UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Ligand Desolvation in Molecular Docking (How, Why, and With What?)

Permalink

https://escholarship.org/uc/item/24t6017j

Author

Mysinger, Michael McLeod

Publication Date

2012

Peer reviewed|Thesis/dissertation

Ligand Desolvation in Molecular Docking (How, Why, and With What?)

by

Michael McLeod Mysinger

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences & Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedicated To My Father, Jim Lee Mysinger

Acknowledgements

Foremost I thank my advisor, Brian Shoichet, the ultimate scientific Zen Master. The breadth of this dissertation I owe to his high expectations; the clarity I owe to his prowess with the written word. As a mentor, he is honest yet authoritative, persuasive yet convincible. I also thank John Irwin, a docking guru constantly available to guide any new padawan. His passion for free chemical tools forms the foundation upon which the final chapter is built.

I am grateful to the guidance of my committee, from which Jim Wells wielded the reality of experiment, whilst Matt Jacobson enveloped me in the mantle of reason. Faced with an experimental void in chapter two, we were rescued by the chemokine expertise of Brian Volkman and his proficient side-kick Josh Ziarek. Yet that story would not have been possible without the equal contribution of Dahlia Weiss, an ideal teammate whose insight is only surpassed by her ability to finish any task. On that same twisting journey, I thank Allison Doak, who persevered to gather a preponderance of evidence against small-molecule aggregation. The proverbial icing on the cake was provided by others collaborators in chapter two, namely Joel Karpiak, Stéphanie Gravel, and Nikolaus Heveker. In his brief stay, Michael Carchia gave our docking code warp speed.

Michael Keiser, as my rotation mentor, introduced me to the brilliance and camaraderie of Shoichet lab culture. Discussions with Jens Carlsson, followed by several chapters of algebra and calculus, forced me to really understand the limits and assumptions of our ligand desolvation method. Kaushik Raha and Johannes