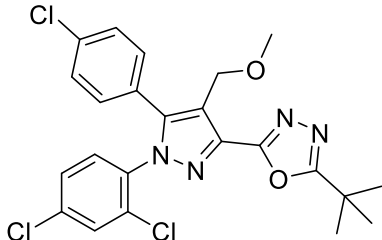
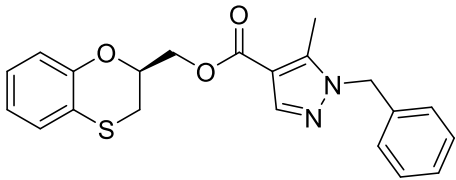
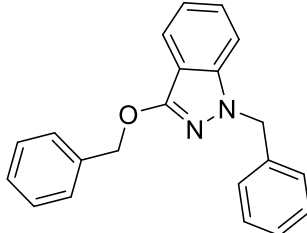
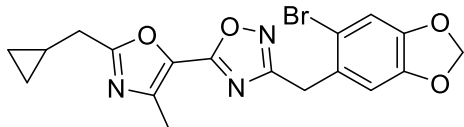
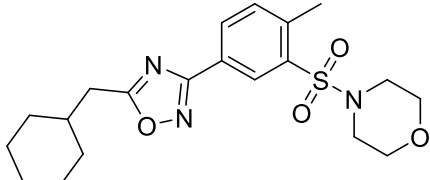
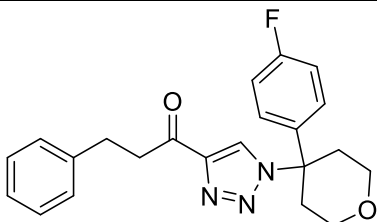
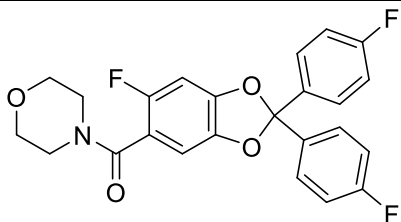


**Figure 5.1. Comparison of properties of predicted and known CB<sub>1</sub> ligands.** Calculated octanol-water partition coefficients (cLogP) and molecular weight (MWT) of known CB<sub>1</sub> ligands (blue) and purchased molecules in the first virtual screen (red, A) and purchased molecules in the second virtual screen (yellow, B).

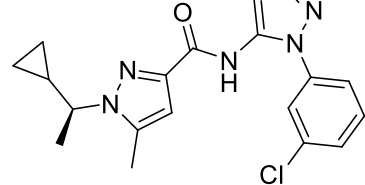
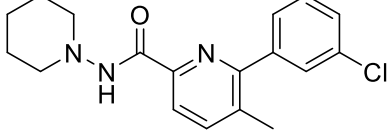
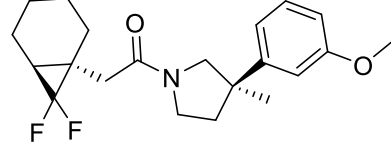
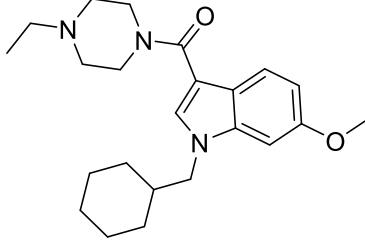
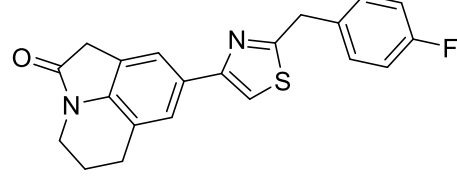
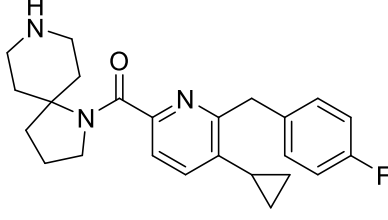


**Figure 5.2. Poses and functional dose response curves of novel ligands.** A) CryoEM pose of MDMB-Fubinaca (PDB: 6N4B), a synthetic cannabinoid agonist, which interacts with both S383<sup>7.39</sup> and H178<sup>2.65</sup> and makes stacking interactions with W279<sup>5.43</sup>. B) Crystallographic pose of AM-841 (PDB: 5XR8), a synthetic phytocannabinoid-like agonist that interacts with S383<sup>7.39</sup>. Docked poses of ZINC1341460450 (C) and ZINC504609243 (D), which both have halogen-containing benzene rings stacking with W279<sup>5.43</sup>. PRESTO-Tango functional assays of ZINC1341460450 (D,E). PRESTO-Tango functional assays of ZINC1341460450 muscarinic acetylcholine receptor 5 (F) and D1 dopamine receptor (G).

**Table 5.1.** Active molecules from single-point radioligand displacement assay.

Active Molecule	Predicted IC <sub>50</sub> ( $\mu\text{M}$ , after 1 point)	Closest Known CB <sub>1</sub> /CB <sub>2</sub> Molecule (ECFP4 Tanimoto Coefficient)
	1	 CHEMBL3922344 (0.30)
	2	 CHEMBL519214 (0.36)
	2	 CHEMBL3116279 (0.28)
	4	 CHEMBL472680 (0.24)
	8	 CHEMBL259699 (0.29)



Active Molecule	Predicted IC <sub>50</sub> ( $\mu$ M, after 1 point)	Closest Known CB <sub>1</sub> /CB <sub>2</sub> Molecule (ECFP4 Tanimoto Coefficient)
 <p>ZINC538517902</p>	8	 <p>CHEMBL3915046</p> <p>0.32</p>
 <p>ZINC618737218</p>	9	 <p>CHEMBL3347301</p> <p>0.31</p>
 <p>ZINC506941038</p>	9	 <p>CHEMBL3890211</p> <p>0.28</p>

## 5.4 Discussion

Though we may have identified an inverse agonist when our goal was to find agonists, it is possible that an inverse agonist may be useful in pain indications. It has been shown that CB<sub>1</sub> antagonists like rimonabant can reduce CFA-induced arthritis pain behavior, as well as reduce thermal hyperalgesia and mechanical allodynia in rodents<sup>34</sup>. Similarly, another antagonist, SR141716, is capable of counteracting neuropathic pain by reducing neurogenic inflammation via downregulation of TNF- $\alpha$  expression<sup>35</sup>.

If these molecules prove to be true hits, we have devised several analog schemes to improve potency and modify functional activity. This includes extending the length between the central scaffold and the moiety interacting with W279<sup>5,43</sup>, as well as changing or adding a halogen on this moiety, which has been shown to increase potency to picomolar affinities<sup>36</sup>. Similarly, we have considered substituting the hydrogen bond donor that interacts with S383<sup>7,39</sup> with various groups as outlined previously<sup>37,38</sup>. Inspection of the antagonist-bound crystal structure also demonstrates a doubling of the binding site volume<sup>22,23</sup>, which, if ZINC1341460450 is an inverse agonist, provides a justification for reducing the size of the moiety interacting with H178<sup>2,65</sup>, such that its analog stays within the agonist binding site volume. This location is partially exposed to solvent, allowing for charged moieties in CB<sub>1</sub> ligands<sup>39</sup>, which could serve as the basis for novel, peripherally restricted CB<sub>1</sub> molecules. Similarly, peripherally restricted cannabinoids have been identified by focusing on compounds with higher calculated polar surface area, such that they do not pass the blood-brain barrier<sup>40</sup>. Overall, this project is still in its early stages, but given the exciting data we

have now, there are a lot of paths forward, which should result in an interesting set of molecules to test *in vivo*.

## 5.5 Methods

### Docking Calculations and Virtual Screens.

In the first screen, a cryoEM structure of the human CB<sub>1</sub> cannabinoid receptor was used in the docking calculations. Atoms of the cryogenic ligand, MDMB-Fubinaca, were used to seed the matching sphere calculation in the orthosteric site. These spheres represent favorable positions for ligand atoms to dock, with 45 total being used. The receptor structure was protonated using REDUCE<sup>41</sup> and assigned AMBER united atom charges<sup>42</sup>. The volume of the low protein dielectric, which defines the boundary between solute and solvent in Poisson-Boltzmann electrostatic calculations, was extended out 0.8 Å from the protein surface. These pseudo-atom positions represent possible ligand atom positions. The desolvation volume of the site was also increased using similar atom positions using a radius of 1.0 Å. Scoring grids were precalculated using CHEMGRID<sup>43</sup> for AMBER van der Waals potential, QNIFFT<sup>44</sup> for Poisson-Boltzmann-based electrostatic potentials, and SOLVMAP<sup>45</sup> for ligand desolvation.

These potential grids and ligand-matching parameters were evaluated for their ability to enrich known CB<sub>1</sub> ligands over property-matched decoys. We extracted 199 known CB<sub>1</sub> ligands – both agonists and antagonists – from the IUPHAR database<sup>46</sup>, ChEMBL24<sup>32</sup>, and ZINC15, and generated 14,929 property-matched decoys using an in-house pipeline. Docking success was judged based on the ability to enrich known ligands over the decoys by docking rank, using adjusted logAUC values. We also

ensured that molecules with extreme physical properties were not enriched, such that we wanted neutral molecules to be prioritized in the best-scoring molecules. The docking setup was also judged for how well it reproduced the expected and known binding modes of the known ligands.

The “lead-like” subset of ZINC15 (<http://zinc15.docking.org>) with calculated octanol-water partition coefficients (cLogP)  $\leq 3.5$  and with molecular mass  $\leq 350$  Da, was docked against the CB<sub>1</sub> orthosteric site using DOCK3.7<sup>31</sup>. This library contained over 225 million molecules, most of which were make-on-demand compounds from the Enamine REAL set<sup>27</sup>. Of these, more than 181 million successfully docked. An average of 3,283 orientations, and for each orientation, an average of 441 conformations was sampled. Overall, about 123 trillion complexes were sampled and scored. The total time was about 70,470 core hours, or 1.96 calendar days on 1,500 cores.

To reduce redundancy of the top scoring docked molecules, the top 300,000 ranked molecules were clustered by ECFP4-based Tanimoto coefficient (Tc) of 0.5, and the best scoring member was chosen as the cluster representative molecule. These 51,365 clusters were filtered for novelty by calculating the ECFP4-based Tanimoto coefficient against >7,000 CB<sub>1</sub> and CB<sub>2</sub> receptor ligands from the ChEMBL24<sup>32</sup> database. Molecules with Tanimoto coefficients  $\geq 0.38$  to known CB<sub>1</sub>/CB<sub>2</sub> ligands were not pursued further.

After filtering for novelty, the docked poses of the best-scoring members of each cluster were filtered by the proximity of their polar moieties to S383<sup>7,39</sup>, T201<sup>3,37</sup>, or H178<sup>2,65</sup>, and manually inspected for favorable geometry and interactions. Of the most visually favorable molecules, all members of its cluster within the top 300,000 molecules

were inspected, and one of these was chosen to replace the cluster representative if they exhibited more favorable poses or chemical properties. Of these, 60 compounds were chosen for testing, 55 of which were successfully synthesized.

In the second screen, a crystal structure of the CB<sub>1</sub> receptor (PDB: 5XR8)<sup>22</sup> was used in the docking calculations. The coordinates of M363<sup>6.55</sup> were modified slightly, while still maintaining the residue within the electron density, and the full structure with MDMA-Fubinaca overlaid into the orthosteric site was minimized with Schrödinger's Maestro. Atoms of the crystal ligand, AM-841, and the cryogenic ligand, MDMA-Fubinaca, were combined and used to seed the matching sphere calculation in the orthosteric site, with 45 total spheres used. As before, the structure was protonated with REDUCE and assigned AMBER united atom force field charges. The volume of the low protein dielectric was increased by 1.5 Å from the protein surface, and the desolvation volume was increased by 1.9 Å. The desolvation volume was removed around S383<sup>7.39</sup> and H178<sup>2.65</sup> to decrease the desolvation cost near these residues and to increase the number of molecules that would form polar contacts with them. As in the first setup, this new docking setup was judged based on its ability to enrich known 199 CB<sub>1</sub> ligands over 14,929 property-matched decoys, to prioritize neutral over charged molecules, and to reproduce the expected and known binding modes of CB<sub>1</sub> ligands.

A larger, greasier subset of ZINC15 with cLogP ranging from 3 to >5 and molecular mass ranging from 350 to >500 was docked against the CB<sub>1</sub> orthosteric site using DOCK3.7. This library contained over 74 million molecules. Of these, more than 18 million successfully docked. An average of 4,713 orientations, and for each orientation, an average of 645 conformations was sampled. Overall, about 63 trillion

complexes were sampled and scored. The total time was about 25,432 core hours, or 0.71 calendar days on 1,500 cores.

As before, the top 300,000 ranked molecules were clustered by ECFP4-based Tanimoto coefficient (Tc) of 0.5, and the best scoring member was chosen as the cluster representative. This resulted in 60,420 clusters, which were filtered for novelty by calculating the ECFP4-based Tanimoto coefficient against >7,000 CB<sub>1</sub> and CB<sub>2</sub> receptor ligands from the ChEMBL24 database. Molecules with Tanimoto coefficients  $\geq$  0.38 to known CB<sub>1</sub>/CB<sub>2</sub> ligands were not pursued further.

The docked poses were again filtered for proximity to S383<sup>7,39</sup>, T201<sup>3,37</sup>, or H178<sup>2,65</sup>, manually inspected for favorable geometry and interactions, and the full cluster within the top 300,000 molecules was inspected for more favorable replacements. Of these, 60 compounds were chosen for testing, 58 of which were successfully synthesized.

### **In vitro pharmacology**

The PRESTO-Tango<sup>47</sup> and GloSensor assays using the human CB<sub>1</sub> cannabinoid receptor construct, were used to determine agonist and inverse agonist activity. Single-point assays were performed as described previously<sup>22,23</sup>, using the agonist, [<sup>3</sup>H]CP55,940 as a positive control.

## References

1. Mechoulam, R. & Ben-Shabat, S. From gan-zi-gun-nu to anandamide and 2-arachidonoylglycerol: the ongoing story of cannabis. *Nat Prod Rep* **16**, 131-143, doi:10.1039/a703973e (1999).
2. Banister, S. D., Krishna Kumar, K., Kumar, V., Kobilka, B. K. & Malhotra, S. V. Selective modulation of the cannabinoid type 1 (CB1) receptor as an emerging platform for the treatment of neuropathic pain. *Medchemcomm* **10**, 647-659, doi:10.1039/c8md00595h (2019).
3. Rubino, T. *et al.* Cellular mechanisms underlying the anxiolytic effect of low doses of peripheral Delta9-tetrahydrocannabinol in rats. *Neuropsychopharmacology* **32**, 2036-2045, doi:10.1038/sj.npp.1301330 (2007).
4. Alonso, M. *et al.* Anti-obesity efficacy of LH-21, a cannabinoid CB(1) receptor antagonist with poor brain penetration, in diet-induced obese rats. *Br J Pharmacol* **165**, 2274-2291, doi:10.1111/j.1476-5381.2011.01698.x (2012).
5. Jones, D. End of the line for cannabinoid receptor 1 as an anti-obesity target? *Nat Rev Drug Discov* **7**, 961-962, doi:10.1038/nrd2775 (2008).
6. Parker, L. A., Rock, E. M. & Limebeer, C. L. Regulation of nausea and vomiting by cannabinoids. *Br J Pharmacol* **163**, 1411-1422, doi:10.1111/j.1476-5381.2010.01176.x (2011).
7. Agarwal, N. *et al.* Cannabinoids mediate analgesia largely via peripheral type 1 cannabinoid receptors in nociceptors. *Nat Neurosci* **10**, 870-879, doi:10.1038/nn1916 (2007).

8. Turgeman, I. & Bar-Sela, G. Cannabis for cancer - illusion or the tip of an iceberg: a review of the evidence for the use of Cannabis and synthetic cannabinoids in oncology. *Expert Opin Investig Drugs* **28**, 285-296, doi:10.1080/13543784.2019.1561859 (2019).
9. Schug, S. A. & Goddard, C. Recent advances in the pharmacological management of acute and chronic pain. *Ann Palliat Med* **3**, 263-275, doi:10.3978/j.issn.2224-5820.2014.10.02 (2014).
10. Turcotte, D. *et al.* Examining the roles of cannabinoids in pain and other therapeutic indications: a review. *Expert Opin Pharmacother* **11**, 17-31, doi:10.1517/14656560903413534 (2010).
11. Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C. & Bonner, T. I. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346**, 561-564, doi:10.1038/346561a0 (1990).
12. Munro, S., Thomas, K. L. & Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365**, 61-65, doi:10.1038/365061a0 (1993).
13. Hazekamp, A., Ware, M. A., Muller-Vahl, K. R., Abrams, D. & Grotenhermen, F. The Medicinal Use of Cannabis and Cannabinoids—An International Cross-Sectional Survey on Administration Forms. *Journal of Psychoactive Drugs* **45**, 199-210, doi:10.1080/02791072.2013.805976 (2013).
14. Cohen, K., Weizman, A. & Weinstein, A. Positive and Negative Effects of Cannabis and Cannabinoids on Health. *Clin Pharmacol Ther* **105**, 1139-1147, doi:10.1002/cpt.1381 (2019).



15. Sachs, J., McGlade, E. & Yurgelun-Todd, D. Safety and Toxicology of Cannabinoids. *Neurotherapeutics* **12**, 735-746, doi:10.1007/s13311-015-0380-8 (2015).
16. Agrawal, A., Nurnberger, J. I., Jr., Lynskey, M. T. & Bipolar Genome, S. Cannabis involvement in individuals with bipolar disorder. *Psychiatry Res* **185**, 459-461, doi:10.1016/j.psychres.2010.07.007 (2011).
17. van Amsterdam, J., Brunt, T. & van den Brink, W. The adverse health effects of synthetic cannabinoids with emphasis on psychosis-like effects. *J Psychopharmacol* **29**, 254-263, doi:10.1177/0269881114565142 (2015).
18. Pertwee, R. G. Targeting the endocannabinoid system with cannabinoid receptor agonists: pharmacological strategies and therapeutic possibilities. *Philos Trans R Soc Lond B Biol Sci* **367**, 3353-3363, doi:10.1098/rstb.2011.0381 (2012).
19. Cheng, Y. & Hitchcock, S. A. Targeting cannabinoid agonists for inflammatory and neuropathic pain. *Expert Opin Investig Drugs* **16**, 951-965, doi:10.1517/13543784.16.7.951 (2007).
20. Yu, X. H. *et al.* A peripherally restricted cannabinoid receptor agonist produces robust anti-nociceptive effects in rodent models of inflammatory and neuropathic pain. *Pain* **151**, 337-344, doi:10.1016/j.pain.2010.07.019 (2010).
21. Burstein, S. H., Karst, M., Schneider, U. & Zurier, R. B. Ajulemic acid: A novel cannabinoid produces analgesia without a "high". *Life Sci* **75**, 1513-1522, doi:10.1016/j.lfs.2004.04.010 (2004).
22. Hua, T. *et al.* Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **547**, 468-471, doi:10.1038/nature23272 (2017).

23. Hua, T. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **167**, 750-762 e714, doi:10.1016/j.cell.2016.10.004 (2016).
24. Li, X. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* **176**, 459-467 e413, doi:10.1016/j.cell.2018.12.011 (2019).
25. Shao, Z. *et al.* High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* **540**, 602-606, doi:10.1038/nature20613 (2016).
26. Krishna Kumar, K. *et al.* Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell* **176**, 448-458.e412, doi:<https://doi.org/10.1016/j.cell.2018.11.040> (2019).
27. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229, doi:10.1038/s41586-019-0917-9 (2019).
28. Stein, R. M. *et al.* Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609-614, doi:10.1038/s41586-020-2027-0 (2020).
29. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185-190, doi:10.1038/nature19112 (2016).
30. Huang, X. P. *et al.* Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **527**, 477-483, doi:10.1038/nature15699 (2015).
31. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992, doi:10.1371/journal.pone.0075992 (2013).
32. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res* **42**, D1083-1090, doi:10.1093/nar/gkt1031 (2014).

33. Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* **59**, 4103-4120, doi:10.1021/acs.jmedchem.5b02008 (2016).
34. Croci, T. & Zarini, E. Effect of the cannabinoid CB1 receptor antagonist rimonabant on nociceptive responses and adjuvant-induced arthritis in obese and lean rats. *Br J Pharmacol* **150**, 559-566, doi:10.1038/sj.bjp.0707138 (2007).
35. Costa, B. *et al.* Effect of the cannabinoid CB1 receptor antagonist, SR141716, on nociceptive response and nerve demyelination in rodents with chronic constriction injury of the sciatic nerve. *Pain* **116**, 52-61, doi:10.1016/j.pain.2005.03.043 (2005).
36. Nikas, S. P. *et al.* The role of halogen substitution in classical cannabinoids: a CB1 pharmacophore model. *AAPS J* **6**, e30, doi:10.1208/aapsj060430 (2004).
37. Schoeder, C. T., Hess, C., Madea, B., Meiler, J. & Muller, C. E. Pharmacological evaluation of new constituents of "Spice": synthetic cannabinoids based on indole, indazole, benzimidazole and carbazole scaffolds. *Forensic Toxicol* **36**, 385-403, doi:10.1007/s11419-018-0415-z (2018).
38. Banister, S. D. & Connor, M. The Chemistry and Pharmacology of Synthetic Cannabinoid Receptor Agonists as New Psychoactive Substances: Origins. *Handb Exp Pharmacol* **252**, 165-190, doi:10.1007/164\_2018\_143 (2018).
39. Amato, G. S. *et al.* Blocking Alcoholic Steatosis in Mice with a Peripherally Restricted Purine Antagonist of the Type 1 Cannabinoid Receptor. *J Med Chem* **61**, 4370-4385, doi:10.1021/acs.jmedchem.7b01820 (2018).

40. Adam, J. M. *et al.* Low brain penetrant CB1 receptor agonists for the treatment of neuropathic pain. *Bioorg Med Chem Lett* **22**, 2932-2937, doi:10.1016/j.bmcl.2012.02.048 (2012).
41. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735-1747, doi:10.1006/jmbi.1998.2401 (1999).
42. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106**, 765-784, doi:10.1021/ja00315a051 (1984).
43. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. Automated docking with grid-based energy evaluation. **13**, 505-524, doi:10.1002/jcc.540130412 (1992).
44. Gallagher, K. & Sharp, K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys J* **75**, 769-776, doi:10.1016/S0006-3495(98)77566-6 (1998).
45. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
46. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res* **44**, D1054-1068, doi:10.1093/nar/gkv1037 (2016).
47. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat Struct Mol Biol* **22**, 362-369, doi:10.1038/nsmb.3014 (2015).

## Chapter 6: Future Directions

So where are we now? What have we learned? What happens next? In the preceding chapters, I have presented some data that generate more questions. Inevitably, some questions that I initially had are still left unanswered, but to be fair, these are difficult questions that don't have straightforward answers. How is it that a weighting of 1.0 for all three scoring function terms generates the best, most reliable performance? How is it that all these different theories, charge models, and parameter choices fit together even when neglecting key energetic terms like entropy and receptor desolvation? How are we still able to find ligands that hit a protein? How is it that our hit rate seems to be increasing? How can we best balance the scoring function if we only rely on the weak, unphysical criterion of enrichment and log AUC values? What is the relationship between enrichment and identification of new ligands? How does docking setup choice affect this relationship? Why have many of my supposedly more physically correct fixes to the DOCK scoring function and pipeline, such as using the all-atom AMBER parameters and charges, diminished performance? A key passage from the DOCK ligand desolvation paper<sup>1</sup>, and also a key refrain in many lab members' presentations ("it's a miracle it works!"), has permanently fixed itself in my mind for the past few years:

"At first blush it may seem surprising that docking programs ever discover new ligands for proteins, so many are the approximations made by their scoring functions. That they do reflects, at least partly, a cancellation of errors among approximations. Whenever a term is improved by making it physically more

correct it is easy to image that the new model may perform worse than the old by upsetting this prior cancellation of errors.”

But how do we identify this “cancellation of errors”? Is it possible to identify how these incomplete approximations of physical phenomena fit together in an incomplete way, yet successfully, in some cases, model reality? Where does approximation end and reality begin? I think these questions address the difficulties we had in Chapters 1, 2, and 3 of incorporating blurry GIST and identifying the correct setup after running parameter scanning in the scoring function weights.

## **6.1 A new methods development pipeline**

Out of necessity and to save our future selves, we created tools to help one from deceiving themselves after preparing a setup – using different decoy backgrounds with various properties, but also using bootstrapping to identify significant differences. I think this is a step in the right direction, as we need more ways to convince ourselves that our results are right for the right reasons.

In the last few months of my PhD, I have been helping Jiankun Lyu, Stefan Gahbauer, and John Irwin with a new ligand building pipeline. We have taken a multi-pronged approach to measuring performance of the ligands built with the new pipeline and the previous pipeline: strain energies, enrichment performance, quantifying number of dockable molecules built, and RMSD values to crystallographic ligand poses and poses we judged to be good. I think this is an effective approach – looking at the problem from different perspectives to identify any problems before we commit to it. I

think this kind of approach should be implemented in a pipeline for new methods development projects. The tests will change with the type of new implementation being tested, of course – for example, the tests we are doing on the ligand building pipeline could be applied to a new ligand charge model, or the inclusion of individual ligand desolvation energies for each conformer of a ligand – ideas that have been, or have been planned to be, toyed with in the lab. This would be a “ligand-based” pipeline.

In terms of a new scoring function term, we could have another set of tests – enrichment as a first step with low dielectric and ligand desolvation thin sphere parameter scanning, weighting the new scoring function term differently to determine change in performance, RMSD calculations to crystallographic and good poses, comparison of energies in DOCK versus actual energies from the program this new scoring function term comes from, an alternative set of benchmarks with different preparations, among other tests. For example, I created a version of 40 DUD-E systems that were all built with chloroform ligand desolvation energies instead of hexadecane to check whether blurry GIST could be incorporated more readily with smaller ligand desolvation energies. This would be a more “protein-based” pipeline. Basically, all the tests that are or should be performed once something new is implemented becomes a pipeline. Just as applications projects follow the common path of matching sphere scanning, thin sphere scanning, charge extrema decoys, Goldilocks decoys, pose viewing, with modifications as necessary, so too should methods projects follow a common path of a series of tests that would get at the heart of whether this new method is right for the right reasons before a prospective screen is run. I think methods development will always have a place in the Shoichet lab and having a battery of tests,

a set of tools, and benchmarks that the methods developer could turn to or take ideas from would be extremely useful and save a lot of time. This is one reason why I have consistently added my scripts to the lab wiki (<http://wiki.docking.org>), so that others may use them and modify them as desired.

## **6.2 A new receptor desolvation method**

What's next for receptor desolvation? I think we would have had more success, had we focused on a buried binding site rather than one that is solvent-exposed, but this is one important and interesting lesson we learned from the blurry GIST work – that it may be system- and binding site-dependent. Thus, I think if blurry GIST would be applied to a different system, potentially a GPCR for which desolvation is important, and that has only partial or no exposure to bulk solvent, and thus little reorganization energy, we might find that blurry GIST has a more beneficial effect. Identifying a protein for which desolvation contributes most of the water energetics, as we did with cytochrome c peroxidase, may lead to better outcomes.

The extra effort involved in running a molecular dynamics simulation and then GIST, followed by “blurring” the grids, and then thin sphere scanning with blurry GIST weight scanning, all for a mild enrichment improvement, as well as the intimidating GIST papers, has limited its acceptance in the lab. Perhaps the more streamlined tools available<sup>2</sup> could give people a better idea of what GIST is capable, and how to more easily use it. If this is not enough, I referenced multiple water location and energy prediction programs in the Introduction, many of which may be easier or faster to use and can be applied to binding sites where water is important. Some of these programs



that predict water energies are grid-based and can, in theory, be easily incorporated into my trilinear interpolation receptor desolvation scheme in the DOCK code. Thus, it should be quick to test other solvation energy methods to determine if they fit into the DOCK scoring function as well as GIST does or better.

Many lab members that choose to include water in their docking setups include key crystallographic waters and minimize them in the presence of the ligand and protein to identify low energy water orientations. This is a simple way to find molecules with water-mediated interactions in a screen but restricts the size of the binding site, as the waters now become part of the protein, and therefore, doesn't allow one to find molecules that displace these water molecules. The goal was to get around this situation by combining GIST with turning waters on and off<sup>3</sup> – by including displaceable waters and also including the desolvation cost of those waters. It wasn't until I had stopped working on this project that I found a bug in the flexible receptor docking code, such that the van der Waals and ligand desolvation grids were being double counted. This bug has been fixed, but I did not have the time to re-test the 10 or so DUD-E systems to which I applied the combination of GIST and displaceable waters. Another issue that has arisen is that Simplex minimization does not work with the flexible receptor docking code. Thus, future developers should first reconcile this in the DOCK source code, as minimization improves enrichment significantly, before trying the combination of displaceable waters and GIST desolvation energies.

Once this is completed, I think continuing with the same trajectory as I had – including crystallographic or computationally-predicted waters and a desolvation grid from GIST or some other program – can be attempted again. Early during this project,

Trent and I had ideas of running GIST on multiple protein-ligand complexes, where the ligands had different chemotypes, and thus should have different mediating water locations, energies, and surrounding solvent shells. Then when docking, multiple GIST grids could be read into DOCK for scoring, thereby extending the flexible receptor docking code to receptor desolvation. This is not physically accurate, as the specific ligand that you run with the protein would have its own set of ligand-specific GIST energies, but in the inaccurate world of docking energies, it's possible that these energies might be transferable.

We have seen that the reorganization energy is very important for protein-ligand binding from the blurry GIST work and from others<sup>4</sup>. We could potentially even include the first shell solvent energies from the different protein-ligand complex GIST calculations, much as we did with the reorganization energy in Chapter 2 – by just adding it on to the desolvation energy. However, as identifying voxels outside of each protein-ligand complex to determine the reorganization energy would suffer from the same issues as the original displacement GIST, careful thought would have to go into how to precompute the first shell energies, and how best to incorporate them during docking, as different ligand poses would undoubtedly overlap with the surrounding solvent that was calculated using a different ligand. The details would need to be worked out, but it might be worth exploring. If we show that reorganization energy is a significant contribution to AmpC binding from our blurry GIST work, it could be used as a great model system to use for testing this method.

As in the Introduction, implementing a new water method in docking boils down to: how can we incorporate water energetics in an approximate, quick way that will meaningfully account for water's role in protein-ligand binding?

### **6.3 A generalized form of combinatorial scoring**

I was quite happy with the combinatorial scoring code from Chapter 2, in which sampling is performed once, the poses are scored with the two scoring functions sequentially, minimized, and then swapped if a better scoring pose is found with the other scoring function. As it is written now, it only works for standard and blurry GIST scoring functions, which only differ in the blurry GIST term. However, with some effort, the code can be modified so that this scheme could be applied to different, new scoring function terms. This ensures that only one screen must be performed and when docking ever-increasing small molecule libraries, would save a lot of time. Perhaps in the future, an INDOCK argument would allow users to specify the new scoring function term they want to compare to the standard scoring function, or maybe even some combination of terms that they want to compare to the standard scoring function, and then my scheme above would be performed during docking. This could be a set of terms separate from, or added onto, the standard scoring function. Then one could get a direct comparison of the benefits of this new scoring function term or set of scoring function terms relative to the standard scoring function.

## References

1. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573, doi:10.1021/ci100214a (2010).
2. Ramsey, S. *et al.* Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J Comput Chem* **37**, 2029-2037, doi:10.1002/jcc.24417 (2016).
3. Huang, N. & Shoichet, B. K. Exploiting ordered waters in molecular docking. *J Med Chem* **51**, 4862-4865, doi:10.1021/jm8006239 (2008).
4. Mahmoud, A. H., Masters, M. R., Yang, Y. & Lill, M. A. Elucidating the multiple roles of hydration for accurate protein-ligand binding prediction via deep learning. *Communications Chemistry* **3**, 19, doi:10.1038/s42004-020-0261-x (2020).

## Appendix A: Supplementary Figures and Tables

### A1. Supplementary Material for Chapter 1

#### Retrospective docking.

*Enrichment.* We quantified enrichment by calculating the area under the curve (AUC) and the log-adjusted AUC (logAUC) values with respect to the receiver operator characteristic (ROC) curves: ligand and property-matched decoys (PMD) were generated based on actives using the DUD-E method. Enrichment studies were performed on 25+1 systems: CcP-ga consisting of 46 ligands and 3,338 decoys and 25 DUD-E systems (AA2AR, ACES, ADA, AMPC, CXCR4, EGFR, FA10, FABP4, GLCM, HIVPR, HMDH, HS90A, ITAL, KIT, KITH, LCK, NRAM, PARP1, PLK1, PPARA, PTN1, PUR2, SRC, THRB, and TRY1) consisting of 6571 ligands and 397,864 decoys in total. See ref for more details of the DUD-E benchmark set.

*Pose reproduction.* We post-processed the ligands from our enrichment calculations and compared their poses to the crystallographic conformations. All crystal complexes were aligned into the docking frame using UCSF Chimera. DOCK6.6 was used to calculate the symmetry-corrected root mean square deviation (RMSD) using the Hungarian algorithm. We looked at two measures of pose fidelity: (1) average RMSD; and (2) the percent docking success ( $\#$  of poses  $<$  RMSD threshold /  $\#$  molecules  $\times$  100).

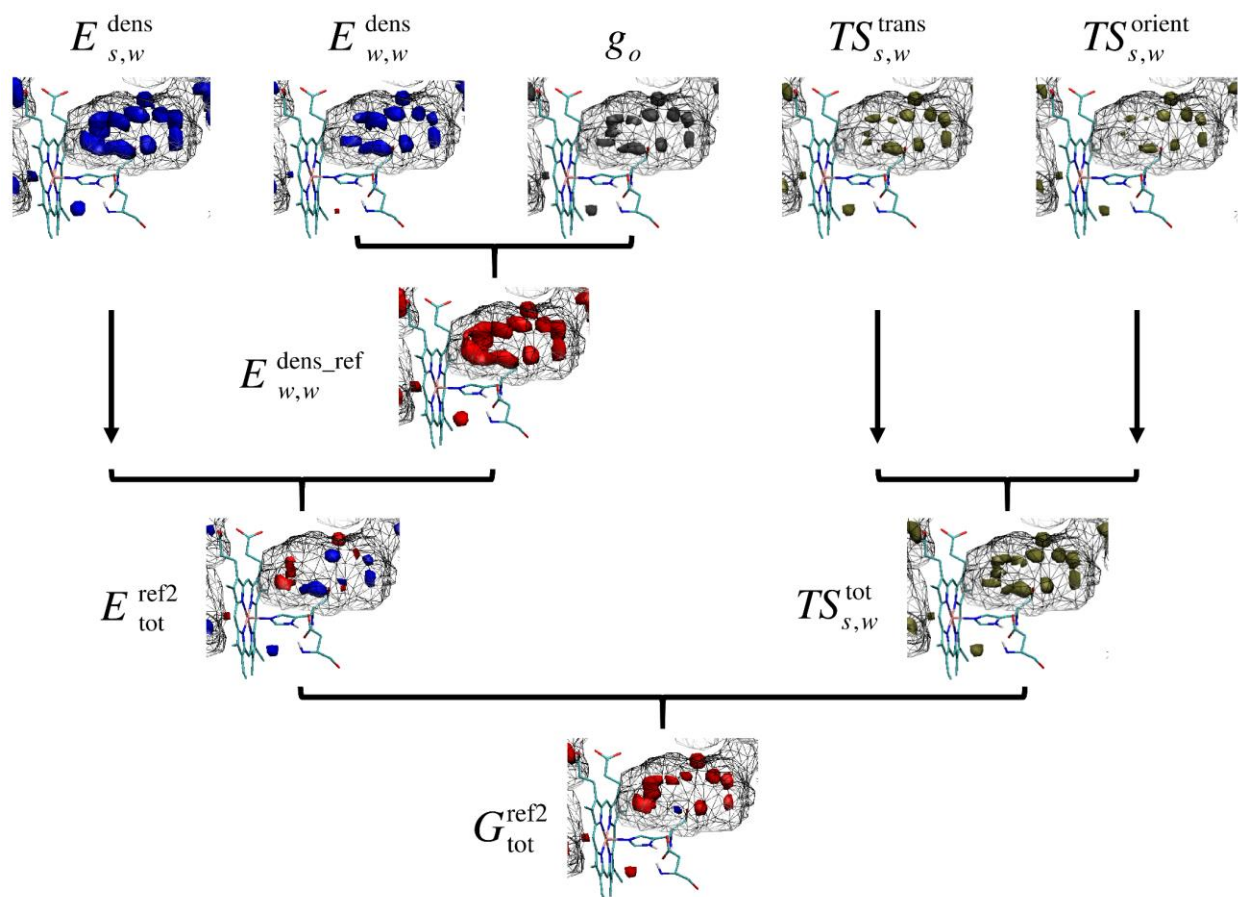
## GIST grids and how to combine them.

In docking, two tasks are performed: sampling and scoring. In this paper the objective is to improve the scoring aspect by adding a receptor desolvation ( $E_{\text{rec,desol}}$ ) term to the DOCK scoring function (eq 1, main document). The receptor desolvation term is estimated by using GIST grids. Here, we focus on how to generate GIST grids for use in docking by combining the five GIST components that are output by the Cpptraj program (cf. Ambertools14):

- Enthalpy between solvent (water) and solute (receptor) ( $E_{s,w}^{\text{dens}}$ );
- Enthalpy of water with water ( $E_{w,w}^{\text{dens}}$ ), also called the two-body term;
- Translational entropy between water and receptor ( $TS_{s,w}^{\text{trans}}$ );
- Orientational entropy between water and receptor ( $TS_{s,w}^{\text{orient}}$ );
- Density of water in the context of the receptor ( $g_o$ ).

The four energy values are in kcal/mol/Å<sup>3</sup>. The density is unitless (density/bulk density). The GIST nomenclature has undergone a development over time, particularly whether the enthalpies are to be scaled by one-half, as discussed previously, and here. The GIST grids used here are obtained using Amber14 and Ambertools14.

We combine the GIST terms (outlined above) in four physically meaningful ways to be used in docking. There are two issues to explore regarding this new GIST term: (1) the best way to combine the GIST components; and (2) the best scaling factor to bring the GIST term into balance with the other scoring function terms.



**Figure A.1.1. GIST Combinations.**

Illustration of how the GIST grids are combined in this work. For enthalpy and free energy contributions  $> 0.5$  kcal/mol/Å<sup>3</sup>, regions are coloured red. For the case  $< -0.5$  kcal/mol/Å<sup>3</sup>, the regions appear blue. Tan colored are regions with entropy contributions  $> 0.5$  kcal/mol/Å<sup>3</sup>. Regions of water density  $g_o > 6.0$  units (6 times that of bulk) are displayed in grey.

To estimate the free energy difference of water transfer (desolvation), we need to subtract the energy of water in bulk from the energy on the surface of the protein. This is done by referencing the water-water term to bulk (eq A.1.1):

$$E_{w,w}^{\text{dens\_ref}}(i) = E_{w,w}^{\text{dens}}(i) + 0.3184 \times g_o(i) \quad (\text{Equation A.1.1})$$

Here, the  $i$  refers to a grid position, a voxel. The constant was calculated using two parameters (taken from the Amber manual): mean energy,  $C_{\text{bulk}} = -9.533$  kcal/mol/water, and number density,  $C_{\text{num\_dens}} = 0.0334$  waters /  $\text{\AA}^3$ .  $C_{\text{bulk}} \times C_{\text{num\_dens}} = -0.3184$  kcal/mol/  $\text{\AA}^3$ .

In this study, we include displacement from all voxels: both high and low occupied sites. In previous IST displacement studies voxels only received a score if the density was above a cutoff. This ignores contributions from low density regions that may have a considerable contribution. Also in prior work, the energy normalized to density (eq A.1.2) was used.

$$E_{w,w}^{\text{norm\_ref}}(i) = \frac{E_{w,w}^{\text{dens}}(i)}{0.0334 \times g_o(i)} + 9.533 \quad (\text{Equation A.1.2})$$

The normalized value is the average energy per water in the voxel and thus the units of normalized energies ( $E_{w,w}^{\text{norm\_ref}}$ ) are in kcal/mol/water. Although we did consider the normalized grid (preliminary enrichment experiments yielded poor results), we chose to use the referenced grid (eq A.1.1). The units also indicate that the un-normalized grids are more compatible with our scoring function.

The GIST grids may be combined to produce the total enthalpy grid (eq A.1.3) and the total free energy grid (eq A.1.4).

$$E_{\text{tot}}^{\text{ref}}(i) = E_{s,w}^{\text{dens}}(i) + E_{w,w}^{\text{dens\_ref}}(i) \quad (\text{Equation A.1.3})$$

$$G_{\text{tot}}^{\text{ref}}(i) = E_{s,w}^{\text{dens}}(i) + E_{w,w}^{\text{dens\_ref}}(i) - (TS_{s,w}^{\text{orient}}(i) + TS_{s,w}^{\text{trans}}(i)) \quad (\text{Equation A.1.4})$$

In addition, we scaled the water-water term by two (eqs A.1.5 and A.1.6, and **Figure A.1.1**).



$$E_{tot}^{ref2}(i) = E_{s,w}^{dens}(i) + 2 \times E_{w,w}^{dens\_ref}(i) \quad (\text{Equation A.1.5})$$

$$G_{tot}^{ref2}(i) = E_{s,w}^{dens}(i) + 2 \times E_{w,w}^{dens\_ref}(i) - (TS_{s,w}^{orient}(i) + TS_{s,w}^{trans}(i)) \quad (\text{Equation A.1.6})$$

In-house Python scripts were used to combine grids and are available at

[https://github.com/tbalius/GIST\\_DX\\_tools](https://github.com/tbalius/GIST_DX_tools).

In eqs A.1.5-A.1.6, the factor of two results from every water interacting with every other water. Each water involved in the interaction retains half the energy (eq A.1.7).

$$E_k = \frac{1}{2} \sum_{\substack{l \in W \\ l \neq k}} E_{k,l} \quad (\text{Equation A.1.7})$$

Here, k and l denote waters and W is the set of all waters. The water-water term in eqs A.1.5 and A.1.6 has the full interaction energy at every voxel.

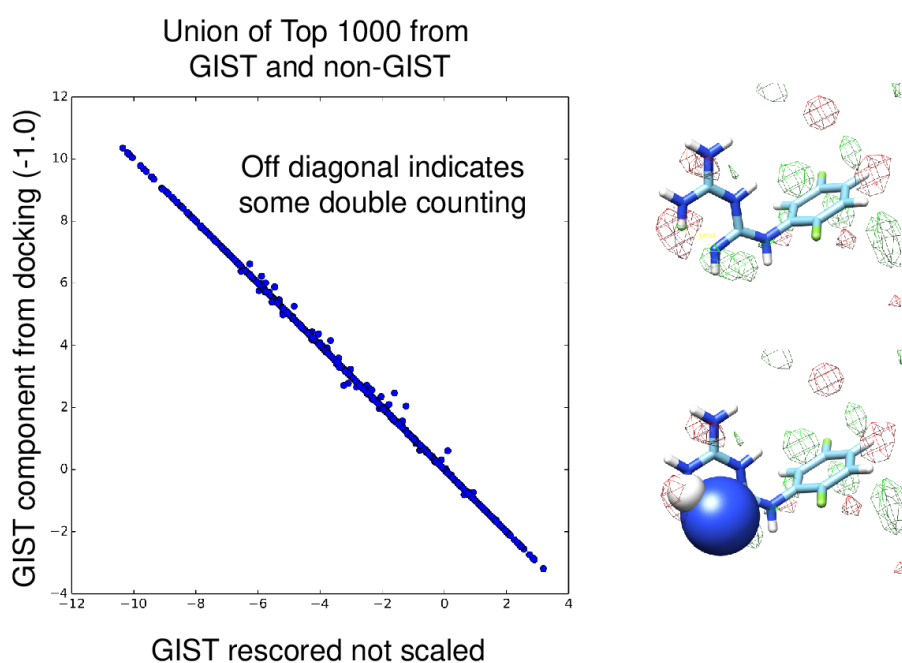
### **GIST Displacement Algorithm.**

To estimate the cost of desolvating the receptor upon binding, we first identify the voxels displaced by the ligand ( $V = \{v_i \mid v_i \in \text{ligand}\}$ ). A voxel is considered to be displaced if it is contained within the van der Waals radius of an atom during the docking calculation. We sum up the energies of those voxels (eq A.1.8) and multiply the sum by the volume of the voxel ( $vol = 0.125 \text{ \AA}^3$ ) to get a value in kcal/mol.

$$E_{rec,desol} = \alpha \times vol \times \sum_{v_i \in V} E_{GIST}(v_i) \quad (\text{Equation A.1.8})$$

Here,  $\alpha$  is a scaling factor. The algorithm is made available in the source code of the new release of the DOCK3.7 program.

To make estimating the GIST component fast and compatible with DOCK 3.7, some approximations were made. Double counting occurs only rarely when non-connected parts of the molecules overlap (**Figure A.1.2**, right panel). We determined that there was very good agreement between the GIST energies calculated with double-counting during docking and the exact GIST energies calculated by a rescoring procedure (**Figure A.1.2**, left panel).



**Figure A.1.2. GIST in docking is a good approximation.**

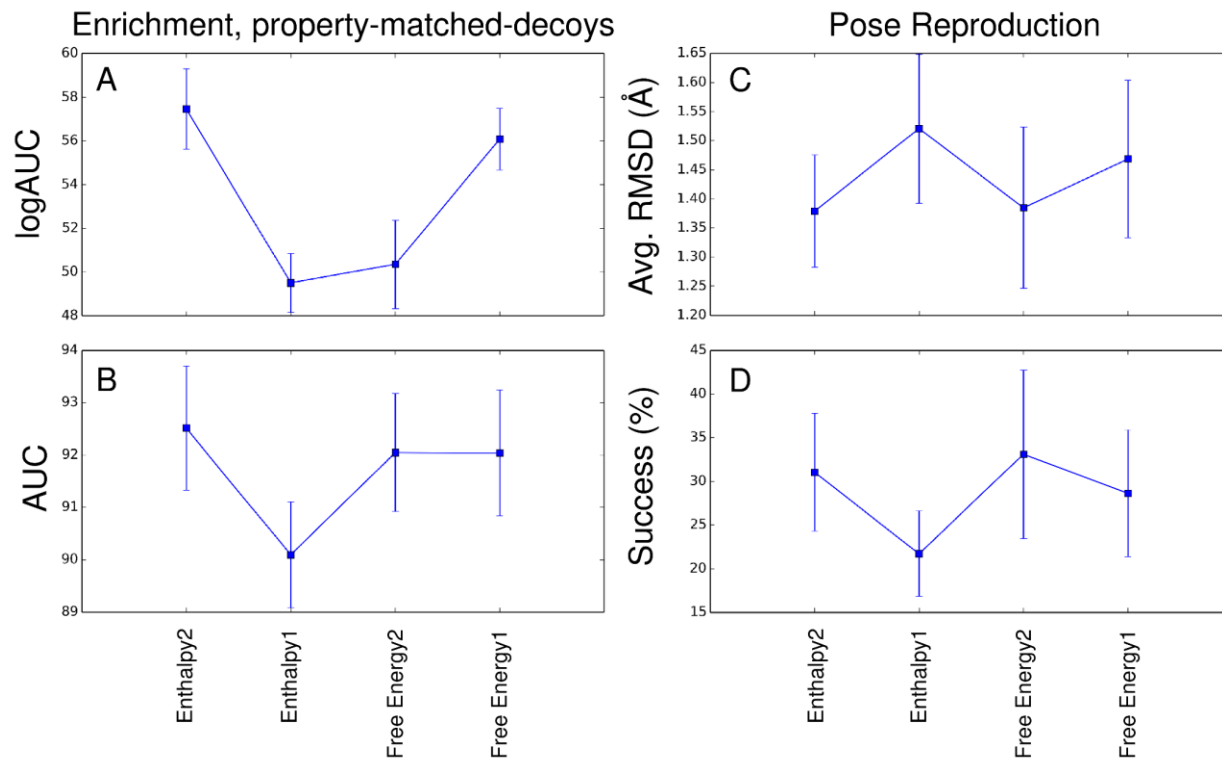
The left panel shows a correlation between the top scoring molecules from two screens, where the poses and scores are taken from the virtual docking screen with the GIST term. The GIST component is taken from the screening results (y-axis) and from rescoring the poses. The right panel shows a molecule for which double counting has occurred.

### **Comparison of GIST combinations.**

We explored which of the four combinations of the GIST components (discussed above) is best for estimating receptor desolvation during docking. We performed retrospective tests on the four GIST grids, Enthalpy1 (eq A.1.3), Free Energy1 (eq A.1.4), Enthalpy2 (eq A.1.5), and Free Energy2 (eq A.1.6), used to estimate the desolvation component (where  $\alpha = 1$  in eq A.1.8).

For each GIST grid we ran ten docking calculations to obtain a mean value and standard deviation. Because DOCK is deterministic, we modified our sampling (by perturbing the spheres used to orient the molecules into the binding site during docking) to obtain different results. Ten runs were used to better gauge the confidence in our results in the same way as performing a wet lab experiment in triplicate.

Here, the Enthalpy2 (eq A.1.8) performed the best with log AUC of 57.46 (**Figure A.1.3** and **Table A.1.1**) followed by Free Energy1 (eq A.1.7) as the second best with log AUC of 56.08. The Enthalpy2 grids were used for the remainder of this study.



**Figure A.1.3. Comparison of GIST combinations.**

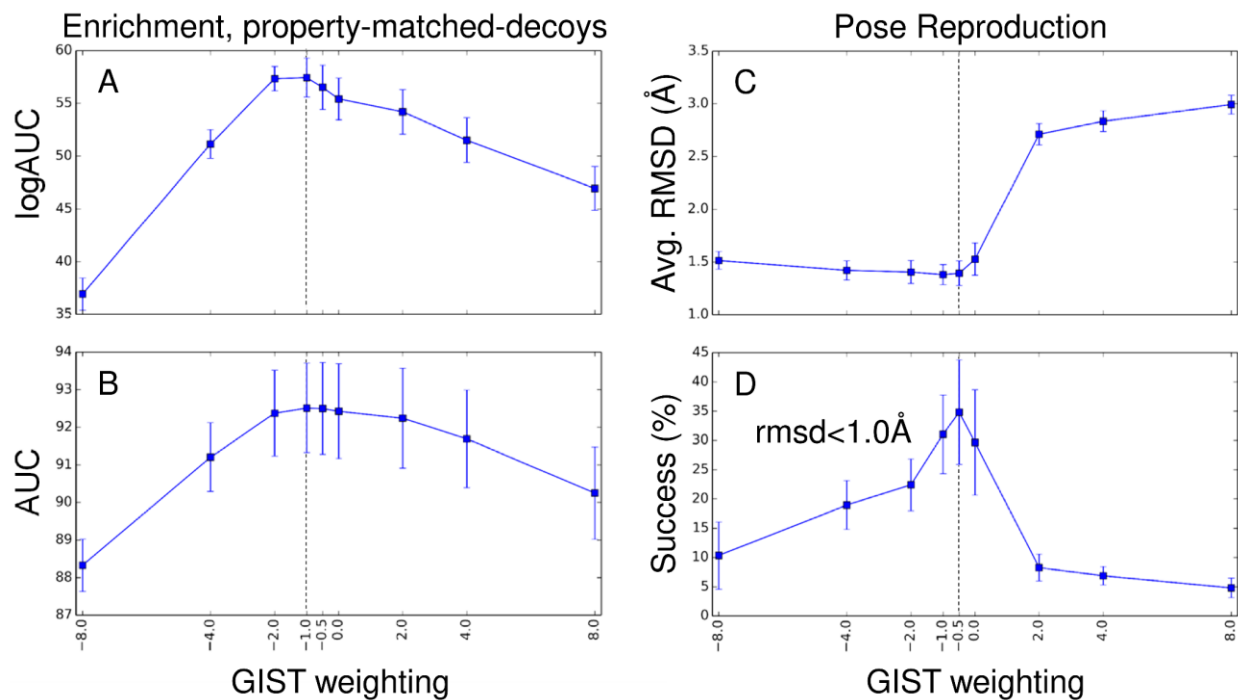
CcP-ga docking enrichment values (panels A and B) and pose reproduction (panels C and D) shown using different combinations of the GIST grids incorporated into the DOCK3.7 scoring function. The error bars are generated by running DOCK3.7 ten times with modified sampling.

**Table A.1.1.** Comparison of GIST combinations.

	LogAUC		AUC		avg RMSD (Å)		success (%) <sup>a</sup>	
	mean	std	mean	std	mean	std	mean	std
Enthalpy2	57.46	1.84	92.51	1.19	1.38	0.10	31.03	6.72
Enthalpy1	49.50	1.34	90.09	1.01	1.52	0.13	21.72	4.89
Free Energy2	50.35	2.02	92.05	1.13	1.38	0.14	33.10	9.66
Free Energy1	56.08	1.42	92.04	1.20	1.47	0.14	28.62	7.24
<sup>a</sup> Success percent of systems with RMSD less than 1.0 Å								

**Retrospective analysis for CcP-ga.**

Next, we explored what the best scaling factor ( $\alpha$  in eq A.1.8) is for weighting the receptor desolvation term in the DOCK3.7 scoring function (main text eq 1). All other terms in eq 1 (besides  $E_{\text{rec,desol}}$ ) have scaling factors of one.



**Figure A.1.4. GIST Weighting Factors.**

Retrospective analysis of CcP-ga is shown. (A, B) Enrichment analysis. Panel (A) shows log AUC. Panel (B) shows the AUC. (C, D) Pose reproduction analysis. Panel (C) shows RMSD averaged over all ligands. Panel (D) shows the success rate (number of ligand with RMSD < 1.0 Å). The blue squares represent the mean of 10 docking runs and the error bars show the standard deviation indicating the variance in distribution of values.

**Table A.1.2.** CcP-ga retrospective analysis for GIST weight.

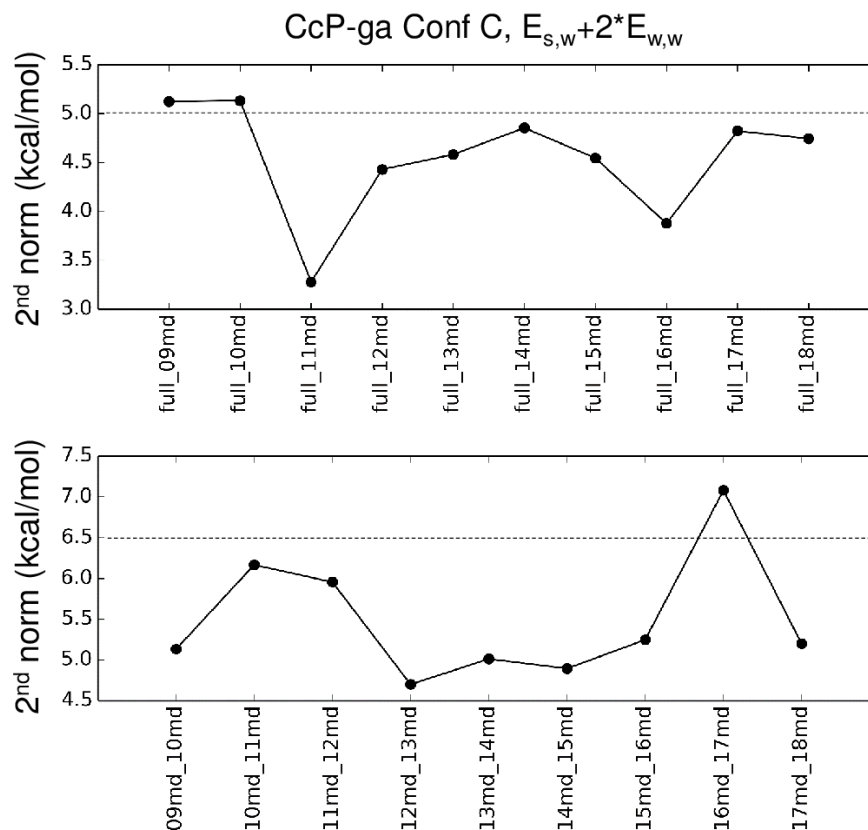
GIST scale ( $\alpha$ )	logAUC		AUC		avg RMSD (Å)		success (%) <sup>a</sup>	
	mean	std	mean	std	mean	std	mean	std
<b>-8.0</b>	36.91	1.52	88.33	0.69	1.51	0.08	10.34	5.77
<b>-4.0</b>	51.16	1.38	91.20	0.91	1.42	0.09	18.97	4.15
<b>-2.0</b>	57.36	1.16	92.38	1.14	1.40	0.11	22.41	4.43
<b>-1.0 (full GIST)</b>	57.46	1.84	92.51	1.19	1.38	0.10	31.03	6.72
<b>-0.5</b>	56.54	2.10	92.50	1.22	1.39	0.12	34.83	8.92
<b>0.0 (non-GIST)</b>	55.43	2.00	92.43	1.26	1.53	0.15	29.66	9.02
<b>2.0</b>	54.20	2.11	92.24	1.33	2.71	0.10	8.28	2.29
<b>4.0</b>	51.52	2.12	91.69	1.30	2.84	0.10	6.90	1.54
<b>8.0</b>	46.94	2.07	90.25	1.23	2.99	0.09	4.83	1.69

<sup>a</sup> Success percent of systems with RMSD less than 1.0 Å

### GIST convergence analysis.

To gauge if we ran the simulations long enough, the full simulation was divided into ten 5ns sub-trajectories and GIST grids were generated for each for comparison. First, we calculated the second-norm between pairs of GIST grids to quantify how similar the corresponding voxels are to one another between two grids; second, we docked to the different GIST grids (as the receptor desolvation component of the scoring function in eq 1) and quantified the variability in enrichment (log AUC).

Sub-trajectory GIST grids were compared to the full simulation GIST grid (**Figure A.1.5**, top panel), and to neighboring sub-trajectory GIST grids (**Figure A.1.5**, bottom). The oscillating behavior in both curves (**Figure A.1.5**) indicates convergence.



**Figure A.1.5. Comparison of GIST grids from sub-trajectories.**

The combined GIST grid of solute-water enthalpy and water-water enthalpy scaled by two are evaluated here. Top, each sub-trajectory is compared to the full simulation. Bottom, each sub-trajectory is compared to its immediate neighbors.

We examined the variance of docking performance when using the sub-trajectory GIST grids (0.19 log AUC units, **Table A.1.3**). As a control, we looked at the variance by modifying the sampling (1.84 log AUC units, **Table A.1.3**). When compared to the modified sampling, the sub-trajectory docking varied little (9.6 times less). These data show that docking with the GIST grids of the 5 ns long simulations gave very similar docking results as the full 50 ns simulation (differing at most by 0.36 log AUC units).



**Table A.1.3.** Impact of modified sampling and subtrajectory on enrichment

Trajectory	Spheres	mean	std	max	min	diff
Sub <sup>a</sup>	original	58.51	0.19	58.76	58.12	0.64
Full <sup>b</sup>	original	58.40	--	--	--	--
Full	modified <sup>c</sup>	57.46	1.84	62.24	55.16	7.08

<sup>a</sup> 10 GIST grids generated from 5 ns sub-trajectory;

<sup>b</sup> One GIST grid from the 50 ns trajectory;

<sup>c</sup> 10 perturbed spheres

### Retrospective analysis for 25 DUD-E systems.

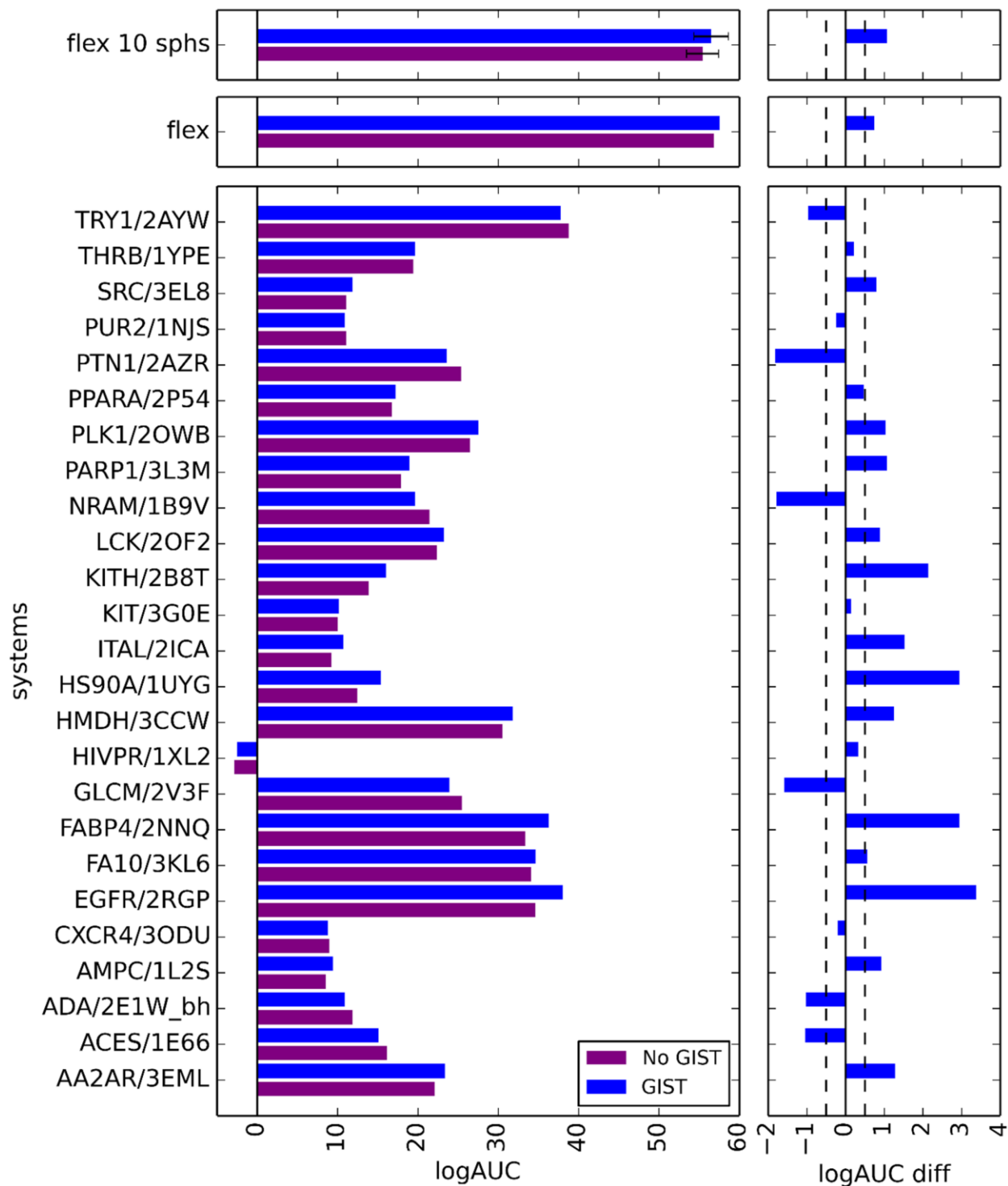
When comparing GIST to no-GIST results across the 25 DUD-E systems, GIST performs worse (average log AUC difference is -1.33, **Table A.1.4**), unlike CcP-ga which performs best with a weighting of -1.0. However, when we lower the weighting of GIST component to -0.5 the results got slightly better than the no-GIST enrichments (avg.  $\Delta \log \text{AUC} = 0.28$ , **Table A.1.4**). When examining the GIST grids, we observed extrema of very high energies at specific voxels. For example, ADA had the most extreme voxel of any system with a value of  $-119.73 \text{ kcal/mol}/\text{\AA}^3$  that if displaced would penalize the score by  $+14.97 \text{ kcal/mol}$ . Such a large penalty seems to be unreasonable in the context of our scoring. Thus, we truncate these peaks to  $\pm 3.0 \text{ kcal/mol}/\text{\AA}^3$  (which remains a high value, 5 to 19 fold higher than the standard deviation of the 210,000 voxels in the grid). This truncation impacts only 0.03% of the voxels, ranging from 17 to 88 for the favorable water voxels and 0 to 10 for unfavorable voxels. When truncation of extrema is combined with a weighting of -0.5 there is an additional improvement of GIST compared with no-GIST (avg.  $\Delta \log \text{AUC} = 0.53$ , **Table A.1.4**, **Figure A.1.6**). AA2AR and AMPC both change classification from same to better when

truncated grids are used, FXA likewise shifts but this is due to very slight change in log AUC. We believe that the extrema are artificially high due to the following: (1) The simulations are run with the protein's heavy atoms strongly restrained (5 kcal/mol/Å<sup>2</sup>). Since waters interact with the restrained atoms, their densities and energies are more concentrated than if the residue/atoms could move. The waters that are interacting with a moving atom would also move smearing the water's densities and energies across more voxels. (2) Entropy is neglected and the positions that have the highest energies are also those position where the waters are most frozen, so there is likely an entropic cost to having the water there.

**Table A.1.4.** DUD-E evaluation of GIST contribution on enrichment calculations.

Analysis of different weighting factors on enrichments. <sup>a</sup>				
	better	Same	worse	avg. ΔlogAUC
weight: -0.5	10	9	6	0.28
weight: -1.0	8	5	12	-1.33
weight: -2.0	5	4	16	-6.55
weight: -0.5, truncate 3.0	13	6	6	0.53
Weight: -1.0, truncate 3.0	11	3	11	-0.39

<sup>a</sup> Each row sums to the 25 systems.



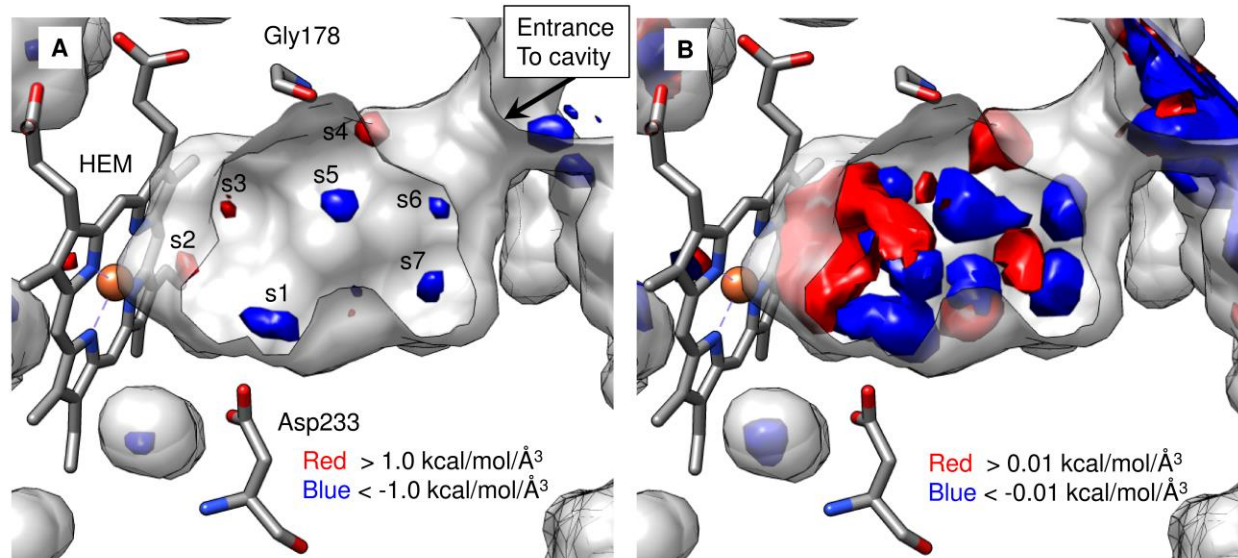
**Figure A.1.6. Enrichment analysis of CcP-ga and 25 DUD-E systems.**

Bar graphs of logAUC values for six docking types are shown: non-GIST in purple and GIST in blue (with the GIST component weighted by -0.5 and the GIST grids truncated at 3.0 kcal/mol/Å<sup>3</sup> results). The bottom panels show the total enrichment values for No-

GIST and GIST, while the top two panels show the difference (GIST - non-GIST). CcP results are shown for 10 perturbed results (error bars show standard deviation as an indication of the distribution of the results) and for the original sphere set. ADA was prepared by hand. All other systems were prepared with an automated procedure.

### **Binding site analysis.**

We examine the CcP-ga closed binding site to understand the nature of solvent in the site. In **Figure A.1.7** the enthalpy with water-water term scaled by two (Enthalpy2, eq A.1.5) is shown. The regions of unfavorable energy for waters ( $>1.0$  kcal/mol/Å<sup>3</sup>) are shown in red, which are favorable to displace according to the GIST scoring function. The favorable regions for water ( $>-1.0$  kcal/mol/Å<sup>3</sup>) are shown in blue, which are unfavorable to displace according the GIST scoring function. The favorable site (s1) proximal to Asp233, is the most favorable water location in the site. The region closest to the heme has two unfavorable water locations (s2 and s3) (**Figure A.1.7**). There is also an unfavorable location (s4) proximal to Gly178. Finally, there is a region close to the cavity entrance that encompasses three additional favorable water locations (s5, s6, and s7). Decreasing the cutoff value to 0.01 kcal/mol/Å<sup>3</sup> reveals the irregular shapes of the hydration sites (**Figure A.1.7**). Note that the majority of the solvation energy is concentrated at these seven sites. However, just accounting for the most intense sites (as WaterMap does) will neglect the lower magnitude regions, which do add up (-1.47, and +2.42, **Table A.1.5**) and contribute to the score.



**Figure A.1.7. Hydration of CcP-ga with the GIST enthalpy grid.**  
**A.** Here, GIST enthalpy grids with a cutoff of  $1.0 \text{ kcal/mol/\AA}^3$  are shown. The only opening to the closed cavity is indicated by an arrow. Seven hydration sites are indicated, s1 through s7. **B.** The cutoff value is decreased to 0.01.

**Table A.1.5. Site energetics of subregions.**

Subsite name	Energies (kcal/mol)
s1	-4.27
s2	2.58
s3	1.63
s4	1.67
s5	-2.36
s6	-2.20
s7	-1.22
Sum positive	5.88
Sum negative	-10.05
Whole site positive	8.30
Whole site negative	-11.52
Total	-3.22
Remainder positive	-1.47
Remainder negative	2.42

<sup>a</sup> Sites are spheres with a radius of 1.4Å located at the centers of intensities of the energies.

### Prospective testing.

The behavior of the 17 tested molecules (**Table A.1.6**) is presented in the following, including ranks and energies. Ligand occupancies are presented in **Table A.1.6**; for compound **14**, MES was not completely removed from the binding site and its partial occupancy is shown in **Figure A.1.8**. Ligand efficacy is determined from the affinity (**Figure A.1.9**) and ranges from -1.0 to -0.22. The ligands that make water-

mediated interactions with Asp233 on average bind more weakly than the molecules that bind with a direct electrostatic interaction (**Table A.1.7**).

From among those molecules substantially changing rank or pose due to including GIST, 17 were purchased for experimental testing. Compounds **3** to **14** were acquired and tested because their ranks improved with GIST, while compounds **15** to **17** were acquired and tested because of better ranks without the GIST term (**Table 1.1**). Molecules that ranked higher by GIST scored more favorably than without GIST by up to -1.8 kcal/mol, but could also be more unfavorable by as much as +2.0 kcal/mol out of a total docking score that ranged from -42.8 to -35.4 kcal/mol among the top-scoring 1000 molecules of VS1. The observation that GIST can improve ranks while reducing scores reflects its global effects on other high-ranking molecules that were affected more substantially still, emphasizing the role of decoy molecules in docking. For molecules whose rank was substantially better without GIST, the GIST term ranged from 8.1 to 8.7 kcal/mol (unfavorable), showing that GIST strongly disfavored these otherwise high-ranking molecules. We also looked for molecules where a substantial pose change occurred between the two scoring functions (e.g. compounds **1** and **2**, **Table 1.1**). Finally, we considered implicit water-mediated interactions to be favorable regions in the GIST grid within hydrogen-bonding distance to ligand and protein, though no explicit water molecules were used. This occurred with compounds **3**, **4**, **5**, and **6** (**Table 1.1**). We now consider the 14 molecules prioritized by including GIST (pro-GIST), and then turn to those 3 prioritized by excluding GIST (anti-GIST).

Intriguingly, GIST penalties on these deprioritized molecules, at around +8 kcal/mol, had a much stronger impact on reducing their ranks than favorable GIST

energies had on improving them; as with most scoring terms in docking, deprioritizing decoys is as or even more important than highly scoring what turns out to be true ligands.

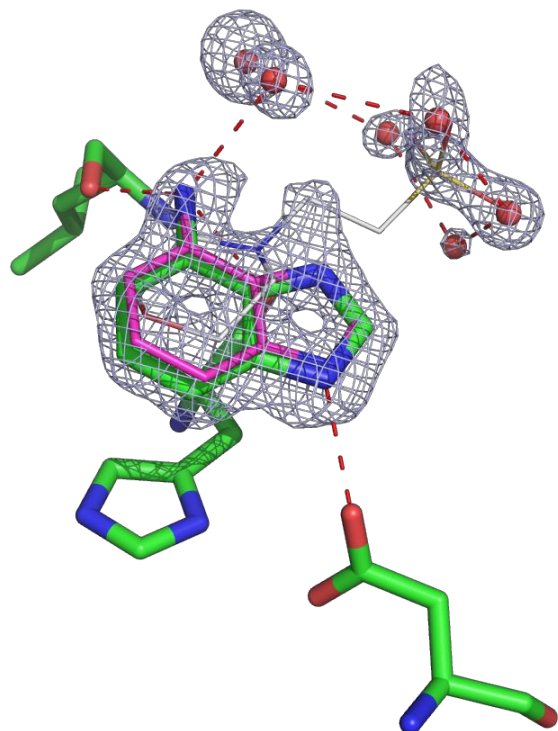
**Table A.1.6.** Detailed properties of selected molecules.

<b>Cmpd #</b>	<b>Name</b>	<b>Rank1 GIST</b>	<b>Rank2 Non-GIST</b>	<b><math>\Delta</math>logrank</b>	<b>RMSD <sup>a</sup></b>	<b><math>k_d</math> (<math>\mu</math>M) <sup>b</sup></b>
3	ZINC4705523	13	249	1.28	0	3472 $\pm$ 172
6	ZINC19439634	91	355	0.59	0	3435 $\pm$ 860
9	ZINC20357620	98	745	0.88	0	522 $\pm$ 21
4	ZINC6869116	112	464	0.62	0	809.7 $\pm$ 99
12	ZINC2389932	118	645	0.74	0	619 $\pm$ 63
13	ZINC39212696	147	1462	1	0	n.d.
11	ZINC161834	358	1212	0.53	0	1.3 $\pm$ 0.03
1	ZINC2564381	490	180	-0.43	3.21	n.d.
8	ZINC42684308	601	1916	0.5	0	1962 $\pm$ 554
--	ZINC95079390	615	2612	0.63	0	n.a.
2	ZINC6557114	664	740	0.05	3.17	154 $\pm$ 19
5	ZINC6855945	869	2550	0.47	0	1606 $\pm$ 287
7	ZINC1827502	5	19	0.58	0	113.7 $\pm$ 20.05
14	ZINC112552	747	4380	0.77	0	29.6 $\pm$ 2.5
10	ZINC74543029	1128	4923	0.64	0	$\sim$ 712 $\pm$ 231
ANTI-GIST						
17	ZINC22200625	6000	577	-1.02	0	n.d.
15	ZINC2534163	9487	906	-1.02	0	NB
16	ZINC156254	14828	1657	-0.95	0	5464 $\pm$ 2694 (NB)

<sup>a</sup> RMSD uses the Hungarian algorithm

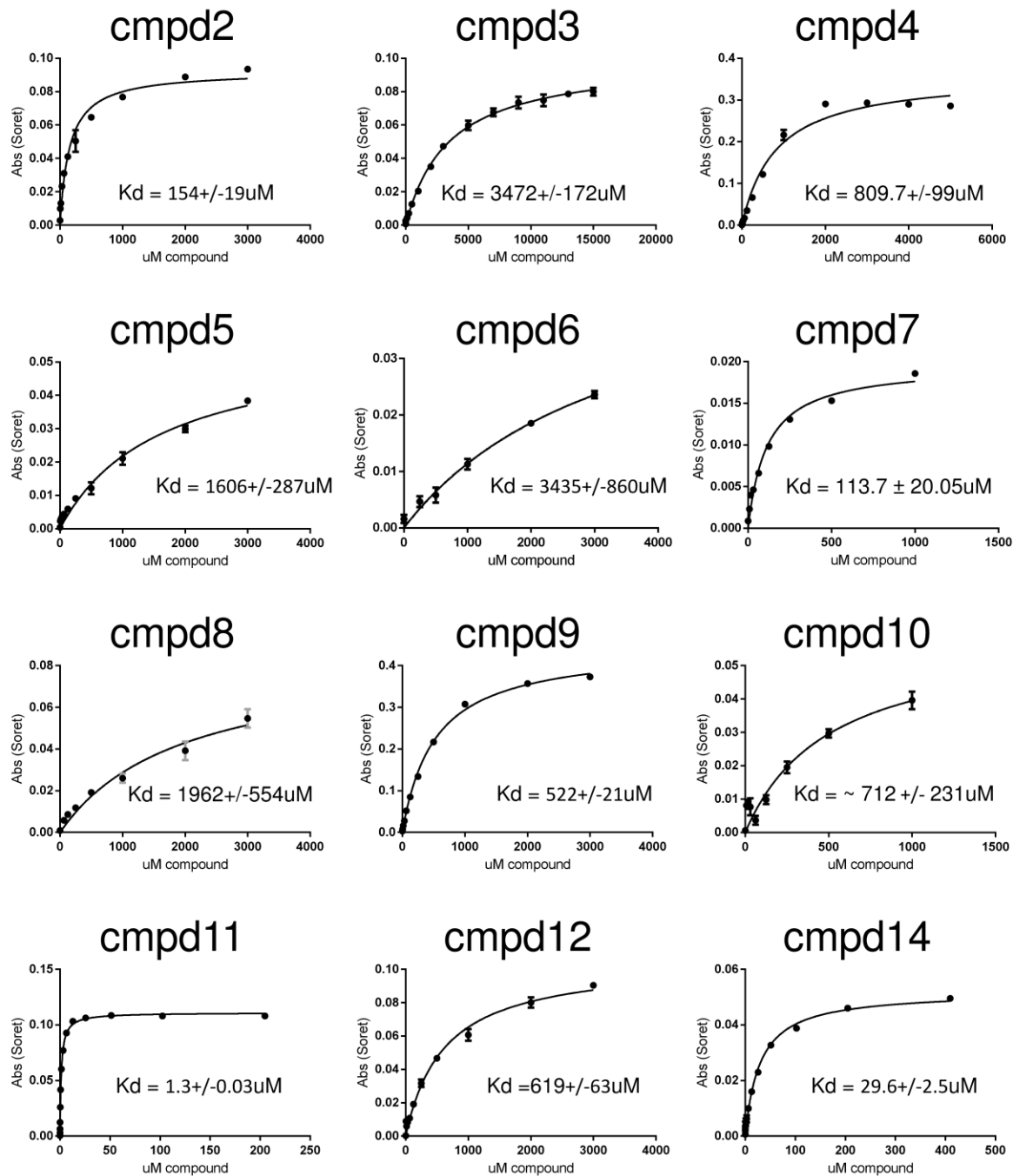
<sup>b</sup> n.a., not available - molecule not in assayable form. n.d., not determinable - compound interference with absorbance peaks. NB, non-binder. " $\sim$ ", assay interference of compound 10 before saturation was reached.





**Figure A.1.8. Compound 14 with MES.**

Compound **14** was refined to 73% in the presence of 26% MES from the crystallization buffer



**Figure A.1.9. Ligand binding curves.**

The Soret band shift is shown as a function of ligand concentration ( $\mu\text{M}$ ).

**Table A.1.7.** Ligand occupancies after automatic refinement.

<b>Cmpd #</b>	<b>Ligand occupancy</b>
<b>1</b>	0.88
<b>2</b>	0.90 (one conformation modeled)
<b>3</b>	0.93
<b>8</b>	0.92
<b>9</b>	0.90
<b>10</b>	0.92
<b>11</b>	0.87
<b>12</b>	0.93
<b>14</b>	0.73 (+MES @ 0.26)

While the occupancy for the major pose of **2** refined to 90%, the alternative pose would sterically clash with a nearby protein loop that has insufficient electron density to allow explicit modeling of alternative conformations.

**Table A.1.8.** Comparison of affinities for compounds with different interactions

<b>WM</b>		<b>NonWM</b>	
<b>Cmpd #</b>	<b>Affinity (<math>\mu</math>M)</b>	<b>Cmpd #</b>	<b>Affinity (<math>\mu</math>M)</b>
<b>1</b>	n.d.	<b>2</b>	154
<b>3</b>	3472	<b>7</b>	114
<b>4</b>	810	<b>8</b>	1962
<b>5</b>	1606	<b>11</b>	1
<b>6</b>	3435	<b>12</b>	619
<b>9</b>	522	<b>14</b>	30
<b>10</b>	712		
<i>average</i>	1759.5	<i>average</i>	480
<i>median</i>	<b>1208</b>	<i>median</i>	<b>134</b>

## Timings.

The GIST-scoring algorithm is more time- and memory-intensive than trilinear interpolation, which is used in the other scoring components. To determine how GIST affects the speed of docking calculations, we ran one set of ligands from each system ten times on the same, dedicated machine (**Table A.1.9**). This results in a 1.5 to 16.4 times (on average six-fold) slowdown in runtime. However, we anticipate that using good GIST approximations will result in no slowdown and little impact on docking quality.

**Table A.1.9.** DOCK3.7 run time slowdown with GIST referenced to non-GIST.

<b>PDB code</b>	<b>DUD-E name</b>	<b>Avg number of heavy atoms</b>	<b>Slowdown <sup>a</sup></b>
1B9V	NRAM	25.34	3.87
1E66	ACES	29.48	2.21
1L2S	AMPC	20.19	5.21
1NJS	PUR2	33.29	7.30
1UYG	HS90A	27.95	4.67
1XL2	HIVPR	41.06	4.37
1YPE	THRB	34.88	7.97
2AYW	TRY1	33.66	16.40
2AZR	PTN1	39.93	12.97
2B8T	KITH	30.24	3.14
2E1W	ADA	24.77	3.78
2ICA	ITAL	36.38	13.27
2NNQ	FABP4	30.30	4.27
2OF2	LCK	34.70	9.34
2OWB	PLK1	33.08	6.76
2P54	PPARA	32.18	2.92
2RGP	EGFR	31.39	4.45

PDB code	DUD-E name	Avg number of heavy atoms	Slowdown <sup>a</sup>
2V3F	GLCM	27.26	1.15
3CCW	HMDH	36.66	4.05
3EL8	SRC	34.62	4.43
3EML	AA2AR	31.97	2.65
3G0E	KIT	38.77	2.44
3KL6	FA10	33.52	9.94
3L3M	PARP1	30.30	5.34
3ODU	CXCR4	26.67	5.16
	CcP-ga	12.01	1.51
Average		31.18	5.75

<sup>a</sup> Slowdown = (timing from GIST docking) / (timings from non-GIST docking)

## Supplemental Methods.

**Experimental affinities and structures.** The protein was purified and crystallized as described. To reach high ligand occupancies, crystals were transferred into increasing ligand concentrations up to 100 mM (compound solubility permitting) and soaked for several minutes in each drop containing 25% 2-Methyl-2,4-pentanediol (MPD) as a cryoprotectant.

Diffraction images of flash-frozen crystals were collected at beamline 8.3.1. at the Advanced Light Source, Berkeley CA, and processed automatically with the Xia2 pipeline. Initial phases were obtained by Phaser molecular replacement using a model structure lacking several flexible residues and the loop region (residues 186-194). To avoid bias these regions were also excluded from early rounds of refinement using phenix.refine. The ligand and binding site water molecules were only added in the final

stage of crystallographic refinement and their occupancies were set to a value below 1 to automatically refine to their final values via phenix.refine without manual intervention. Ligand restraint dictionaries were generated from SMILES strings via phenix.elbow, using either automatic or CSD-Mogul geometry optimization. Composite 2mFo-DFc OMIT maps excluding the ligand fraction were calculated using phenix.composite\_omit\_map and converted to 2mFo-DFc FFT maps in ccp4 format in order to generate figures using PyMOL.

Crystallographic models were tested with phenix, Coot and the PDB validation tool before depositing the protein-ligand complexes at the PDB as 5U60 (1), 5U5W (2), 5U5Z (3), 5U61 (8), 5U5Y (9), 5UG2 (10), 5U5X (11), 5U5U (12), 5U5V (14) (**Table A.1.7**).

Experimental affinities were measured at least in duplicate by monitoring the shift of the heme Soret band on ligand binding and plotted using a one-site binding least squares fitting method (GraphPad Prism 6.03).

**Preparing the receptor for MD.** The protein preparation is described in the main text's method section, but further details are explained here. The proteins were assigned FF12SB (CcP-ga protein) or FF14SB (all DUD-E proteins) force field parameters. At the time CcP-ga simulations were run, the FF14SB parameters were not yet released. The proteins were placed in a box of TIP3P waters such that every atom of protein was 10 Å from the boundary of the box. The number of waters is presented in **Table A.1.10**. For CcP-ga (4NVA, the apo structure), ten crystallographic waters were retained for the simulation. No crystallographic waters were retained for

the simulations of the DUD-E systems. For CcP-ga, use of these crystallographic waters alters the GIST grids, particularly for occluded water locations. Some cofactors and structural ions were kept and disulfide bonds were defined (**Table A.1.10**).

Tutorials, which describe (1) running Molecular dynamics for GIST grid generation; and (2) docking with GIST grids, are available at

[\[http://wiki.docking.org/index.php/DOCK\\_3.7\\_with\\_GIST\\_tutorials\]](http://wiki.docking.org/index.php/DOCK_3.7_with_GIST_tutorials).

For CcP-ga, the heme force field was downloaded from the web. The heme parameters were originally prepared for hemoglobin and myoglobin, and thus needed to be adapted for Cytochrome c Peroxidases. The heme parameters were modified by adding a positive charge to the iron (iron Fe III has a 1.25 charge). Amber preparation (prep and frcmod) files for the heme are available at

[https://github.com/tbalius/GIST\\_DX\\_tools](https://github.com/tbalius/GIST_DX_tools).

**Table A.1.10. CcP-ga and DUD-E simulation details**

<b>Protein name</b>	<b>PDB code</b>	<b>Residues</b>	<b>Waters</b>	<b>Atoms</b>	<b>Ions / cofactor / disulfides / capping groups <sup>a</sup></b>
CcP-ga	4NVA (closed)	290	11,013	4614	Heme
AA2AR	3EML	290	14514	4569	Disulfides, caps
ACES	1E66	532	16481	8346	Disulfides
ADA	2E1W	349	9775	5536	ZN
AMPC	1L2S	358	12080	5581	
CXCR4	3ODU	306	15546	4988	Disulfides, caps
EGFR	2RGP	257	12374	4120	Caps
FA10	3KL6	282	13069	4331	Disulfides
FABP4	2NNQ	131	5372	2059	
GLCM	2V3F	497	14611	7765	Disulfides, caps
HIVPR	1XL2	198	7841	3128	
HMDH	3CCW	842	36285	12608	
HS90A	1UYG	209	8014	3295	
ITAL	2ICA	179	6917	2901	
KIT	3G0E	332	13892	5298	
KITH	2B8T	206	11994	3290	
LCK	2OF2	271	12925	4392	
NRAM	1B9V	391	11140	5979	Disulfides, Ca ion
PARP1	3L3M	348	12689	5510	
PLK1	2OWB	294	16083	4828	
PPARA	2P54	267	11020	4282	



Protein name	PDB code	Residues	Waters	Atoms	Ions / cofactor / disulfides / capping groups <sup>a</sup>
PTN1	2AZR	297	12120	4811	
PUR2	1NJS	200	9464	3056	
SRC	3EL8	263	9783	4200	Caps
THRB	1YPE	250	8567	4023	Disulfides, caps
TRY1	2AYW	223	8042	3221	Disulfides

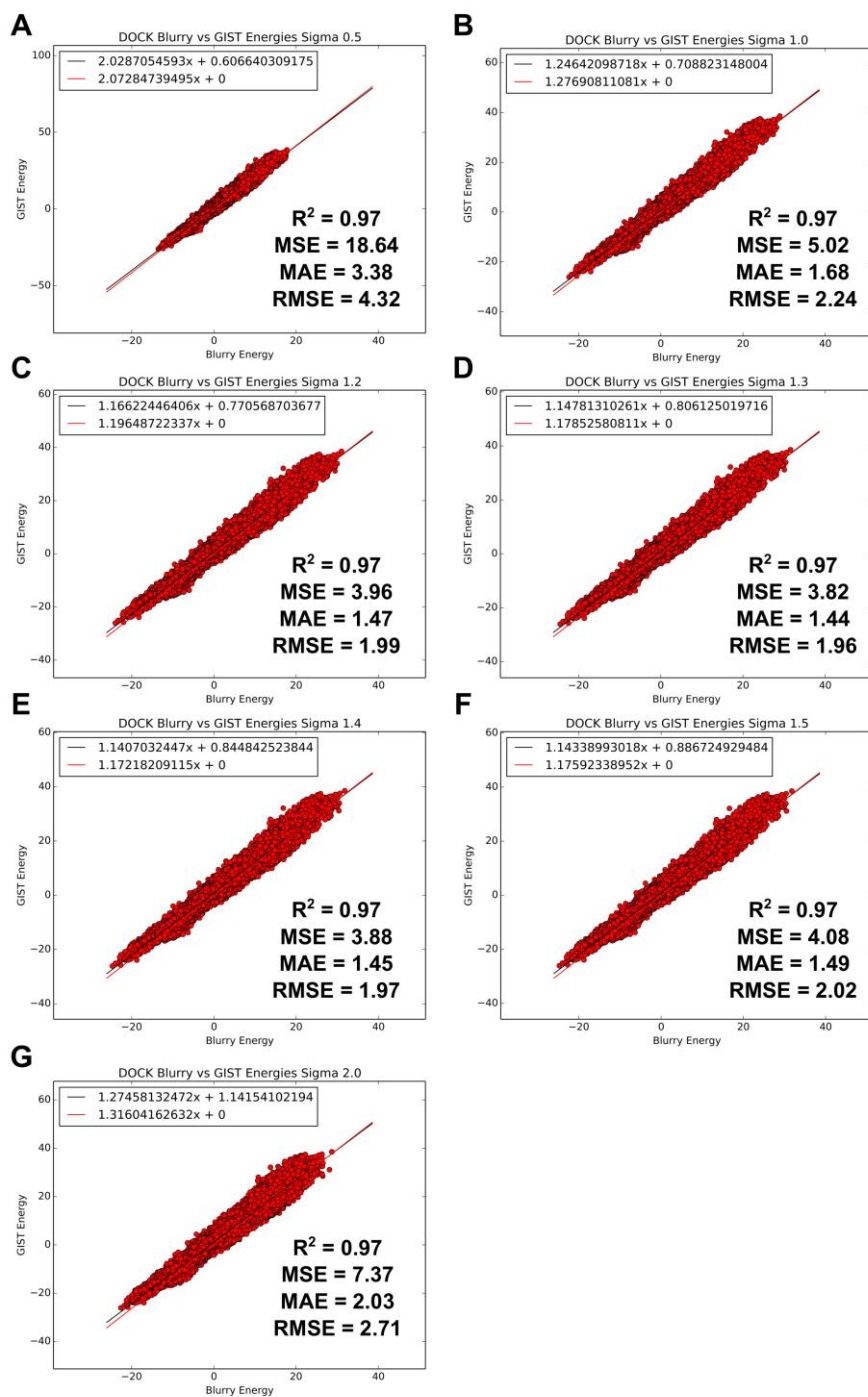
<sup>a</sup> NME and ACE were added to cap breaks (missing residues).

**Docking.** Scripts and programs in the DOCK3.7 distribution were used to prepare the receptors and ligand databases for docking and to carry out the library screens. *Blastmaster.py* was used to prepare the protein: hydrogens were added with *Reduce*, spheres were generated with *sphgen* and by converting the crystallographic ligand atoms to spheres (spheres are used to orient molecules into the binding site); electrostatic grids were generated by solving the Poisson-Boltzmann equation with the *Qniff* program; van der Waals grids were calculated using *Chemgrid*, the ligand desolvation grids were produced with *solvmap*, all distributed within the DOCK3.7 program suite. A GIST component to the scoring function was integrated in a new release of DOCK3.7 (**Figure A.1.2**). Default parameters were otherwise used for docking. CcP-ga was prepared as a flexible receptor with 16 different conformations, as described. All other systems used a single receptor conformation. To use GIST, proteins were aligned using Chimera into the simulation's frame of reference before DOCK preparation.

**Enrichment calculations.** Log AUC is described in Mysinger and Shoichet. We specify a lower bound of 0.001 FPR to avoid infinitely negative values of  $\log(0)$ . The maximum area under the curve is 3, we then convert this value to a percent (maximum area) and subtract the area under the random curve. Thus, Log AUC ranges from -14.5 to 85.5 where 0 is random and anything above 0 is better than random, and below, worse. Note that these values will change for other lower bounds (the lambda parameter in Mysinger et al.). The CcP-ga ligand databases were generated as described below at pH 4, while the DUD-E databases were obtained from the Autodude webpage (<http://autodude.docking.org>). Protein structures were prepared for docking described above (docking section).

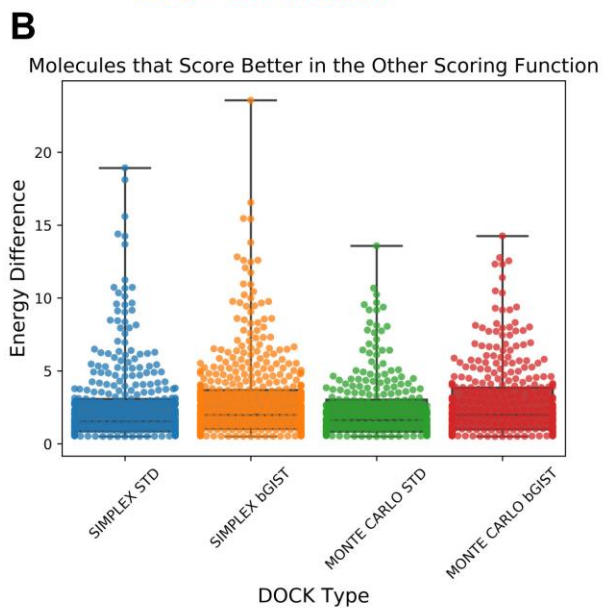
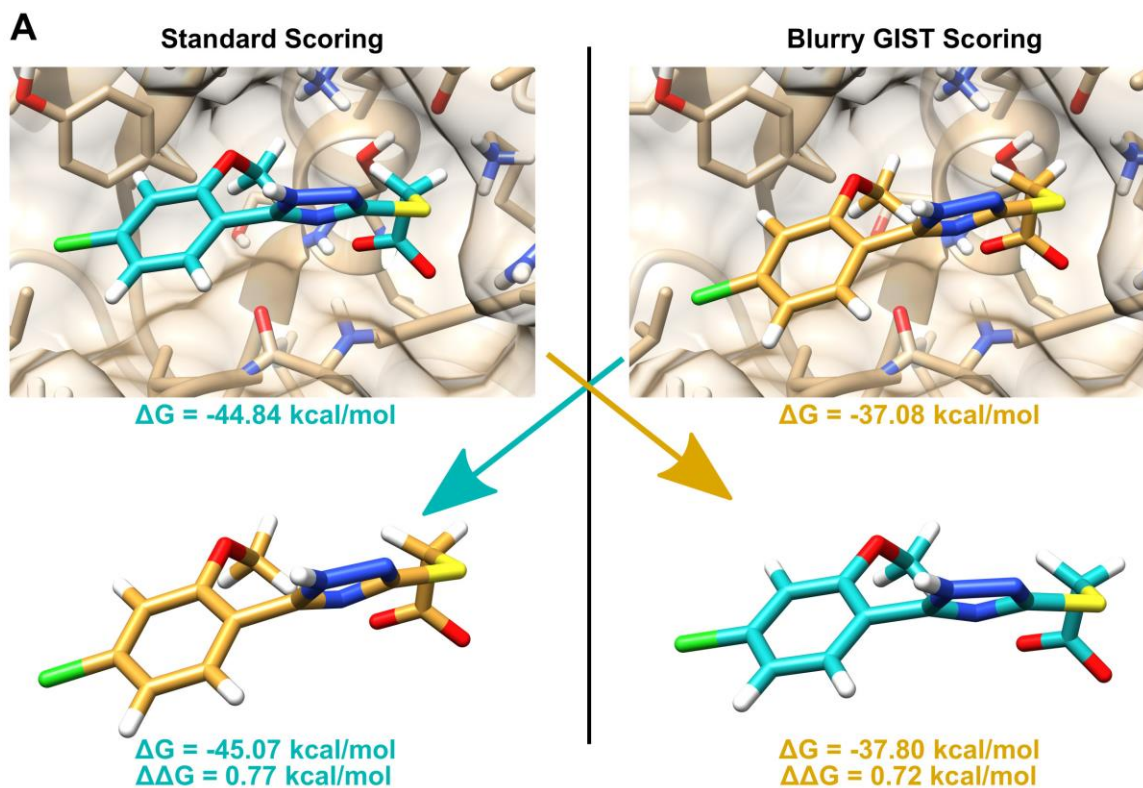
**Database generation.** The databases were generated using the DOCK3.7 ligand generation pipeline. ChemAxon (molconvert) was used to generate a 3D molecule from SMILES. The protonated states of the ligands are generated using Marvin of ChemAxon. Protonation states of the molecule were generated at pH 4.0 (greater than 20% occupancy). AMSOL7.1 was used to calculate the partial charges and per atom decomposition of ligand desolvation, Openeye Omega was used to generate an ensemble of conformations of each ligand. These conformations are stored in db2 format using the db2 generation program distributed with DOCK 3.7. Ligand databases downloaded from ZINC15 used the same pipeline but were generated at pH 6.4.

## A2. Supplementary Material for Chapter 2



**Figure A.2.1. Correlations between GIST energies.** Roughly 297,000 ligand and decoy poses from the 40 DUD-E systems were rescored outside of DOCK using the displacement GIST scoring scheme and the blurry GIST scoring scheme for sigma ( $\sigma$ )

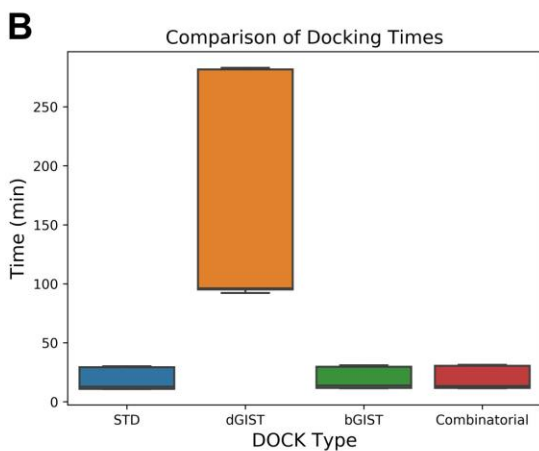
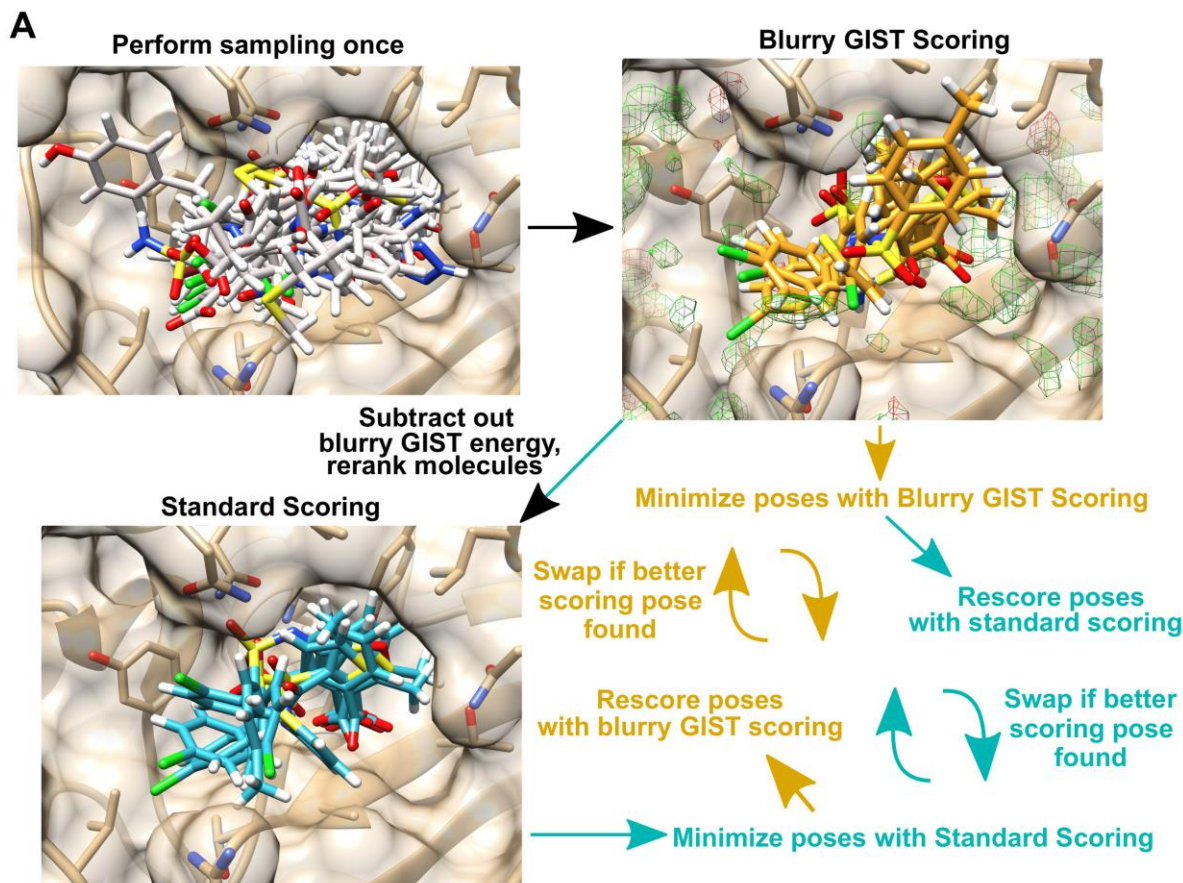
values of pseudo-atom radius / 0.5 (A), pseudo-atom radius / 1.0 (B), pseudo-atom radius / 1.2 (C), pseudo-atom radius / 1.3 (D), pseudo-atom radius / 1.4 (E), pseudo-atom radius / 1.5 (F), pseudo-atom radius / 2.0 (G). The pseudo-atom radius is 1.0 Å for hydrogen atoms and 1.8 Å for heavy atoms. Line equations,  $R^2$  values, mean absolute errors (MAE), mean squared error (MSE) and root mean squared error (RMSE) are reported.



**Figure A.2.2. Insufficient minimization scrambles best scoring poses.**

A) Two different poses are reported as the best scoring pose for this specific molecule. However, the standard pose scores better for the blurry GIST scoring function, and the blurry GIST pose scores better for the standard scoring function

with DOCK energy differences of 0.72 kcal/mol and 0.77 kcal/mol, respectively. B) Hundreds of molecules exhibit this behavior for the 3000 molecule AmpC DUD-E set after docking for Simplex minimization and Monte Carlo optimization, with some of these energy differences rising over 20 kcal/mol. Temperature for Monte Carlo optimization was set at 1 K.

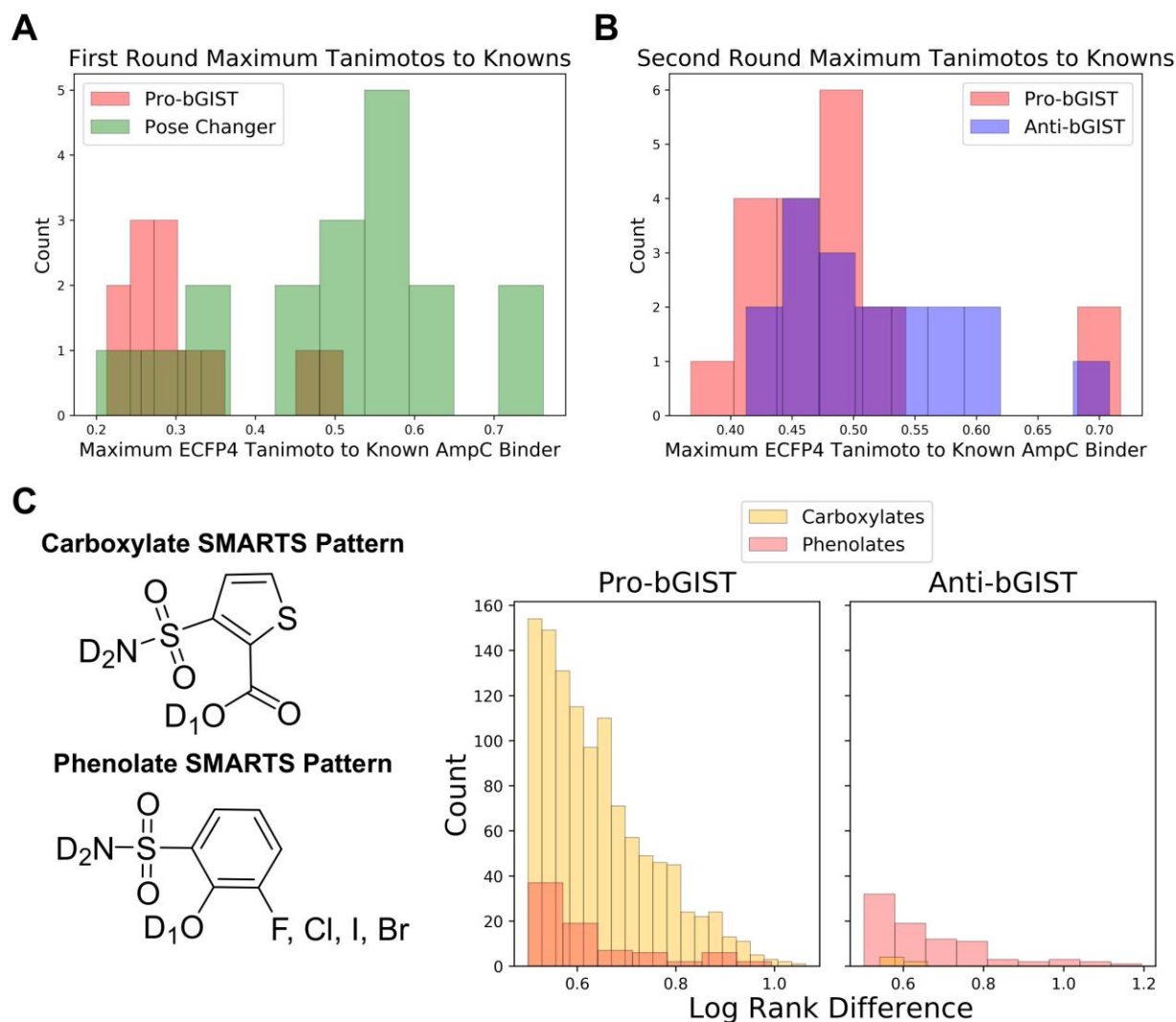


**Figure A.2.3. A new scoring scheme fixes insufficient minimization.**

In the previous implementation of GIST, we performed two screens – one with the standard scoring function, one with the GIST scoring function – where the exact same sampling is performed twice. A) In this new scheme, the sampling is only done once. Molecules are first scored for the blurry GIST scoring function and sorted by energy.

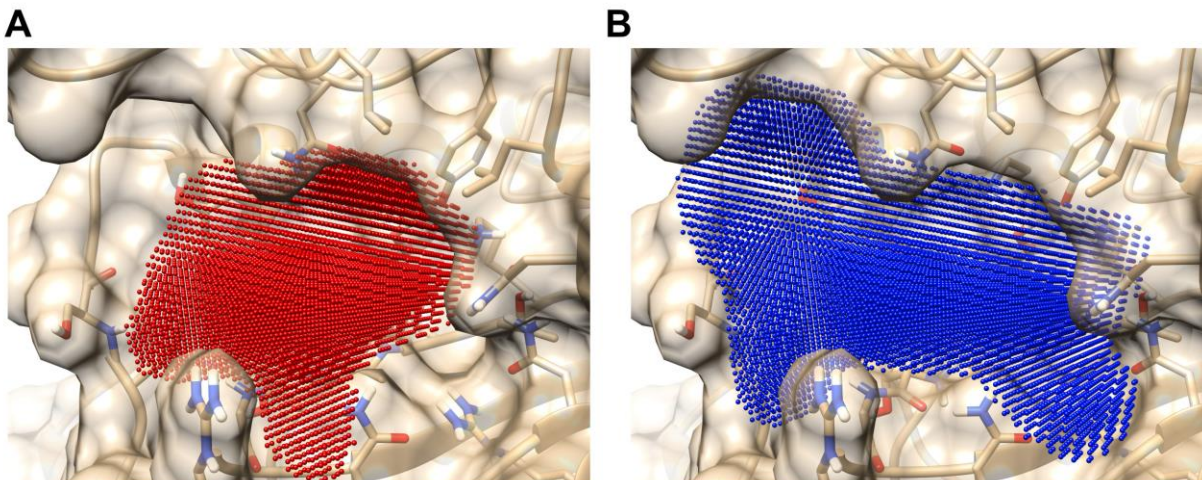
These blurry GIST poses are minimized with the blurry GIST scoring function. To obtain the standard scoring function poses, the blurry GIST score is subtracted from the poses initially scored by blurry GIST. These standard poses are then sorted by the standard energy and minimized with the standard scoring function. The minimized poses from both scoring functions are then rescored with the other scoring function, and if a better energy pose is found, that pose now becomes the best scoring pose for that scoring function. In this case, it does not matter which scoring function generated the pose, as all poses generated are scored with both scoring functions and each scoring function takes its best scoring pose. B) Docking of roughly 2,000 molecules to AmpC with nine replicates. Combinatorial docking performs with the same speed as the standard or blurry GIST scoring functions alone, but produces the output of both, thus cutting the docking time in half.





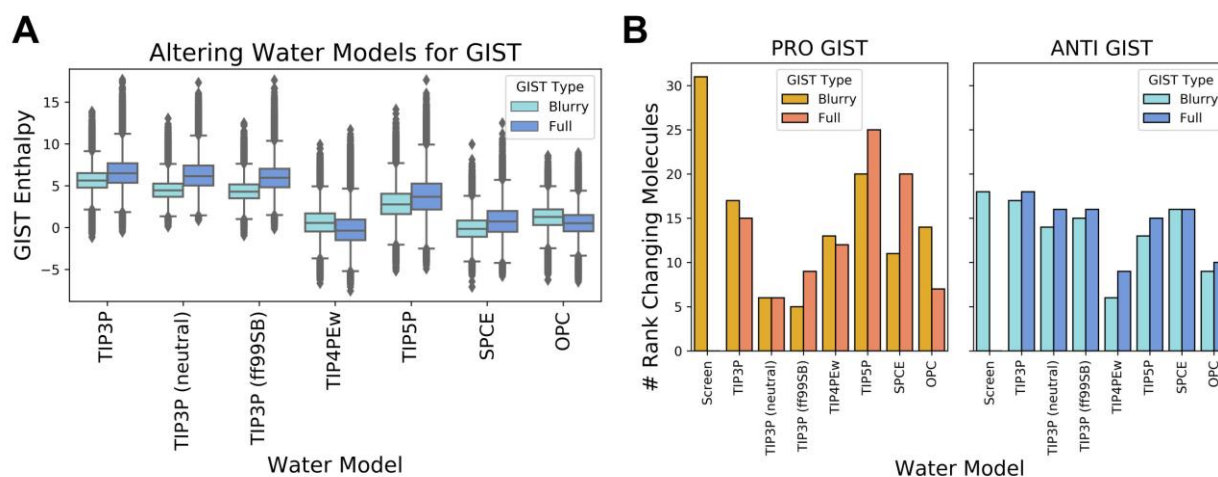
### Figure A.2.4. Choosing molecules similar to known AmpC inhibitors

A) ECFP4 Tanimoto coefficients to known AmpC inhibitors for pro-bGIST and pose-changing molecules from the first round of testing. B) ECFP4 Tanimoto coefficients to known AmpC inhibitors for pro-bGIST and anti-bGIST molecules from the second round of testing. C) Molecules with the carboxylate and phenolate SMARTS patterns were retrieved from ZINC15, docked, and resorted into the original docking hit lists. Molecules were purchased from this subset. This included 1129 carboxylates and 79 phenolates that were prioritized by blurry GIST (pro-bGIST) and 6 carboxylates and 85 phenolates that were deprioritized by blurry GIST (anti-bGIST).



**Figure A.2.5. Volume occupation of pro- and anti-bGIST molecules**

A) Most frequently displaced voxels from 154,256 pro-bGIST molecules (A) and 159,071 anti-bGIST molecules (B). Voxels were counted if they were contained within the van der Waals radii of a molecule's pose and then binned based on frequency of displacement.

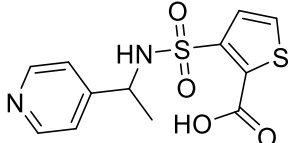
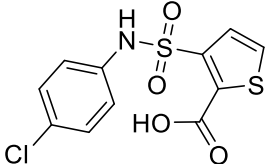
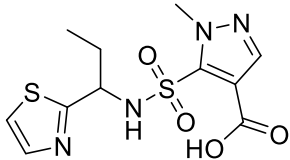
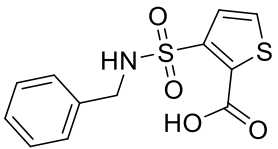
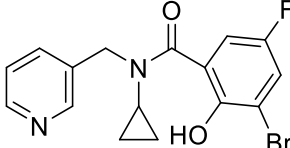
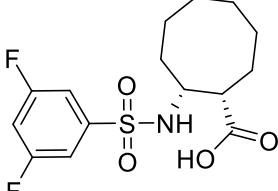
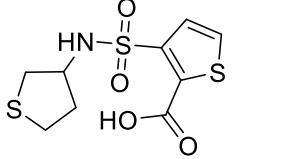
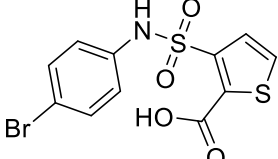
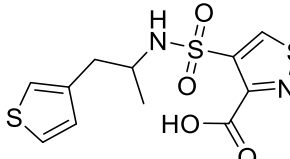
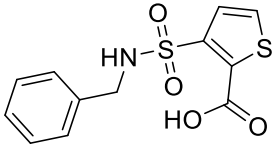
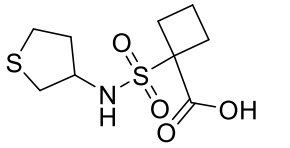
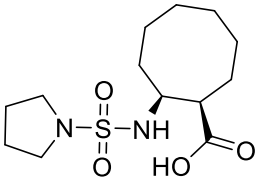


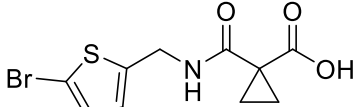
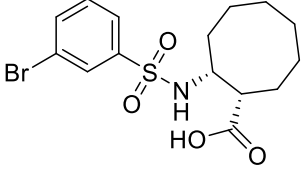
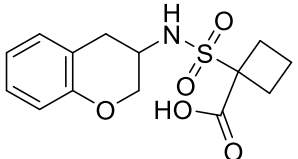
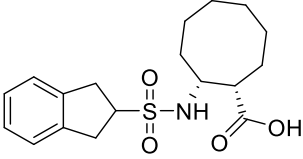
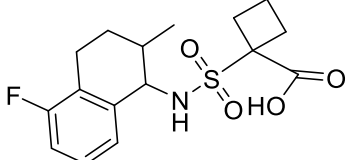
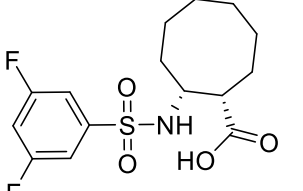
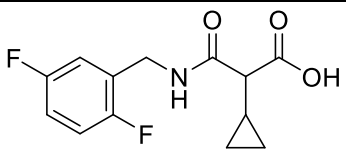
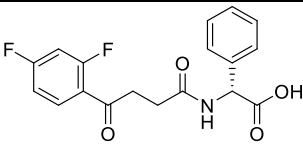
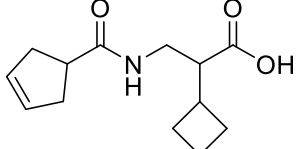
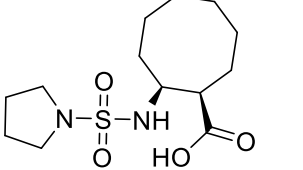
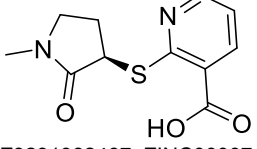
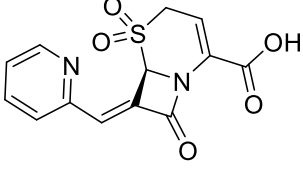
**Figure A.2.6. Parameter and solvent choice do not affect rank changing molecules.**

A) The 50ns molecular dynamics simulation was initially performed with the TIP3P solvent model and ff14SB force field, but was extended to a neutralized TIP3P setup with 3 chloride ions, TIP3P with the ff99SB force field, TIP4PEw, TIP5P, SPCE, and OPC solvent models. The GIST enthalpies show the medians and interquartile ranges after rescoring the top 150,000 poses outputted from the blurry GIST scoring function screen using the displacement (Full) or blurry GIST (blurry) scoring schemes using rescoring scripts. B) Number of molecules that change ranks (pro- or anti-bGIST)

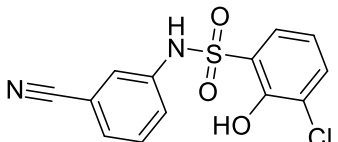
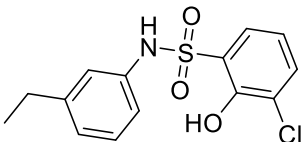
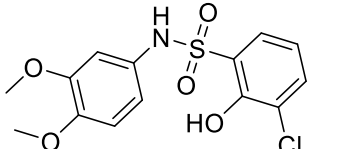
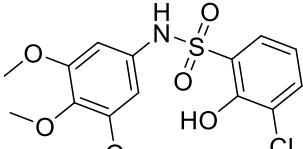
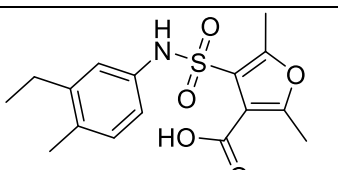
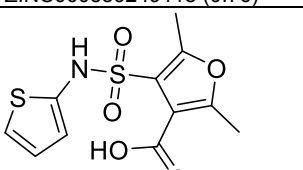
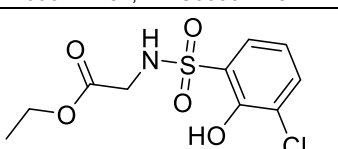
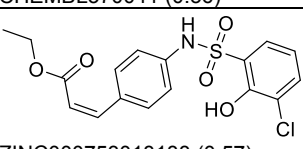
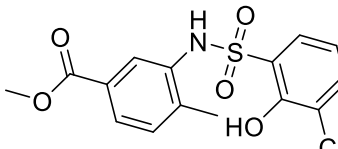
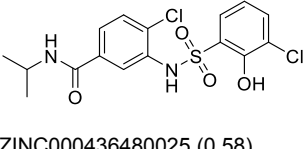
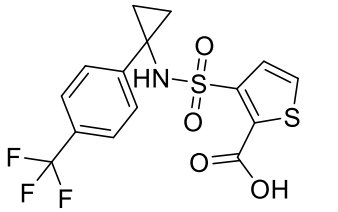
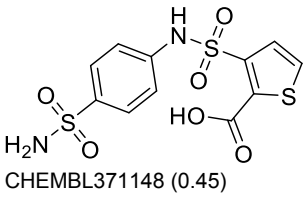
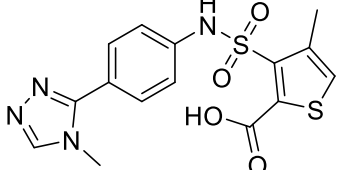
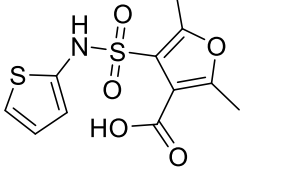
with a 0.5 log order rank difference after rescoring the top 150,000 poses outputted from the blurry GIST scoring function screen with different molecular dynamics water models and parameter choices. Even after altering the parameter choices, the same molecules that were chosen from the screen ("Screen") tend to have 0.5 log order rank differences and would have been chosen again. This suggests that the choice of parameters in the MD simulation is unlikely to have changed our results substantially.

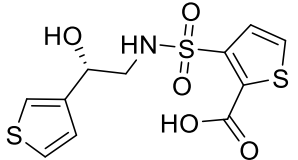
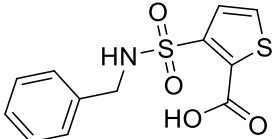
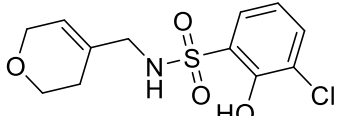
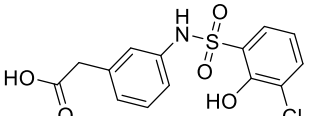
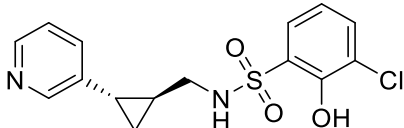
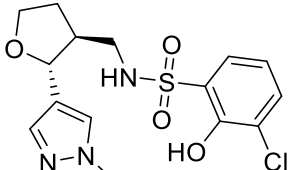
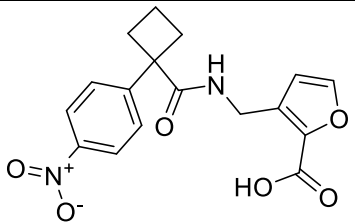
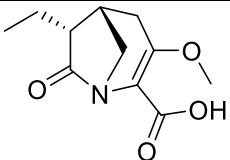
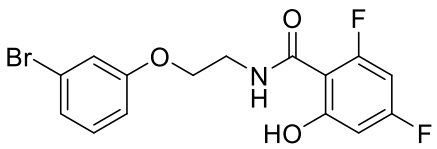
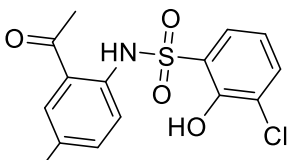
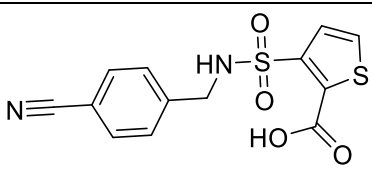
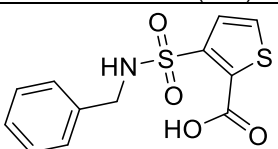
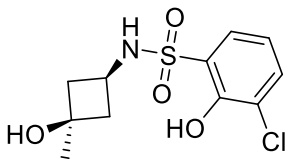
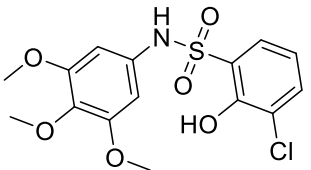
**Table A.2.1. All molecules tested against AmpC.**

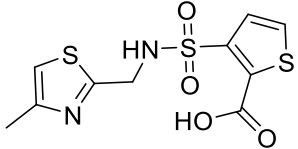
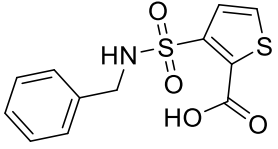
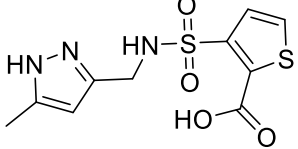
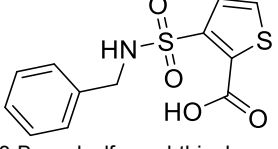
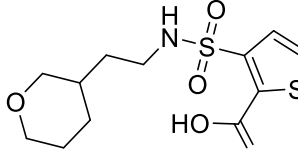
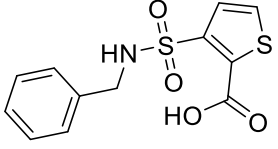
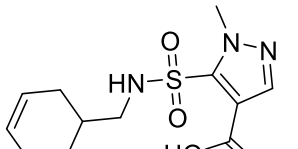
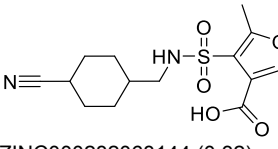
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
<b>PRO GIST</b>					
 Z227878108, ZINC000035126609	31.04	4947	1469	0.53	 3-(4-Chloro-phenylsulfamoyl)-thiophene-2-carboxylic_acid (0.48)
 Z2721494698, ZINC000530155910	9.08	35000	6345	0.74	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.26)
 Z1835990482, ZINC000436994974	1.13	35724	8940	0.6	 ZINC000580868636 (0.25)
 Z2437416709, ZINC000231345804	38.12	38046	7900	0.68	 3-(4-Bromo-phenylsulfamoyl)-thiophene-2-carboxylic_acid (0.51)
 Z2437289226, ZINC000516925327	16.76	38055	9121	0.62	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.30)
 Z2903948290, ZINC000905038806	4.13	51683 / 512614	15297 / 63688	0.53 / 0.91	 ZINC000547933290 (0.27)

Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z1614639933, ZINC000070600835	27.12	63066	14354	0.64	 ZINC000548260732 (0.21)
 Z2903947513, ZINC000905035036	11.13	90516	17239	0.72	 ZINC000681580748 (0.35)
 Z2903948616, ZINC000905040387	65.21	127809	27133	0.67	 ZINC000580868636 (0.29)
 Z2607540718, ZINC000663035453	-4.72	203696	44934	0.66	 ZINC000237985875 (0.33)
 Z2607839654, ZINC000663152888	-0.47	295024	47632	0.79	 ZINC000547933290 (0.23)
 Z3231982467, ZINC000071182697	-0.9	1973615	382284	0.71	 CHEMBL289526 (0.29)

**POSE CHANGERS**

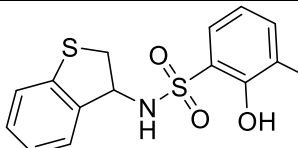
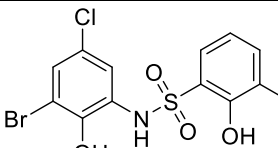

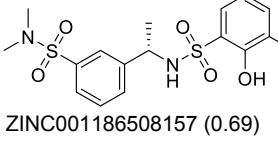
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	RMSD	Closest Known (ECFP4 Tanimoto Coefficient)
 Z2027054565, ZINC000339208618	49.23	81	84	1.4	 ZINC000581714578 (0.61)
 Z2027054051, ZINC000339202812	81.77	296	244	1.5	 ZINC000559249118 (0.76)
 Z1993712482, ZINC000324284771	98.61	865	1047	1.1	 CHEMBL370041 (0.56)
 Z2773172198, ZINC000230467629	16.71	871	524	2.0	 ZINC000753018188 (0.57)
 Z2027055215, ZINC000550110611	99.6	1496	2051	2.7	 ZINC000436480025 (0.58)
 Z2476040032, ZINC000650447472	80.6	2072	1012	3.9	 CHEMBL371148 (0.45)
 Z1796548044, ZINC000327717846	61.61	2151	1642	2.0	 CHEMBL370041 (0.37)

Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	RMSD	Closest Known (ECFP4 Tanimoto Coefficient)
 Z1971843402, ZINC000231978561	27.28	3540	2180	2.0	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.55)
 Z2774472128, ZINC000632389912	40.17	3728	1872	2.3	 ZINC000339204163 (0.43)
 Z2851435096, ZINC000641595024	65.46	4796	4824	3.7	 ZINC000632456968 (0.52)
 Z2755451606, ZINC000600870692	40.19	7849	10101	4.1	 CHEMBL84953 (0.2)
 Z910652810, ZINC000066048697	-1.1 (50uM)	9573	38827	2.8	 ZINC000753016232 (0.30)
 Z3228473727, ZINC000037748240	57.11	9793	5631	3.7	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.71)
 Z2774693635, ZINC000632470191	0.21	10945	5008	1.4	 ZINC000559249118 (0.49)

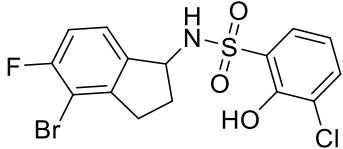
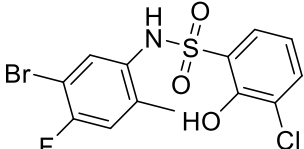
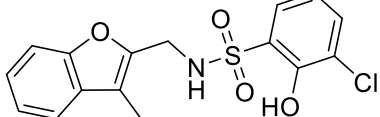
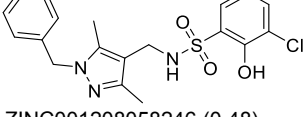
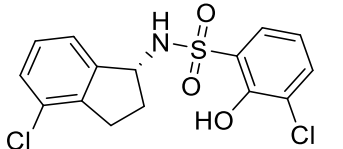
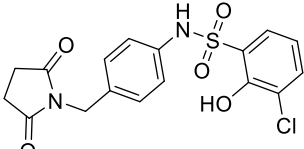
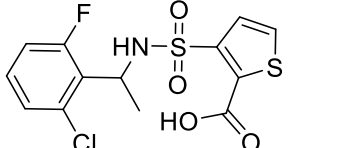
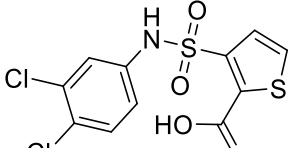
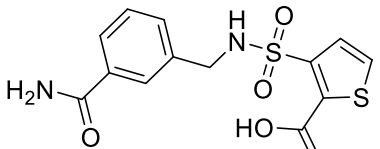
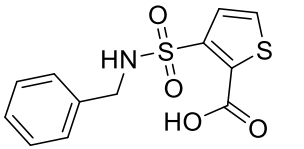
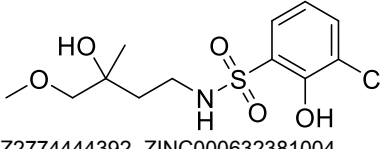
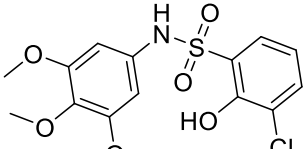
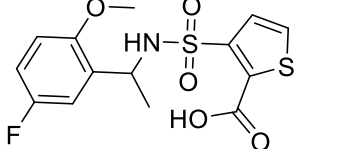
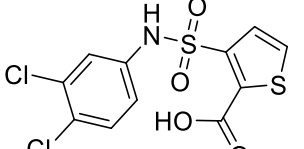
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	RMSD	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3226605788, ZINC000038090806	57.15	11707	4879	2.2	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.59)
 Z2827899976, ZINC000716800583	15.92	14586	6715	2.6	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.60)
 Z2721503949, ZINC000530153418	37.28	15422	8091	2.1	 3-Benzylsulfamoyl-thiophene-2-carboxylic_acid (0.48)
 Z2721488292, ZINC000530149216	4.99	15595	7477	1.9	 ZINC000282068144 (0.32)

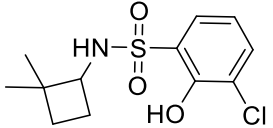
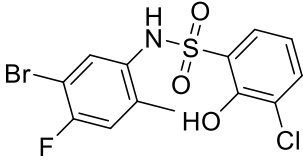
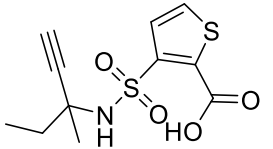
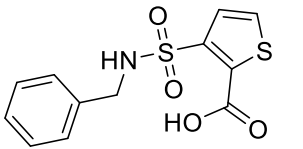
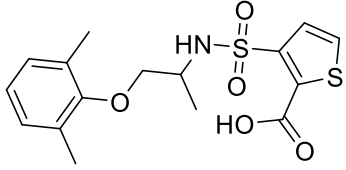
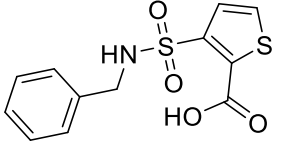
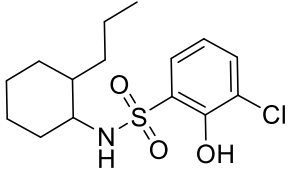
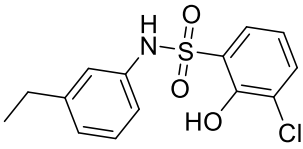
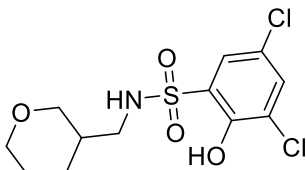
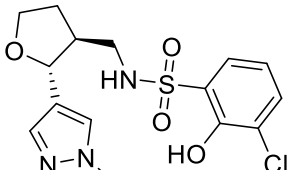
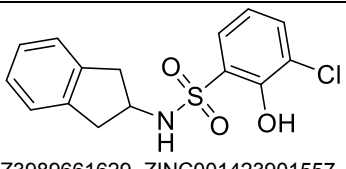
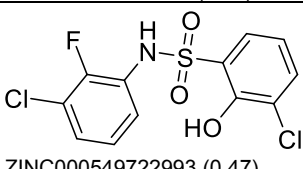
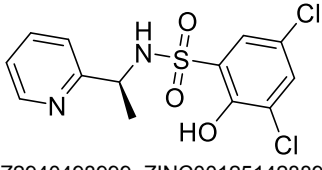
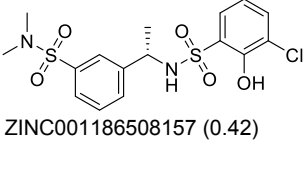
## SECOND ROUND

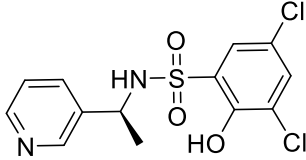
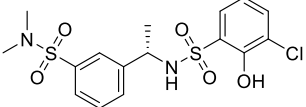
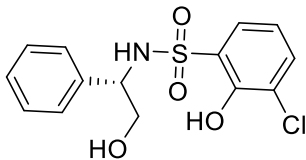
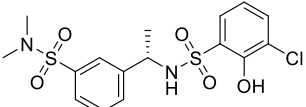
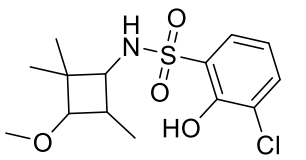
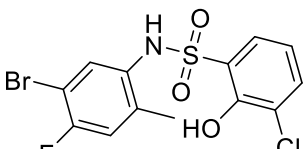
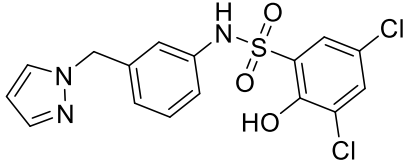
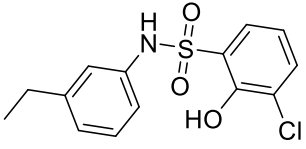
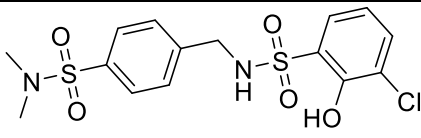
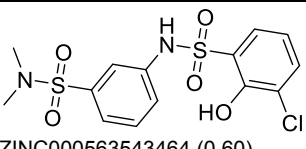
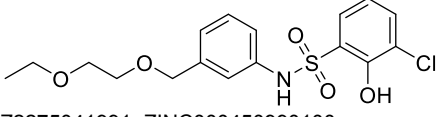
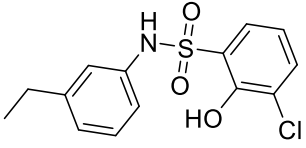
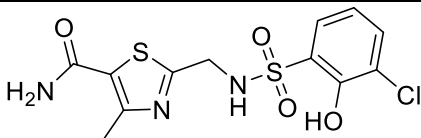
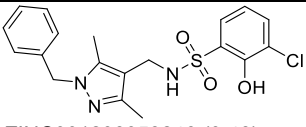
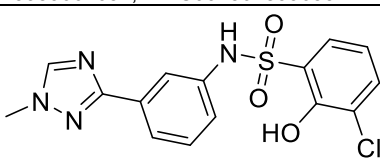
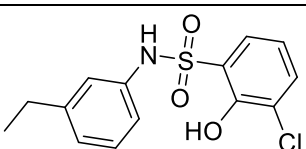
### PRO-bGIST

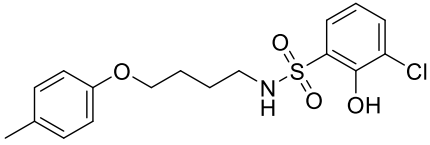
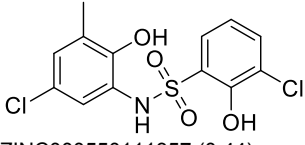
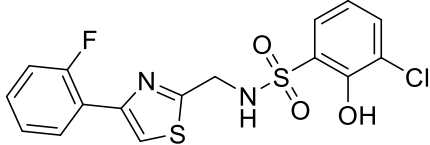
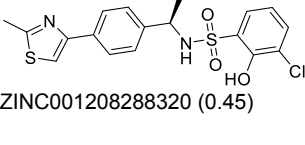
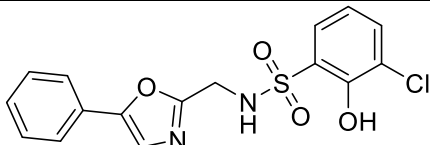
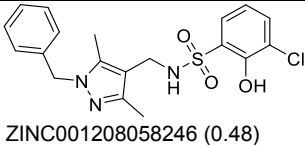
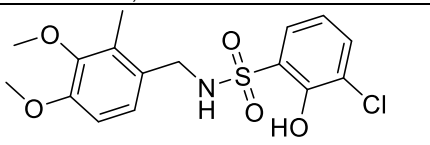
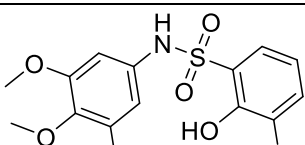
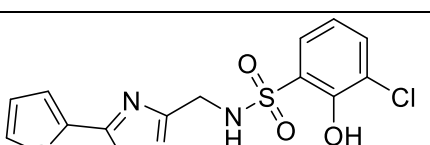
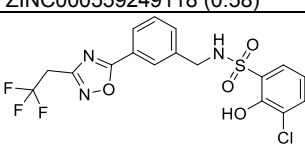
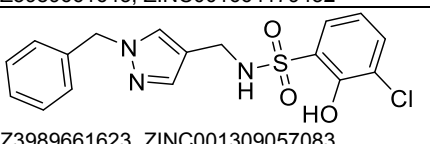
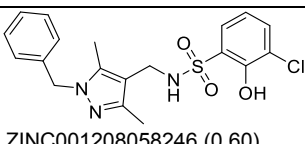
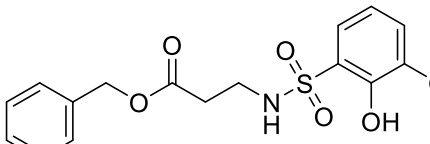
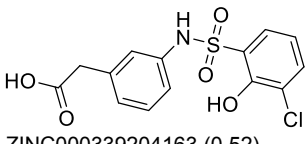
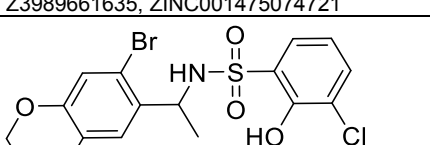
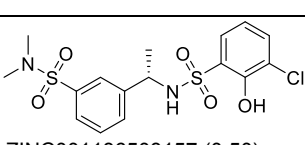
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3989663601, ZINC001474992853	74.13	182	50	0.56	 ZINC000549719284 (0.43)
 Z3989663634, ZINC001666572192	12.87	226	41	0.74	 ZINC001186508157 (0.69)

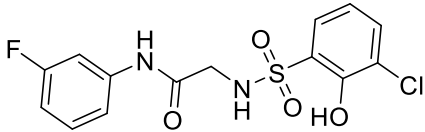
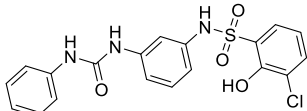
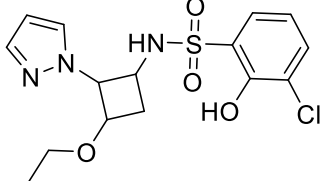
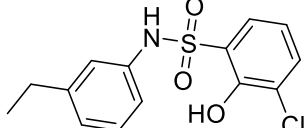
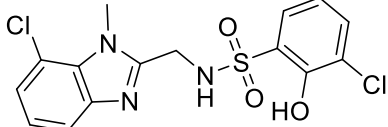
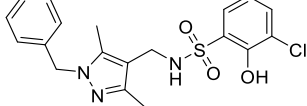
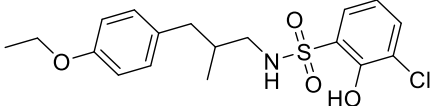
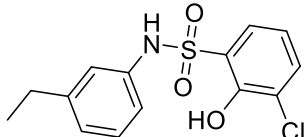
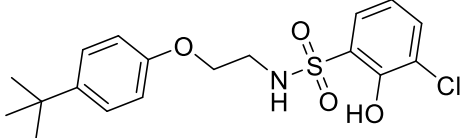
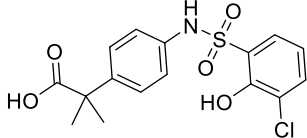


Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3989663625, ZINC001662044468	23.23 (50uM)	424	118	0.56	 ZINC000436479530 (0.47)
 Z3989661636, ZINC001561256162	33.71	448	127	0.56	 ZINC001208058246 (0.48)
 Z3989661646, ZINC001666656536	29.50 (100uM)	739	212	0.54	 ZINC000559252749 (0.44)
 Z3989661608, ZINC001593345874	25.71	46134	12194	0.58	 3-(3,4-Dichloro-phenylsulfamoyl)-thiophene-2-carboxylic acid (0.47)
 Z355256356, ZINC001608246713	44.05	67792	16238	0.62	 3-Benzylsulfamoyl-thiophene-2-carboxylic acid (0.72)
 Z2774444392, ZINC000632381004	11.56	68587	20483	0.52	 ZINC000559249118 (0.45)
 Z445512790, ZINC000319717798	16.13	78880	21981	0.55	 3-(3,4-Dichloro-phenylsulfamoyl)-thiophene-2-carboxylic acid (0.46)

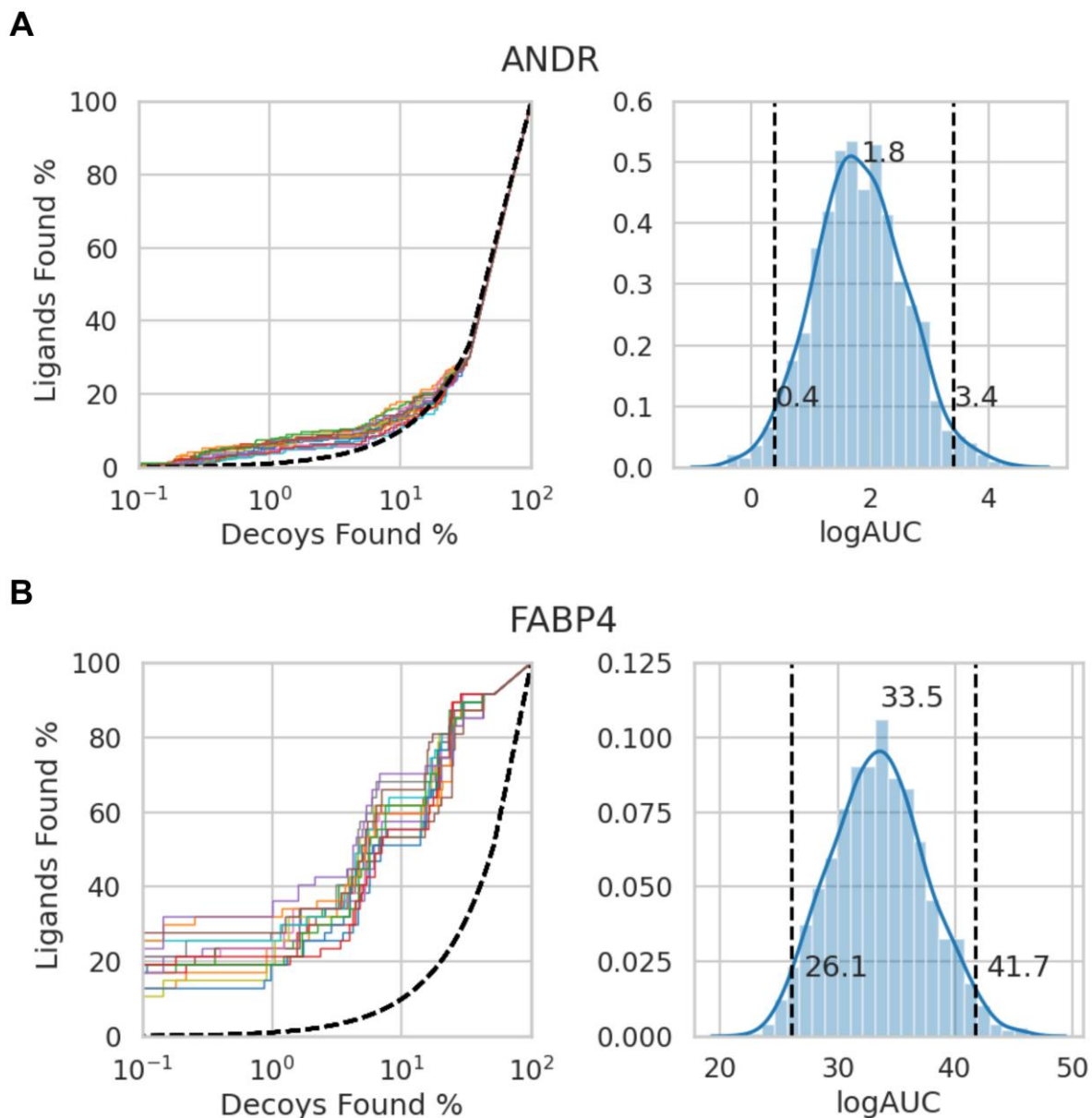
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3989663574, ZINC001364842415	6.97	111315	32955	0.53	 ZINC000436479530 (0.48)
 Z3989661610, ZINC000384192320/ ZINC000384192319	9.42	113040 / 94882	28140 / 15185	0.6 / 0.8	 3-Benzylsulfamoyl-thiophene-2- carboxylic_acid (0.51)
 Z3989663523, ZINC001607844684	32.08	113043	16393	0.84	 3-Benzylsulfamoyl-thiophene-2- carboxylic_acid (0.47)
 Z3989663551, ZINC001209231438	11.92 (50uM)	118216	33214	0.55	 ZINC000581714578 (0.44)
 Z2940234968, ZINC001251429491	14.77	135153	42125	0.51	 ZINC000632456968 (0.37)
 Z3989661629, ZINC001423901557	27.76	177139	44621	0.6	 ZINC000549722993 (0.47)
 Z2940498999, ZINC001251428895	13.94	246886	68196	0.56	 ZINC001186508157 (0.42)

Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z2940312243, ZINC001251387861	12.57	263612	71846	0.56	 ZINC001186508157 (0.49)
 Z3989661628, ZINC001414219141	7.59	267329	61449	0.64	 ZINC001186508157 (0.53)
 Z3989663580, ZINC001434557893	1.29	272507	85961	0.5	 ZINC000436479530 (0.44)
<b>ANTI-bGIST</b>					
 Z2940307649, ZINC001251419289	77.01	80	427	0.73	 ZINC000581714578 (0.52)
 Z3989661624, ZINC001309062817	67.44	110	481	0.64	 ZINC000563543464 (0.60)
 Z2275041991, ZINC000450990100	87.59	165	1600	0.99	 ZINC000581714578 (0.71)
 Z3989661637, ZINC001561899653	80.03	170	691	0.61	 ZINC001208058246 (0.48)
 Z2234688146, ZINC000436478328	75.02	240	963	0.6	 ZINC000581714578 (0.55)

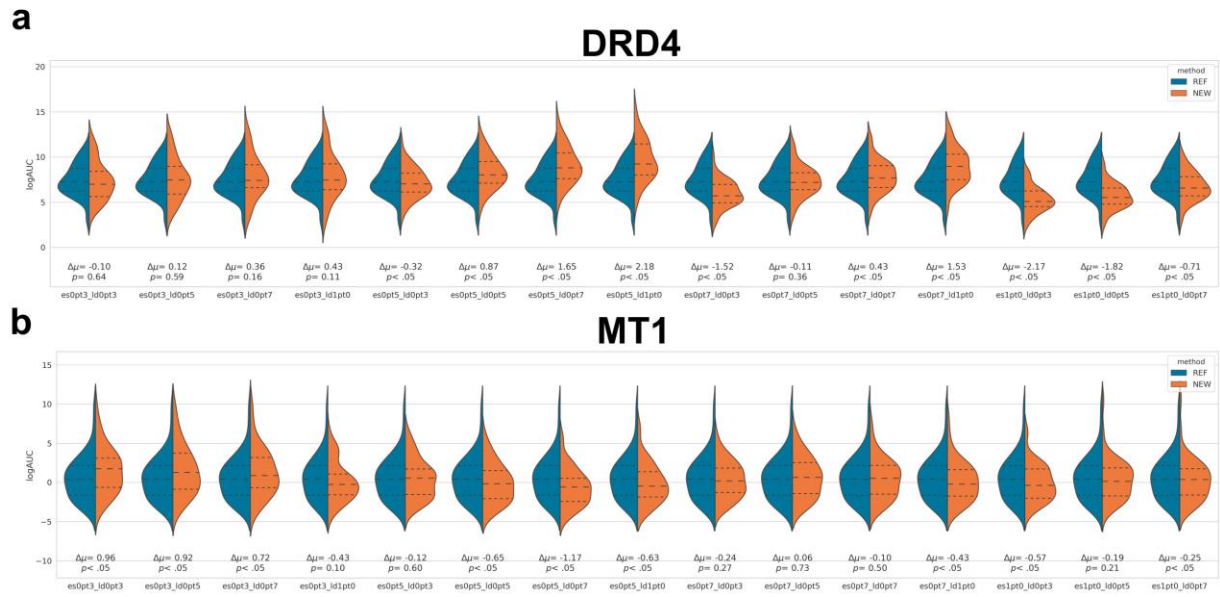
Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3989661639, ZINC001653645310	20.58 (50uM)	268	3367	1.1	 ZINC000550111357 (0.44)
 Z3989661641, ZINC001655188929	77.11 (100uM)	318	1500	0.67	 ZINC001208288320 (0.45)
 Z3989661632, ZINC001463030415	18.8 (100uM)	326	1037	0.5	 ZINC001208058246 (0.48)
 Z3989661621, ZINC001195214804	59.22	354	1296	0.56	 ZINC000559249118 (0.58)
 Z3989661643, ZINC001664179432	23.16	360	1429	0.6	 ZINC001190919234 (0.58)
 Z3989661623, ZINC001309057083	35.14	379	2275	0.78	 ZINC001208058246 (0.60)
 Z3989661635, ZINC001475074721	27.1	400	1651	0.62	 ZINC000339204163 (0.52)
 Z3989663559, ZINC001309078396	3.89 (100uM)	431	2104	0.69	 ZINC001186508157 (0.50)

Enamine ID, ZINC ID	Inhibition @ 300uM (%)	STD Rank	GIST Rank	Log Diff Change	Closest Known (ECFP4 Tanimoto Coefficient)
 Z3989661638, ZINC001621709274	20.5	434	1485	0.53	 ZINC000555530101 (0.54)
 Z2774428723, ZINC000632377095	-2.26	475	2125	0.65	 ZINC000581714578 (0.41)
 Z3989661626, ZINC001309413210	29.33 (100uM)	618	3201	0.71	 ZINC001208058246 (0.47)
 Z3989663630, ZINC001664302212	43.22 (100uM)	1859	7590	0.61	 ZINC000581714578 (0.46)
 Z3989661631, ZINC001462737746	8.52 (50uM)	2329	11521	0.69	 ZINC000563498328 (0.46)

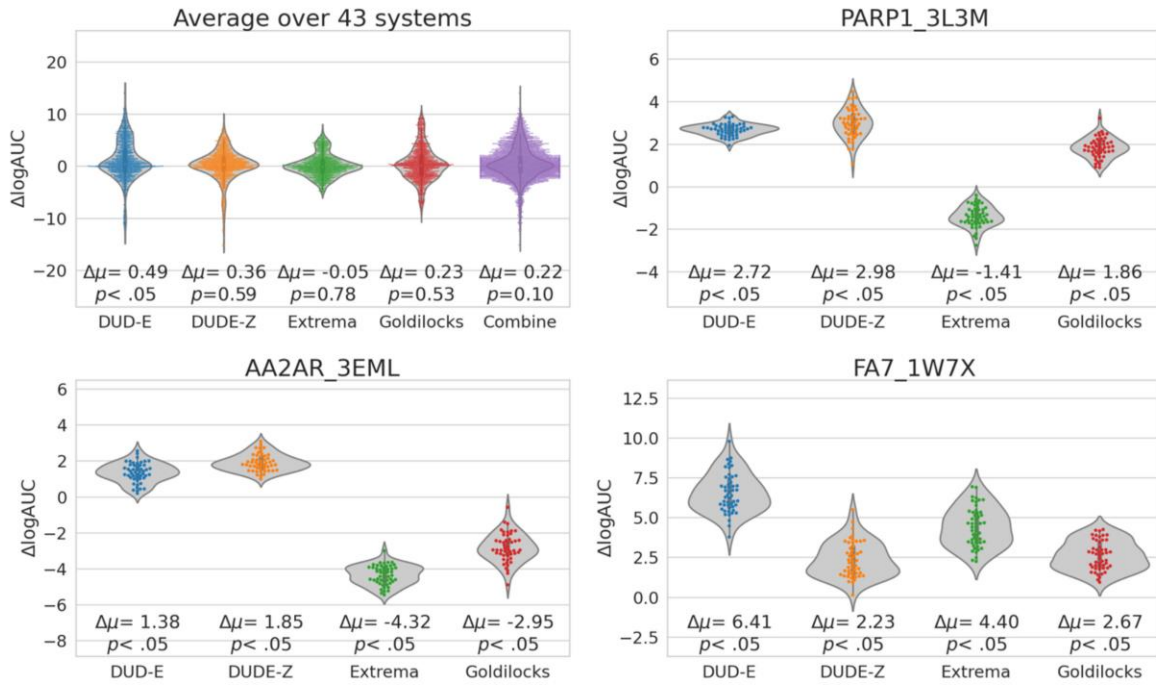
### A3. Supplementary Material for Chapter 3



**Figure A.3.1. Examples of bootstrapping enrichment distribution.** ROC curves with 15 bootstrap replicas are shown on the left. Tight distribution for Androgen Receptor (ANDR, **a**) where 95% confidence interval is 3 adjusted log AUC units. Wider distribution for Fatty acid binding protein adipocyte (FABP4, **b**) with 95% confidence interval of 15.6 adjusted log AUC units.

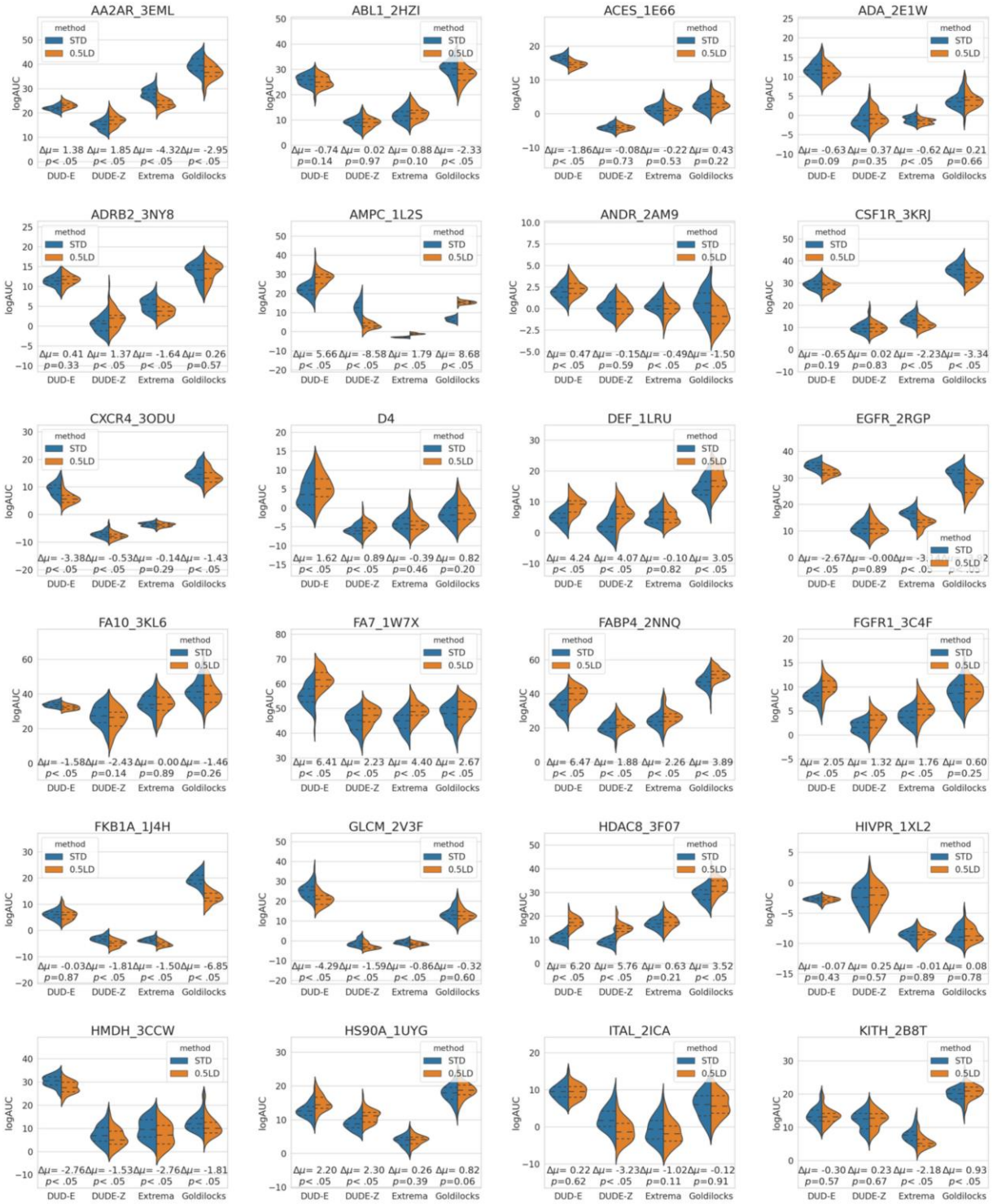


**Figure A.3.2. Bootstrapping on Binders/Nonbinders.** Bootstrapping enrichment distributions of all scoring function coefficient combinations for binders and nonbinders for **a**) D4 dopamine (81, 486), and **b**) MT1 melatonin (105, 65) receptors. The left panels (REF, blue) are different bootstrapping enrichment distributions of the standard scoring function whereas the right panels (NEW, orange) represent the bootstrapped enrichment distribution of the scoring function coefficient combination labeled. Mean log AUC differences and p-values are reported below.



**Figure A.3.3. Bootstrapping Enrichment Differences.** Examples of bootstrapping enrichment distribution where the difference for each the pairs of log AUC is calculated and then the distribution is plotted, and the z-test performed comparing to the distribution about zero.





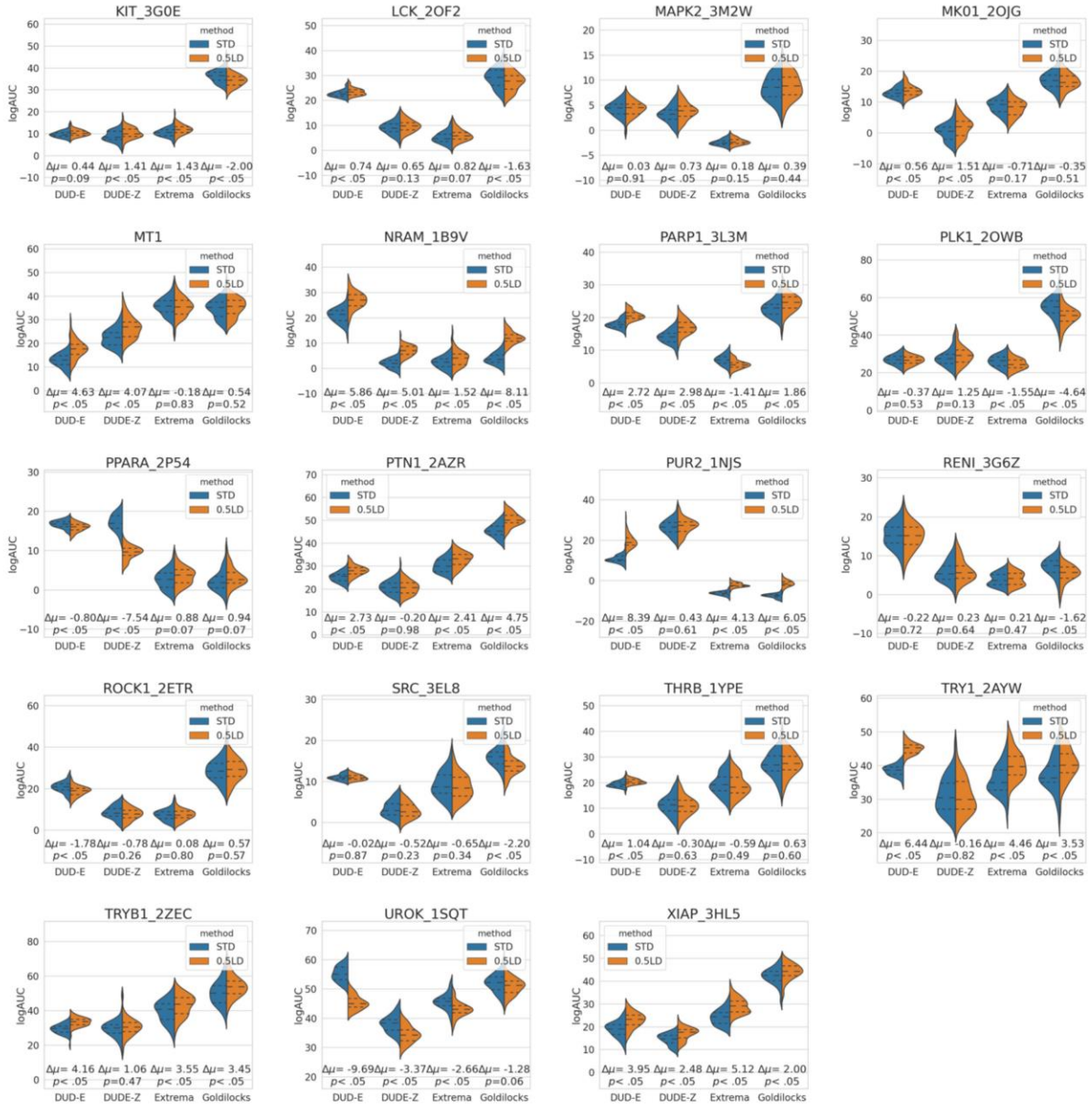
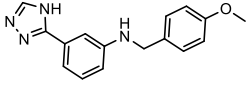
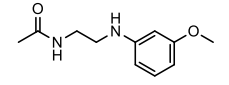
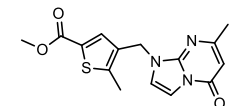
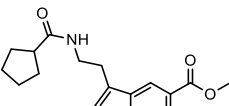
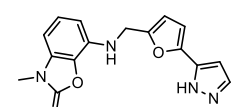
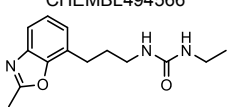
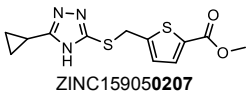
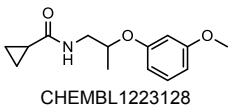
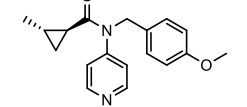
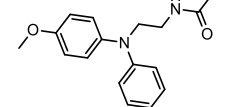
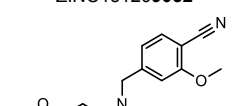
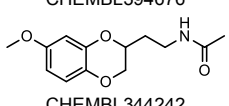
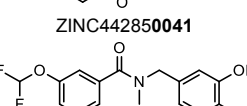
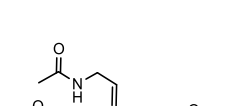
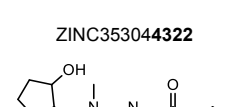
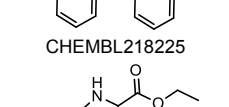
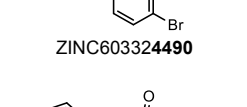
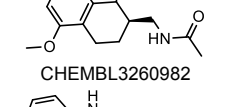


Figure A.3.4. Bootstrapping statistics for all 43 systems.

## A4. Supplementary Material for Chapter 4

**Table A.4.1. Active molecules from the initial docking screen.**

Compound	Cluster rank <sup>a</sup> (global rank)	hMT <sub>1</sub> <sup>b</sup> pEC50 (% Emax) <i>n</i>	hMT <sub>2</sub> <sup>c</sup> pEC50 (% Emax) <i>n</i>	Tc <sup>d</sup>	Nearest ChEMBL23 <sup>e</sup> MT <sub>1</sub> /MT <sub>2</sub> Ligand
 ZINC157665999	167 (197)	4.89±0.38 (63±6) <i>n</i> =3	Inverse 7.29±0.16 (Inverse 90±16) <i>n</i> =3	0.33	 CHEMBL398017
 ZINC419113878	396 (522)	5.20±0.08 (84±4) <i>n</i> =4	< 4.5 <i>n</i> =4	0.22	 CHEMBL494566
 ZINC433313647	875 (1242)	6.81±0.32 (42±2) <i>n</i> =3	7.77±0.02 (96±5) <i>n</i> =3	0.19	 CHEMBL125226
 ZINC159050207	1559 (2474)	9.00±0.15 (99±1) <i>n</i> =4	8.70±0.25 (83±3) <i>n</i> =4	0.24	 CHEMBL1223128
 ZINC151209032	1981 (3583)	5.70±0.11 (88±4) <i>n</i> =4	< 4.5 <i>n</i> =4	0.31	 CHEMBL394676
 ZINC442850041	4123 (7872)	7.91±0.04 (99±3) <i>n</i> =3	9.33±0.33 (97 ± 2) <i>n</i> =3	0.29	 CHEMBL344242
 ZINC353044322	5764 (28,258)	5.48±0.05 (87±6) <i>n</i> =4	< 4.5 <i>n</i> =4	0.33	 CHEMBL218225
 ZINC603324490	7612 (53,767)	Inverse 5.92±0.29 <i>n</i> =3	Inverse 6.20±0.08 Inverse (202±30) <i>n</i> =4	0.27	 CHEMBL3260982
 ZINC182731037	7840 (17,095)	5.30±0.09 (82±2) <i>n</i> =4	< 4.5 <i>n</i> =4	0.29	 CHEMBL3612457
ZINC92585174	1836 (3010)	7.80±0.17 (98±1) <i>n</i> =4	7.68±0.14 (74±8) <i>n</i> =4	0.23	CHEMBL1760949
ZINC432154404	1849 (3035)	6.63±0.17 (95±2) <i>n</i> =4	7.00±0.17 (74±4) <i>n</i> =4	0.27	CHEMBL1760956
ZINC664088238	2248 (3816)	< 5 <i>n</i> =4	5.85±0.06 (75±8) <i>n</i> =4	0.20	CHEMBL435032
ZINC576887661	4161 (14,292)	7.10±0.19 (83±0) <i>n</i> =4	7.28±0.36 (68±5) <i>n</i> =4	0.27	CHEMBL491605

Compound	Cluster rank <sup>a</sup> (global rank)	hMT <sub>1</sub> <sup>b</sup> pEC <sub>50</sub> (% Emax) <i>n</i>	hMT <sub>2</sub> <sup>c</sup> pEC <sub>50</sub> (% Emax) <i>n</i>	Tc <sup>d</sup>	Nearest ChEMBL23 <sup>e</sup> MT <sub>1</sub> /MT <sub>2</sub> Ligand
ZINC301472854	5033 (10,022)	6.03±0.10 (95±5) <i>n</i> =4	7.00±0.21 (88±6) <i>n</i> =4	0.26	CHEMBL115444
ZINC580731466	8503 (19,003)	5.70±0.13 (71±3) <i>n</i> =4	7.55±0.10 (98±5) <i>n</i> =4	0.26	CHEMBL115444

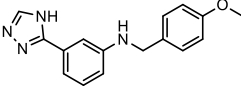
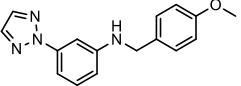
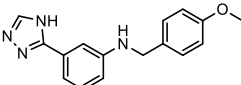
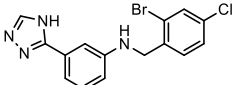
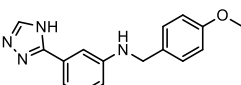
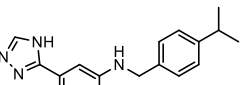
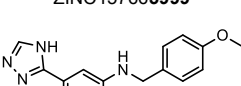
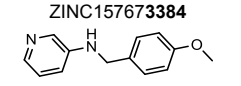
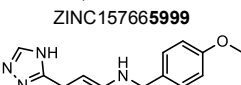
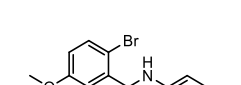
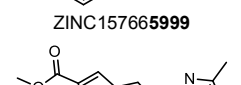
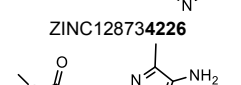
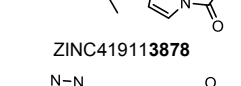
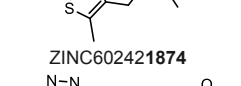
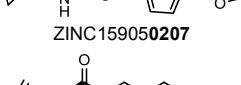
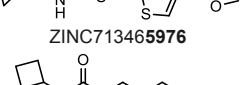
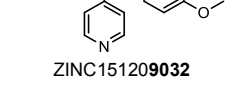
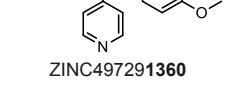
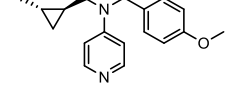
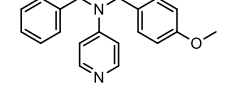
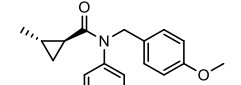
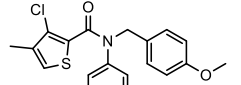
**a.** Cluster rank, Global rank (Methods)

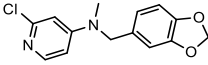
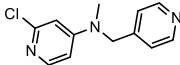
**b.** The log half maximal concentration (pEC<sub>50</sub>) for inhibition of isoproterenol-stimulated cAMP production on hMT<sub>1</sub> or hMT<sub>2</sub> melatonin receptors transiently expressed in HEK cells. Values in parenthesis represent the percentage of the maximal inhibition normalized to % melatonin response, except for inverse agonists, indicated by (Inverse), where data is normalized to % basal induced by isoproterenol. Data represent mean ± S.E.M. from the indicated number (*n*) of biologically independent experiments run in triplicate.

**d.** ECFP4 Tanimoto coefficient (Tc) to the most similar known MT<sub>1</sub> or MT<sub>2</sub> ligand in ChEMBL23.

**e.** MT<sub>1</sub>/MT<sub>2</sub> ligand in ChEMBL23 most similar to docking active.

**Table A.4.2. Some of the potent analogs from initial hits**

Initial Hit <sup>a</sup>	Analog <sup>b</sup>	hMT <sub>1</sub> <sup>c</sup> pEC50 (% Emax) <i>n</i>	hMT <sub>2</sub> <sup>d</sup> pEC50 (% Emax) <i>n</i>
 ZINC157665999	 ZINC864032792	7.49 ± 0.04 (57 ± 3) <i>n</i> =3	Inverse 6.66 ± 0.08 (Inverse 35 ± 5) <i>n</i> =3
 ZINC157665999	 ZINC555417447	Inverse 7.39 ± 0.10 (Inverse 62 ± 13) <i>n</i> =8	Inverse 5.66 ± 0.10 (Inverse 84 ± 9) <i>n</i> =8
 ZINC157665999	 ZINC157673384	Inverse 7.68 ± 0.09 (Inverse 47 ± 12) <i>n</i> =13	Inverse 6.18 ± 0.04 (Inverse 153 ± 14) <i>n</i> =12
 ZINC157665999	 ZINC5586789	6.81 ± 0.72 (37 ± 8) <i>n</i> =3	8.07 ± 0.15 (51 ± 3) <i>n</i> =4
 ZINC157665999	 ZINC128734226	6.83 ± 0.17 (79 ± 3) <i>n</i> =4	8.15 ± 0.09 (89 ± 3) <i>n</i> =4
 ZINC419113878	 ZINC602421874	4.70 ± 0.11 (51 ± 3) <i>n</i> =4	5.35 ± 0.10 (66 ± 7) <i>n</i> =4
 ZINC159050207	 ZINC713465976	7.75 ± 0.22 (101 ± 0) <i>n</i> =4	8.23 ± 0.11 (94 ± 3) <i>n</i> =4
 ZINC151209032	 ZINC497291360	7.05 ± 0.10 (92 ± 2) <i>n</i> =4	7.48 ± 0.05 (75 ± 5) <i>n</i> =4
 ZINC151209032	 ZINC151192780	5.18 ± 0.22 (54 ± 4) <i>n</i> =4	7.13 ± 0.12 (95 ± 5) <i>n</i> =4
 ZINC151209032	 ZINC485552623	< 5 <i>n</i> =4	5.80 ± 0.06 (107 ± 5) <i>n</i> =4
 ZINC442850041	 ZINC608506688	9.78 ± 0.13 (99 ± 1) <i>n</i> =4	8.60 ± 0.10 (89 ± 3) <i>n</i> =4

Initial Hit <sup>a</sup>	Analog <sup>b</sup>	hMT <sub>1</sub> <sup>c</sup> pEC <sub>50</sub> (% Emax) <i>n</i>	hMT <sub>2</sub> <sup>d</sup> pEC <sub>50</sub> (% Emax) <i>n</i>
 ZINC301472854	 ZINC223593565	6.40 ± 0.18 (86 ± 4) <i>n</i> =4	6.45 ± 0.20 (58 ± 5) <i>n</i> =4

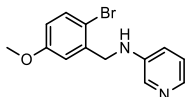
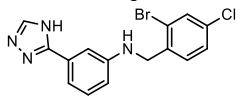
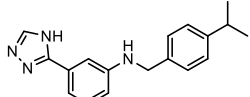
**a.** Compound selected directly from the primary docking screen and found to be active on in vitro testing

**b.** Analog from initial hit

**c, d.** The log half maximal concentration (pEC<sub>50</sub>) for inhibition of isoproterenol-stimulated cAMP production on hMT<sub>1</sub> or hMT<sub>2</sub> melatonin receptors transiently expressed in HEK cells. Values in parenthesis represent the percentage of the maximal inhibition normalized to % melatonin response, except for inverse agonists, indicated by (Inverse), where data is normalized to % basal induced by isoproterenol. Data represent mean ± S.E.M. from the indicated number (*n*) of biologically independent experiments run in triplicate.

**UCSF7447 ('7447), UCSF3384 ('3384), UCSF4226 ('4226)**

**Table A.4.3. Pharmacokinetics of three melatonin receptor type-selective ligands**

Compound	pIC <sub>50</sub> (Emax %) pEC <sub>50</sub> (IA)	C <sub>max</sub> <sup>a</sup> (ng/mL)	AUC <sup>b</sup> (hr*ng/mL)	T <sub>1/2</sub> <sup>c</sup> (hr)	CL <sup>d</sup> (mL/min/kg)	V <sub>ss</sub> <sup>e</sup>	Brain/Plasma ratio
 ZINC128734226 MT <sub>2</sub> -selective agonist	<b>pIC<sub>50</sub></b> MT <sub>1</sub> – 6.8 (48%) MT <sub>2</sub> – 8.2 (80%)	1922.8	282.1	0.29	117.9	1.11	1.58 (30')
 ZINC555417447 MT <sub>1</sub> -selective inverse agonist	<b>pEC<sub>50</sub></b> MT <sub>1</sub> – 7.4 (IA) MT <sub>2</sub> – 5.8 (IA)	1948.6	494.5	0.27	67.11	1.11	3.03 (30')
 ZINC157673384 MT <sub>1</sub> -selective inverse agonist	<b>pEC<sub>50</sub></b> MT <sub>1</sub> – 7.7 (IA) MT <sub>2</sub> – 6.2 (IA)	1299.6	563.8	0.32	58.48	1.38	1.43 (30')

**a.** C<sub>max</sub>: Maximum concentration

**b.** AUC: Area under plasma concentration-time curve

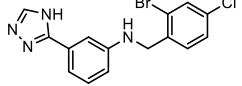
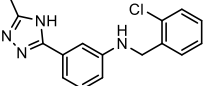
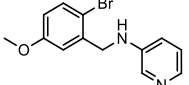
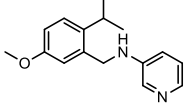
**c.** Half-life

**d.** Clearance

**e.** Volume of distribution at steady-state

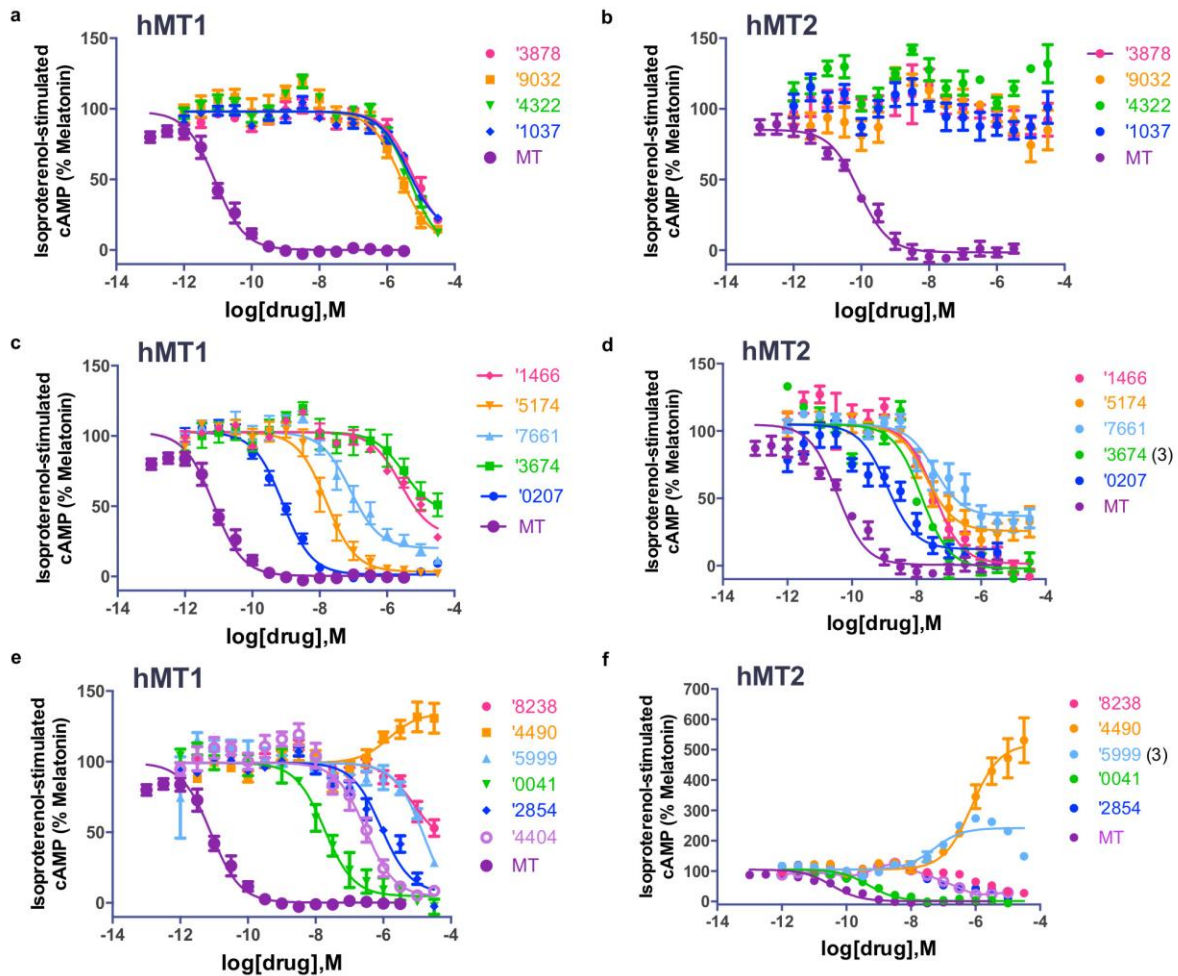
**UCSF4226 ('4226), UCSF7447 ('7447), UCSF3384 ('3384)**

**Table A.4.4: Probe pairs of *in vivo* tested molecules**

Active Selective Probe (Sigma RefCode)	hMT <sub>1</sub> pEC <sub>50</sub> <sup>a</sup> (% Emax) <i>n</i>	hMT <sub>2</sub> pEC <sub>50</sub> <sup>b</sup> (% Emax) <i>n</i>	Inactive analog (Sigma RefCode)	hMT <sub>1</sub> pEC <sub>50</sub> <sup>a</sup> <i>n</i>	hMT <sub>2</sub> pEC <sub>50</sub> <sup>b</sup> <i>n</i>
 ZINC555417447 (SML2751)	Inverse 7.4 ± 0.10 (Inverse 62 ± 13) <i>n</i> =8	Inverse 5.7 ± 0.10 (Inverse 84 ± 9) <i>n</i> =8	 ZINC37781618 (SML2752)	< 4.5 <i>n</i> =3	< 4.5 <i>n</i> =3
 ZINC128734226 (SML2753)	6.8 ± 0.2 (79 ± 3) <i>n</i> =4	8.2 ± 0.1 (89 ± 3) <i>n</i> =4	 Z3670677764 (SML2754)	< 4.5 <i>n</i> =3	< 4.5 <i>n</i> =3

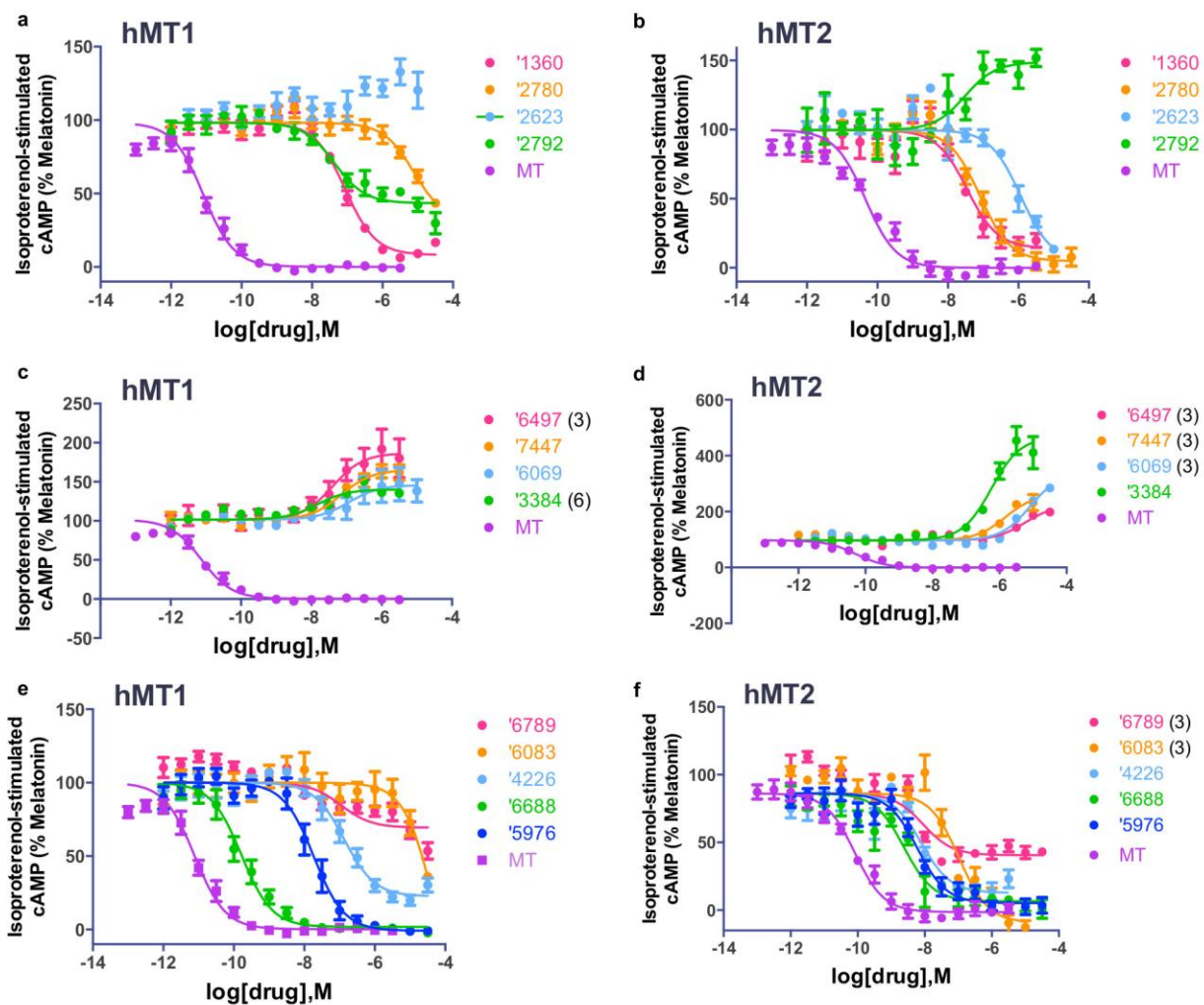
**a, b.** The log half maximal concentration (pEC<sub>50</sub>) for inhibition of isoproterenol-stimulated cAMP production on hMT<sub>1</sub> or hMT<sub>2</sub> melatonin receptors transiently expressed in HEK cells. Values in parenthesis represent the percentage of the maximal inhibition normalized to % melatonin response for **4226**, and to % basal activity for **7447**. Compounds were tested at concentrations up to 30μM. Data represent mean ± S.E.M. from the indicated number (N) of biologically independent experiments run in triplicate. **UCSF4226** (**4226**), **UCSF7447** (**7447**)



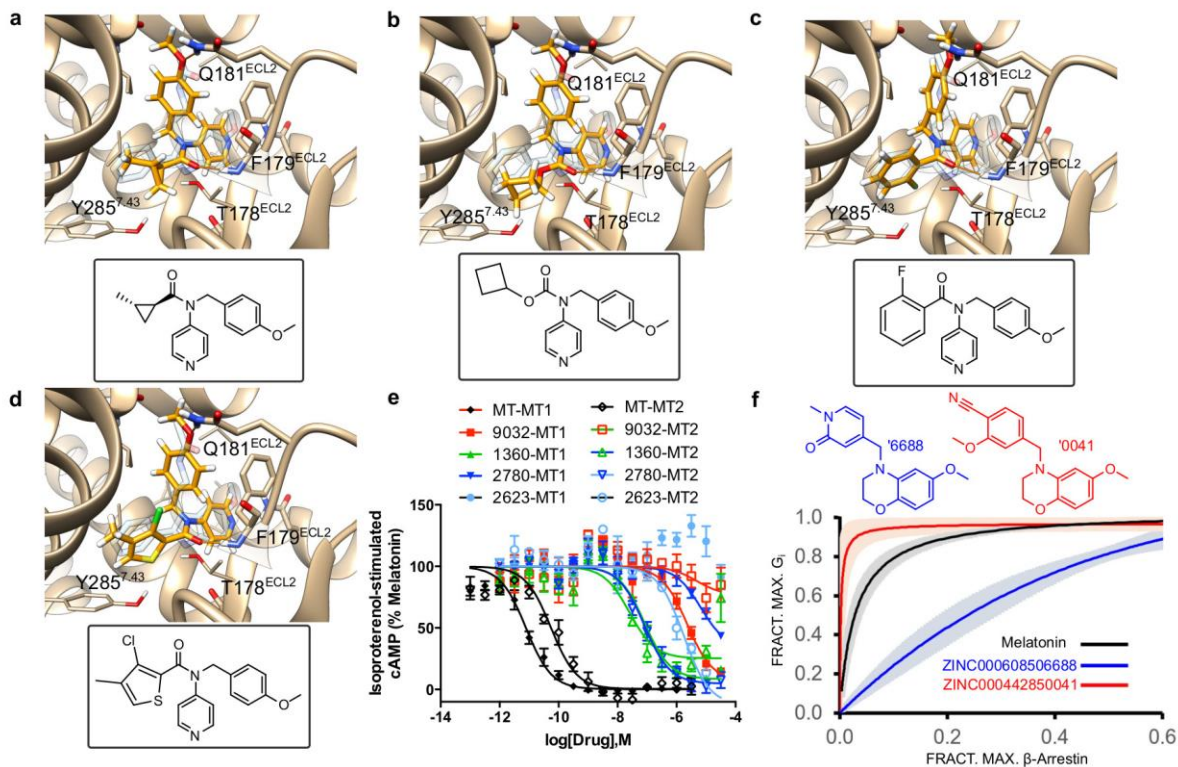


**Figure A.4.1. Concentration-response curves of initial 15 compounds.** hMT<sub>1</sub>- (a,c,e) or hMT<sub>2</sub>-mediated (b,d,f) inhibition of isoproterenol-stimulated cAMP in HEK cells by melatonin and 15 initial compounds. Data normalized to melatonin response represent mean  $\pm$  s.e.m. of four biologically independent experiments (n=4) run in triplicate, unless otherwise indicated, which is indicated in parenthesis next to each compound name.

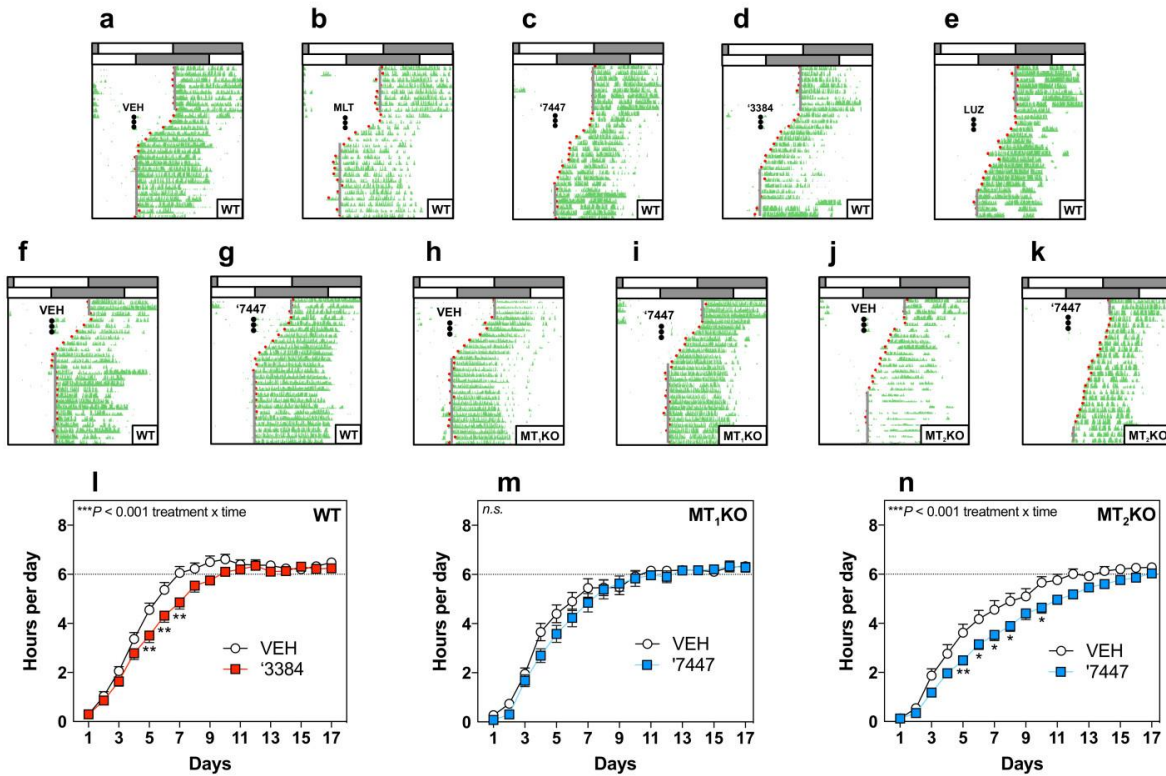




**Figure A.4.2. Concentration-response curves of interesting analogs.** hMT<sub>1</sub>- (a,c,e) or hMT<sub>2</sub>-mediated (b,d,f) inhibition of isoproterenol-stimulated cAMP in HEK cells by melatonin and select analogs. Data normalized to melatonin response represent mean ± s.e.m. of four biologically independent experiments (n=4) run in triplicate, unless otherwise indicated, which is indicated in parenthesis next to each compound name.



**Figure A.4.3. Small ligand changes have large effects on activity and selectivity.** **a**, Docked pose of '9032, an MT<sub>1</sub>-selective direct docking hit. **b**, Docked pose of '1360, a close analog of '9032 that switches 2-fold selectivity for MT<sub>2</sub> over MT<sub>1</sub>. **c**, Docked pose of '2780, an analog where MT<sub>2</sub> selectivity climbs to 89-fold over MT<sub>1</sub>. **d**, Docked pose of '2623, which adds a bulkier 2-chloro-3-methylthiophene into a proposed MT<sub>2</sub>-selective hydrophobic cleft, resulting in a fully MT<sub>2</sub>-selective agonist without detectable MT<sub>1</sub> activity. All docked poses are overlaid onto the crystallographic pose of 2-phenylmelatonin in transparent blue. **e**, Concentration-response curves the four analogs at MT<sub>1</sub> and MT<sub>2</sub>. Data normalized to melatonin response represent mean ± s.e.m. of four biologically independent experiments (n=4) run in triplicate. **f**, Bias plots of '0041 and '6688 relative to melatonin signaling. Mean values (Table A.3.6) are presented as solid lines and the 95% confidence interval for the line is shaded. Data are normalized to melatonin response and represent mean ± s.e.m. of three biologically independent experiments (n=3) run in triplicate, except for '6688 for G<sub>i</sub> activation (n=4).

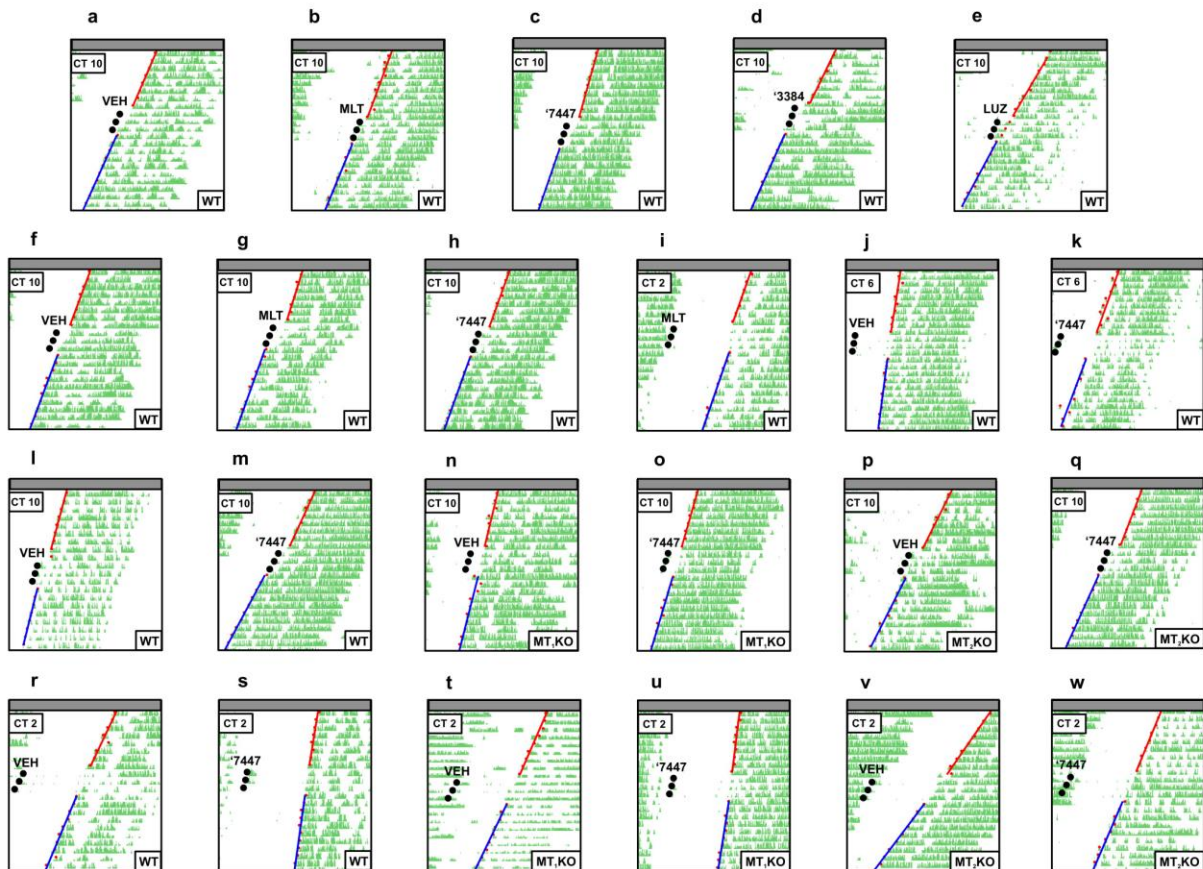


**Figure A.4.4. MT<sub>1</sub>-selective inverse agonists decelerate re-entrainment rate *in vivo*.**

**a - e, Representative actograms** of running wheel (RW) activity in wild type (WT) C3H/HeN (C3H) mice treated with VEH (**a**), 30  $\mu\text{g}/\text{mouse}$  MLT (**b**), **UCSF7447** (**c**), **UCSF3384** (**d**), as well as 300  $\mu\text{g}/\text{mouse}$  LUZ (**e**) just prior to the new dark onset (black dots) following an abrupt 6h advance of dark onset in a 12:12 light-dark cycle (gray: dark phase; white: light phase). Compounds were administered once a day for 3 days (see Methods for additional details). Corresponding quantification found in **Fig. 3b,c. f - k**, Representative actograms of RW activity for VEH [WT (**a**), MT<sub>1</sub>KO (**c**), MT<sub>2</sub>KO (**e**)] or inverse agonist '**7447** [WT (**b**), MT<sub>1</sub>KO (**d**), MT<sub>2</sub>KO (**f**)] treated C3H mice following a 6 h advance of dark onset. Mice were kept in a 12:12 light-dark cycle. '**7447** (30  $\mu\text{g}/\text{mouse}$ ) was administered for 3 consecutive days just prior to the new dark onset (black dots). **l**, Inverse agonist '**3384** decelerates the rate of re-entrainment of RW activity rhythm onset in C3H WT mice. Data expressed in hours advanced each day for VEH vs. '**3384** (two-way repeated measures ANOVA; treatment x time interaction:  $F_{16,647} = 1.99$   $P = 0.0122$ ). **m**, Inverse agonist '**7447** does not modulate the rate of re-entrainment of RW activity rhythm onset in C3H MT<sub>1</sub>KO mice. Data expressed in hours advanced each day for MT<sub>1</sub>KO mice treated with VEH vs. '**7447** (mixed-effect two-way repeated measures ANOVA; treatment x time interaction:  $F_{16,474} = 1.44$   $P = 0.117$ ). **n**, Inverse agonist '**7447** decelerates the rate of re-entrainment of RW activity rhythm onset in C3H MT<sub>2</sub>KO mice. Data expressed in hours advanced each day for MT<sub>2</sub>KO mice treated with VEH vs. '**7447** (mixed-effect two-way repeated measures ANOVA; treatment x time interaction:



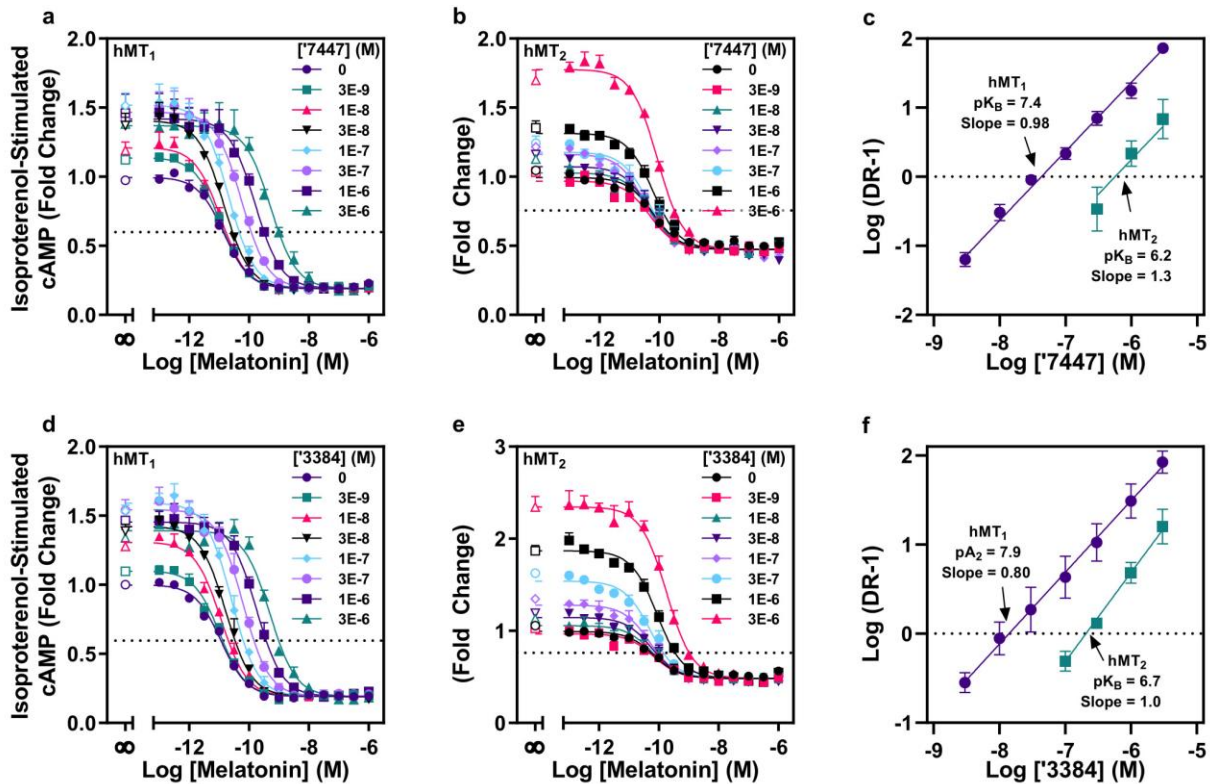
$F_{16,683} = 2.57$   $P = 0.000686$ . Data represents mean  $\pm$  s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ , for multiple comparisons by Tukey's post test ( $P < 0.05$ ). Dotted line in **j - k** refers to the new dark onset. Additional details of all statistical analyses as well as  $n$  for each condition can be found in Methods (Statistics & Reproducibility). Vehicle (VEH), melatonin (MLT), luzindole (LUZ), **UCSF7447** ('7447'), **UCSF3384** ('3384'). All treatments were given via s.c. injection.



**Figure A.4.5. MT<sub>1</sub>-selective inverse agonists phase advance circadian activity at MT<sub>1</sub>.**

**a - e**, Representative actograms of RW activity from individual C3H WT mice kept in constant dark (gray bars) treated with VEH (**a**), MLT (**b**), **UCSF7447** (**c**), **UCSF3384** (**d**) or LUZ (**e**). All treatments were 30  $\mu$ g/mouse except for LUZ which was 300  $\mu$ g/mouse as described in Methods. Mice were treated at dusk (CT 10; 2 hours prior to onset of RW activity) for three consecutive days (black dots). Red lines indicate best-fit line of pre-treatment onsets and blue lines indicate best-fit line of post treatment onsets both used for phase shift determinations (see Methods for more details). Corresponding quantification found in **Fig. 3.3d**. **f - h**, Representative actograms of RW activity from individual C3H WT mice kept in constant dark treated with VEH (**f**), MLT (**g**), or '7447

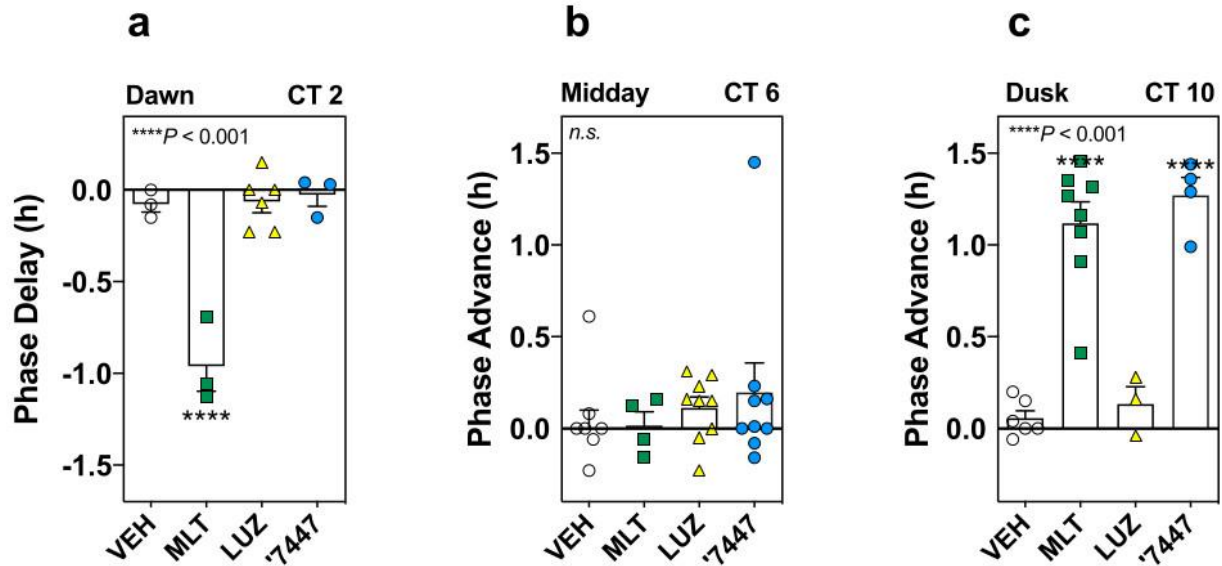
(h, all treatments 0.9  $\mu\text{g}/\text{mouse}$ ) at CT 10. Corresponding quantification found in **Fig. 3.3d. i - k**, Representative actograms of RW activity from individual C3H WT mice kept in constant dark treated with MLT (i) at CT 2 (10 hours prior to RW onset) or VEH (j) vs. '7447(k, all treatments at 30  $\mu\text{g}/\text{mouse}$ ) at CT 6 (6 hours prior to RW onset). Corresponding quantification found in **Fig. A.3.7. l - q**, Representative actograms of running wheel (RW) activity from individual C3H WT (l, m), MT<sub>1</sub>KO (n, o), and MT<sub>2</sub>KO (p, q) mice kept in constant dark treated with VEH (white; l, n, p) or UCSF7447(blue; m, o, q; 30  $\mu\text{g}/\text{mouse}$ ) at CT 10. Corresponding quantification found in **Fig. 3.3e. r - w**, Representative actograms of RW activity from individual C3H WT (r, s), MT<sub>1</sub>KO (t, u), and MT<sub>2</sub>KO (v, w) mice kept in constant dark treated with VEH (white; r, t, v) or UCSF7447(blue; s, u, w; 30  $\mu\text{g}/\text{mouse}$ ) at CT 2. Corresponding quantification found in **Fig. 3.3f**. Vehicle (VEH), melatonin (MLT), luzindole (LUZ), UCSF7447 ('7447), UCSF3384 ('3384). All treatments were given via s.c. injection.



**Figure A.4.6. Concentration-response curves of the inverse agonists.**

**a-d**, Modulation of hMT<sub>1</sub>- (**a,d**) or hMT<sub>2</sub>- (**b,e**) mediated inhibition of isoproterenol-stimulated cAMP in HEK cells by melatonin in the presence of '7447 (**a,b**) or '3384 (**d,e**) over a range of concentrations. Data normalized to effect of isoproterenol alone represent mean  $\pm$  s.e.m. of three biologically independent experiments (n=3) run in triplicate. **c,f**, Schild plots depicting competitive antagonism of melatonin by '7447 (**c**) and '3384 (**f**). Schild analysis at hMT<sub>1</sub> (purple) and hMT<sub>2</sub> (teal) reveal competitive

antagonism for '7447 (hMT<sub>1</sub> pK<sub>B</sub>: 7.4 ± 0.1, slope: 0.98 ± 0.03; hMT<sub>2</sub> pK<sub>B</sub>: 6.2 ± 0.1, slope: 1.3 ± 0.4) (c), and '3384 (hMT<sub>1</sub> pA<sub>2</sub>: 7.9 ± 0.1, slope: 0.80 ± 0.04; hMT<sub>2</sub> pK<sub>B</sub>: 6.7 ± 0.1, slope: 1.0 ± 0.1) (f). Data represent mean ± s.e.m. of three biologically independent experiments (n=3) run in triplicate. UCSF7447 ('7447), UCSF3384 ('3384)



**Figure A.4.7. Phase shift profiles of '7447, melatonin, and luzindole.**

a - c, C3H/HeN mice were kept in constant dark and treated with VEH, MLT, LUZ, or '7447 (all treatments 30 µg/mouse except for LUZ which was 300 µg/mouse, s.c.). Mice were treated at CT 2, 6, or 10 (10, 6, or 2 hours prior to onset of RW activity) for three consecutive days (see details in Methods). **a**, CT 2 phase shift data was compared via one-way ANOVA ( $F_{3,11} = 28.16$   $P = 1.85 \times 10^{-5}$ ). **b**, CT 6 phase shift data was compared via one-way ANOVA ( $F_{3,26} = 0.61$   $P = 0.61$ ). **c**, CT 10 phase shift data was compared via one-way ANOVA ( $F_{3,17} = 35.13$   $P = 1.66 \times 10^{-7}$ ). All multiple comparisons made to VEH using Dunnet's post hoc test ( $P < 0.05$ ).

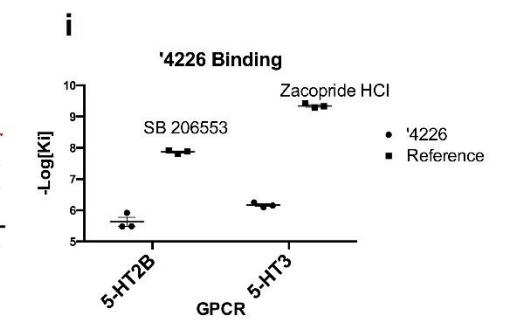
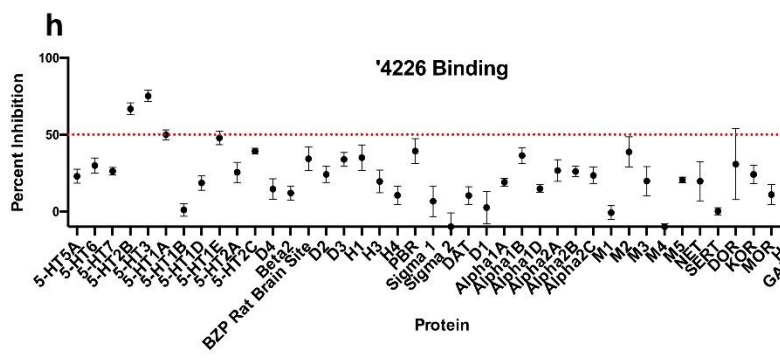
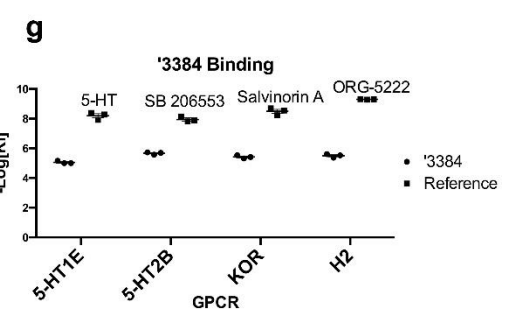
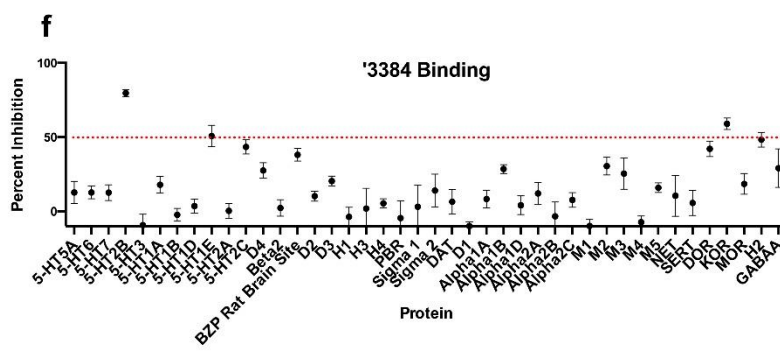
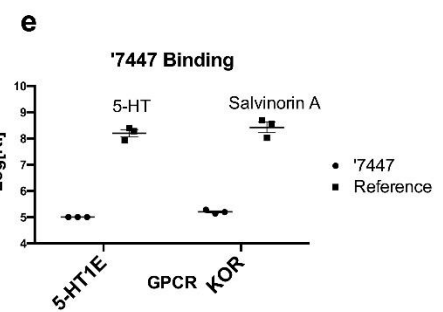
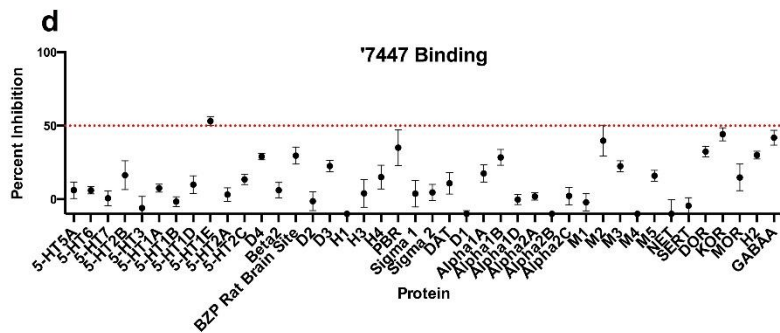
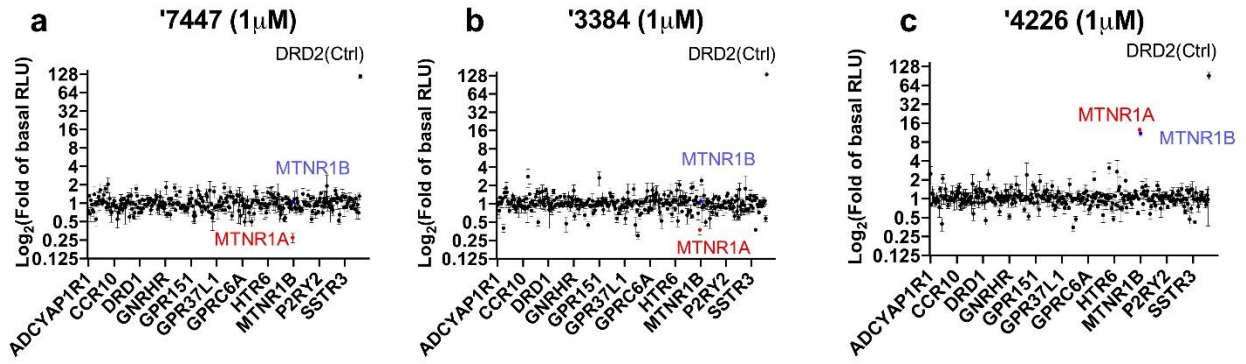
Values for MLT & '7447 at CT 10 pooled from previous data for comparison to LUZ. Data shown represent mean ± s.e.m. \*\*\*\* $P < 0.0001$  for comparisons with VEH. Vehicle (VEH), melatonin (MLT), luzindole (LUZ), UCSF7447 ('7447). All treatments were given via s.c. injection.

**Table A.4.5. Purity information of potent MT<sub>1</sub>/MT<sub>2</sub> compounds & probe pairs**

ZINC ID	Vendor ID	Purity (%)
ZINC000037781618	Z1480757072	100
Not assigned	Z3670677764	100
ZINC000301472854	Z1329100179	99
ZINC000159050207	Z1407773472	99
ZINC000353044322	Z1424428911	99
ZINC000157665999	Z1480758141	93
ZINC000157673384	Z1480758218	99
ZINC000555417447	Z1514261971	99
ZINC000092585174	Z1576575036	97
ZINC000128734226	Z1610979660	99
ZINC000182731037	Z1643514089	90
ZINC000151209032	Z1711618470	99
ZINC000433313647	Z1918348063	99
ZINC000442850041	Z1997668124	94
ZINC000580731466	Z2091863999	97
ZINC000432154404	Z2214014068	94
ZINC000419113878	Z2365412762	99
ZINC000576887661	Z2589323319	99
ZINC000664088238	Z2613947763	99
ZINC000603324490	Z2850918676	98
ZINC000864032792	Z2748808877	99
ZINC000157676497	Z1480758254	100
ZINC000516666069	Z1514262713	100
ZINC000005586789	Z1405137567	97
ZINC000037781620	Z1481194448	100
ZINC000091496083	Z1601049071	92
ZINC000602421874	Z2824473301	96
ZINC000713465976	Z2769133462	100
ZINC000497291360	Z2277188345	100
ZINC000151192780	Z1250569092	100
ZINC000485552623	Z1848085028	100
ZINC000608506688	Z1289702161	100
ZINC000223593565	Z1329102065	95
ZINC000502746614	Z512511068	100
ZINC000342894794	Z2223030428	100
ZINC000448569837	Z2252737042	100
ZINC000533223031	Z1463977047	95
ZINC000278402888	Z1873642200	96
ZINC000153979406	Z1660206648	98
ZINC000679873307	Z1643528542	97
ZINC000427909834	Z1747659963	100
ZINC000935325763	Z2958373897	95
ZINC000782844129	Z1645267832	98
ZINC000053552068	Z805386112	96
ZINC000111617751	Z1159133201	95
ZINC000795260077	Z1958698304	97
ZINC000092689376	Z1576993936	100
ZINC000771256264	Z1576993627	100
ZINC000657415258	Z2589999587	100
ZINC000433339262	Z1918345355	100
ZINC000171411960	Z1021618304	91
ZINC000295670104	Z2273909585	100
ZINC000343738565	Z231663050	100
ZINC000440646486	Z1376027177	98
ZINC000362026289	Z2191432527	95
ZINC000603335604	Z2851050870	100
ZINC000603329297	Z2850976570	100
ZINC000603283923	Z2850356925	100
ZINC000603288243	Z2850437507	100
ZINC000769913802	Z1570930696	97
ZINC000463058770	Z1694152724	96
ZINC000527535107	Z2654515526	100
ZINC000075955186	Z1289701628	100
ZINC000268578884	Z1421406889	100
ZINC000340193755	Z1289700543	100
ZINC000467388371	Z1804092730	90
ZINC000283765277	Z2034608220	100
ZINC000596286623	Z2613821040	90
ZINC000883020057	Z2852703286	97
ZINC000574060358	Z2365411510	100
ZINC000713466047	Z2769131474	100
ZINC000412984585	Z2344648963	100
ZINC000713485663	Z2769133977	100
ZINC000286577892	Z2092146945	100
ZINC000769901394	Z1570866145	99

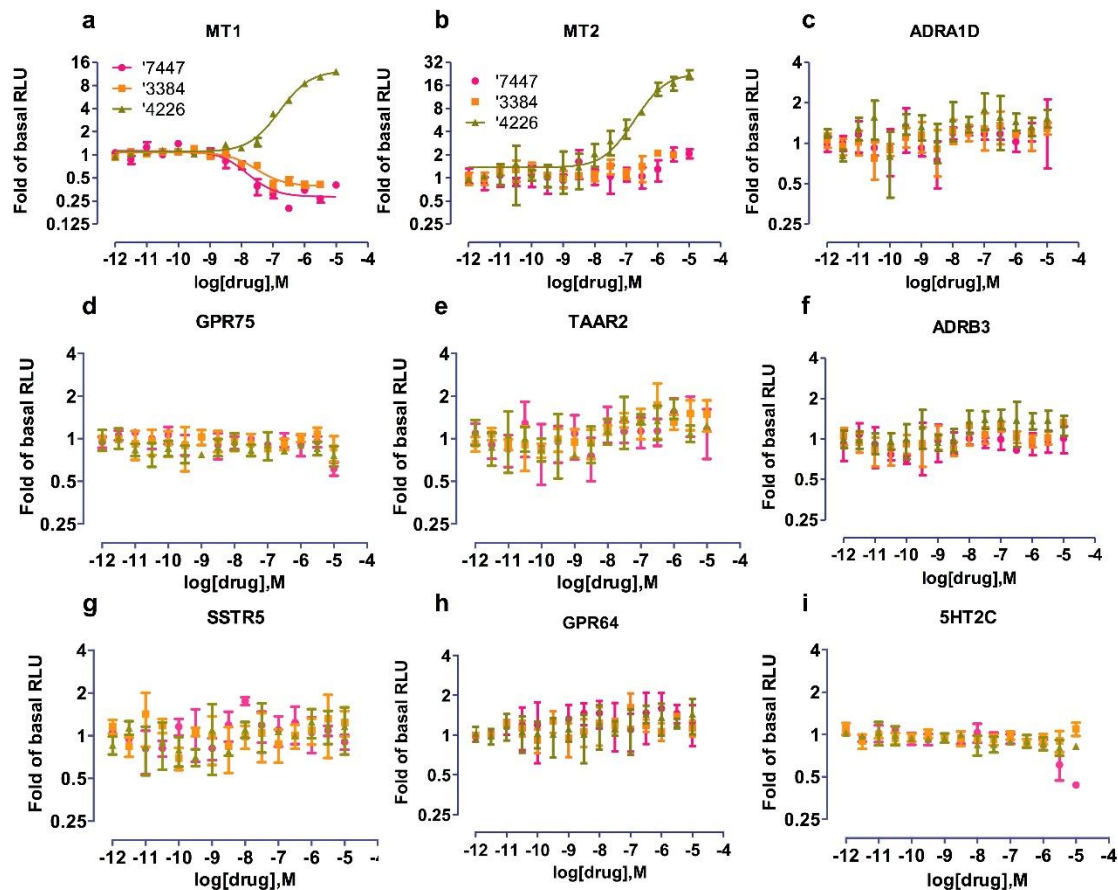
ZINC ID	Vendor ID	Purity (%)
ZINC000092827989	Z1567844459	98
ZINC000157686563	Z1573248969	97
ZINC000019884129	Z1405137410	100
ZINC000037678131	Z1405138337	100
ZINC000075955166	Z1289701578	91
ZINC000428250445	Z2206671038	100
ZINC000294945150	Z2261597750	100





**Figure A.4.8. PRESTO-Tango GPCRome & off-target screening.**

'7447 (a), '3384 (b) and '4226 (c) were screened against 320 non-olfactory GPCRs for agonism in the arrestin recruitment Tango assay. Data were normalized to the basal level of luminescence and represent mean  $\pm$  s.e.m of a single representative biological replicate using technical quadruplicates, and a second confirmatory biological replicate (again using technical quadruplicates) was also run for each compound. For the primary binding assay, each compound was tested at 10 $\mu$ M final concentration against 42 molecular targets and data (% inhibition) represent mean  $\pm$  s.e.m. of 4 biologically independent experiments (d,f,h). Targets with <50% inhibition at 10,000 nM indicate IC<sub>50</sub> values >10,000 nM. For the targets >50% inhibition, Ki was determined in full concentration responses and data (-Log(Ki)) represent mean  $\pm$  s.e.m. of 3 biologically independent experiments run in triplicate (e, g, i). (See Methods). **UCSF7447 ('7447)**, **UCSF3384 ('3384)** and **UCSF4226 ('4226)**.

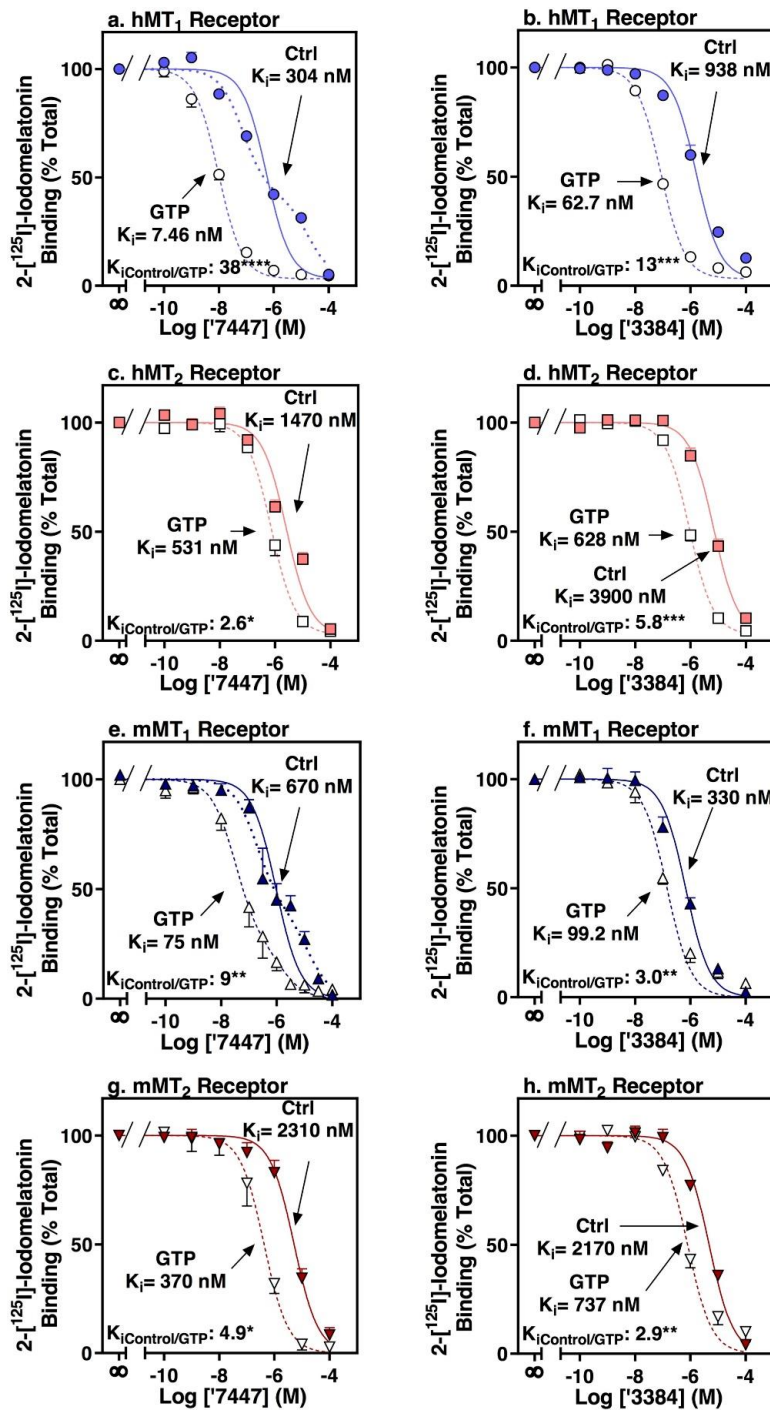


**Figure A.4.9. Dose-response curves for off-target receptors.**

'7447 (red circles), '3384 (orange squares), and '4226 (green triangles) were screened against MT<sub>1</sub> (a), MT<sub>2</sub> (b), and GPCRs that showed activity less than 0.5 fold of basal (RLU) (c-h) or greater than 3.0 fold of basal (RLU) (i) in the PRESTO-TANGO GPCRome. Targets include ADRA1D (c), GPR75 (d), TAAR2 (e), ADRB3 (f), SSTR5 (g), GPR64 (h), and 5HT2C (i). Data were normalized to the basal level of luminescence and represent the mean ± S.E.M. of three biologically independent experiments run in triplicate. **UCSF7447** ('7447), **UCSF3384** ('3384), and **UCSF4226** ('4226).

**Table A.4.6. Biased Analogs**

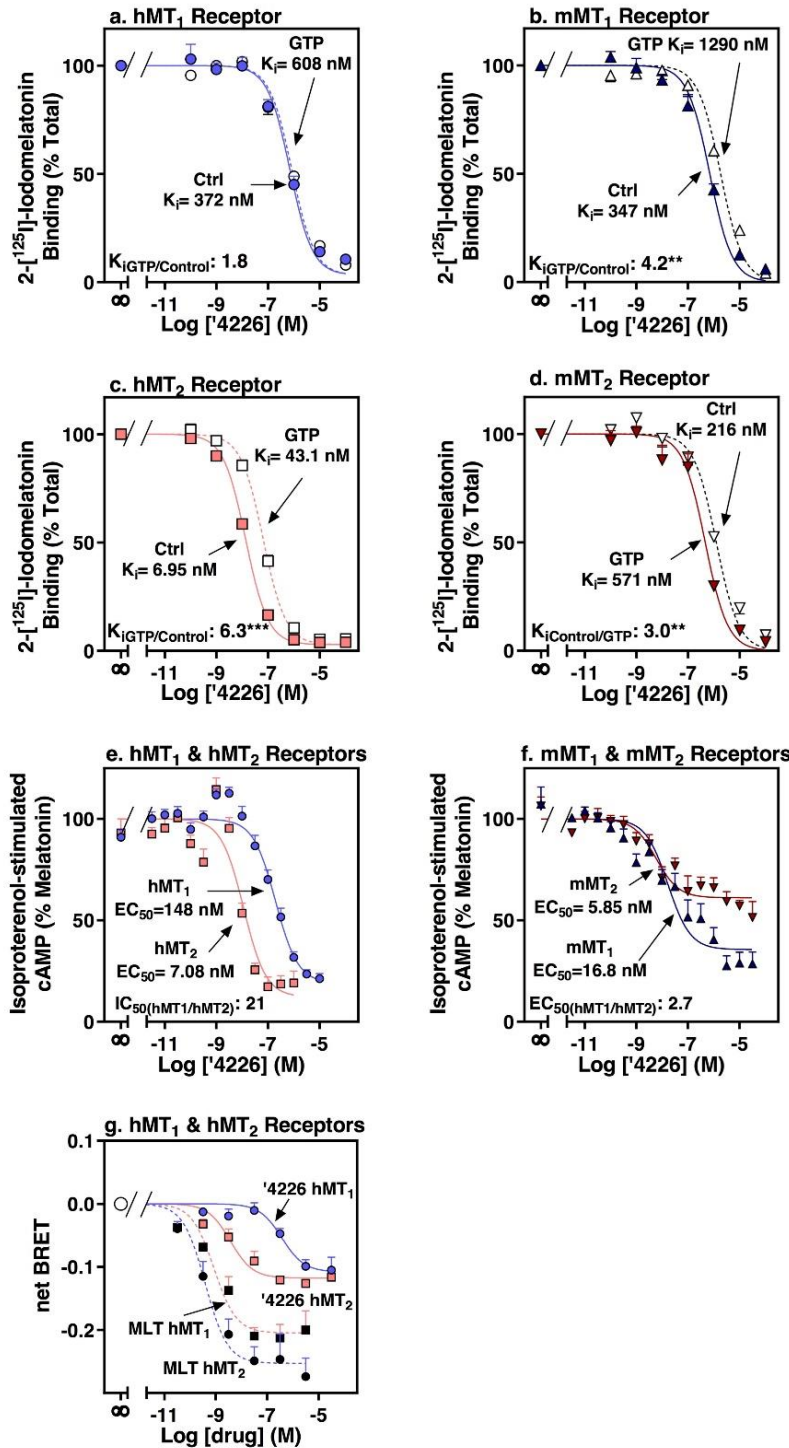
Compound	G <sub>i</sub>	β-arrestin		
	Log(Emax/EC50)	Log(Emax/EC50)	ΔΔLog(Emax/EC50)	β-arrestin Bias
Melatonin (reference)	10.10 (9.85~10.3)	8.56 (8.3~8.8)		
ZINC442850041	9.32 (9.20~9.56)	6.50 (6.2~6.7)	-1.34 (-0.89~-1.8)	0.046 (0.016~0.13)
ZINC608506688	8.60 (8.3~8.8)	7.90 (7.7~8.2)	0.92 (0.46~1.37)	8.2 (2.9~23.4)



**Figure A.4.10. Competition binding of inverse agonists against melatonin receptors.**

Competition of compounds '7447 (a,c,e,g) and '3384 (b,d,f,h) for 2-[<sup>125</sup>I]-iodomelatonin binding to hMT<sub>1</sub> (a,b), hMT<sub>2</sub> (c,d), mMT<sub>1</sub> (e,f), or mMT<sub>2</sub> (g,h) receptors stably expressed in CHO cells in the absence (closed symbols) and presence (open symbols) of 100 μM GTP, 1 mM EDTA.Na<sub>2</sub>, and 150 mM NaCl ('7447: hMT<sub>1</sub> pK<sub>i</sub>: 6.55 ± 0.08;

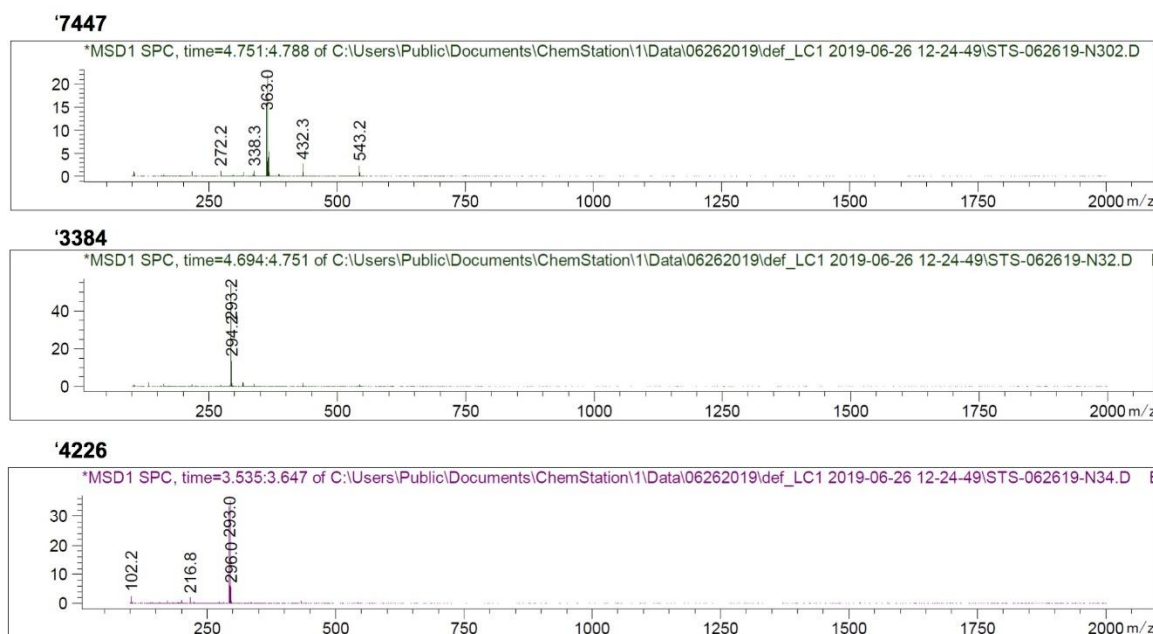
hMT<sub>1</sub>-GTP pK<sub>i</sub>: 8.15 ± 0.06; hMT<sub>2</sub> pK<sub>i</sub>: 5.85 ± 0.07; hMT<sub>2</sub>-GTP pK<sub>i</sub>: 6.30 ± 0.07; mMT<sub>1</sub> pK<sub>i</sub>: 6.54 ± 0.12; mMT<sub>1</sub>-GTP pK<sub>i</sub>: 7.64 ± 0.24; mMT<sub>2</sub> pK<sub>i</sub>: 5.66 ± 0.08; mMT<sub>2</sub>-GTP pK<sub>i</sub>: 6.58 ± 0.21; **'3384**: hMT<sub>1</sub> pK<sub>i</sub>: 6.07 ± 0.09; hMT<sub>1</sub>-GTP pK<sub>i</sub>: 7.21 ± 0.03; hMT<sub>2</sub> pK<sub>i</sub>: 5.43 ± 0.08; hMT<sub>2</sub>-GTP pK<sub>i</sub>: 6.21 ± 0.04; mMT<sub>1</sub> pK<sub>i</sub>: 6.51 ± 0.07; mMT<sub>1</sub>-GTP pK<sub>i</sub>: 7.01 ± 0.04; mMT<sub>2</sub> pK<sub>i</sub>: 5.67 ± 0.03; mMT<sub>2</sub>-GTP pK<sub>i</sub>: 6.17 ± 0.08). pK<sub>i</sub> values were derived from competition curves fitted to a one-site model (control: solid lines, GTP: dashed lines), however a comparison of fits determined that a two-site model (dotted lines) was preferred for **'7447** binding to the hMT<sub>1</sub> (**a**: pIC<sub>50Hi</sub>: 7.12 ± 0.10, pIC<sub>50Lo</sub>: 4.75 ± 0.15) and mMT<sub>1</sub> (**e**: pIC<sub>50Hi</sub>: 6.71 ± 0.15, pIC<sub>50Lo</sub>: 4.87 ± 0.31) in control conditions. Leftward shift in affinity for G protein-decoupled (due to GTP and Na<sup>+</sup>) versus coupled receptors indicates inverse agonist apparent efficacy for competitive compounds. Data represent mean ± s.e.m. of five independent determinations **UCSF7447 ('7447), UCSF3384 ('3384)**.



**Figure A.4.11. Affinity, efficacy, and potency of MT<sub>2</sub>-selective agonist.**

(a-d) Competition of '4226 for 2-[<sup>125</sup>I]-iodomelatonin binding on CHO cells stably expressing either the hMT<sub>1</sub> (a), hMT<sub>2</sub> (c), mMT<sub>1</sub>(b), or mMT<sub>2</sub> (d) receptors in the absence (hMT<sub>1</sub> pK<sub>i</sub>: 6.46±0.07; hMT<sub>2</sub> pK<sub>i</sub>: 8.16± 0.03; mMT<sub>1</sub> pK<sub>i</sub>: 6.49 ± 0.08; mMT<sub>2</sub> pK<sub>i</sub>:

6.69 ± 0.07) and presence (hMT<sub>1</sub>-GTP pK<sub>i</sub>: 6.23 ± 0.05; hMT<sub>2</sub>-GTP pK<sub>i</sub>: 7.38 ± 0.05; mMT<sub>1</sub>-GTP pK<sub>i</sub>: 5.91 ± 0.06; mMT<sub>2</sub>-GTP pK<sub>i</sub>: 5.99 ± 0.03) of 100 μM GTP, 1 mM EDTA.Na<sub>2</sub>, and 150 mM NaCl. GTP and Na<sup>+</sup> uncouples G proteins from melatonin receptors promoting inactive conformations. Inactive receptor conformations lower affinity for agonists (rightward shifts). Data represent mean ± S.E.M. of five independent determinations. **(e,f)** Concentration-response curves on hMT<sub>1</sub> or hMT<sub>2</sub> receptors **(e)** and mMT<sub>1</sub> or mMT<sub>2</sub> **(f)** transiently-expressed in HEK cells, monitoring isoproterenol-stimulated cAMP production for hMT<sub>1</sub> (pEC<sub>50</sub>: 6.83 ± 0.17, E<sub>max</sub>: 79 ± 3 %; n = 4), hMT<sub>2</sub> (pEC<sub>50</sub>: 8.15 ± 0.09, E<sub>max</sub>: 89 ± 3 %; n = 4), mMT<sub>1</sub> (pEC<sub>50</sub>: 7.77 ± 0.11, E<sub>max</sub>: 65 ± 3 %; n = 8), and mMT<sub>2</sub> (pEC<sub>50</sub>: 8.23 ± 0.16, E<sub>max</sub>: 39 ± 2 %; n = 8). Data were normalized to maximal melatonin effect and represent mean ± S.E.M. of indicated number (n) of biologically independent experiments run in triplicate. **(g)** Dose-response curves for G<sub>αi3</sub> activation using BRET2 assay for the endogenous ligand melatonin (MLT) (pEC<sub>50</sub> = 9.33 ± 0.12 and 8.93 ± 0.16 at hMT<sub>1</sub> and hMT<sub>2</sub>, respectively) and for '4226 (pEC<sub>50</sub> = 6.26 ± 0.33 and 8.22 ± 0.27 at hMT<sub>1</sub> and hMT<sub>2</sub>, respectively). Net BRET ratio was calculated by subtracting the GFP/RLuc ratio per well from the GFP/RLuc ratio in wells stimulated with buffer. Data represent mean ± s.e.m. of three biologically independent experiments run in triplicate. **UCSF4226 ('4226)**



**Figure A.4.12. LC/MS of Three In vivo-tested Molecules.**

Expected/observed masses with >95% purity: **a) '7447:** 363.6/363.0 (retention time 4.77 min), **b) '3384:** 292.4/293.2 (retention time 4.73 min), **c) '4226:** 293.2/293.0 (retention time 3.59 min)



## 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:

*Reed Stein*

248AF93E43C9469...

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

6/8/2020

\_\_\_\_\_  
Date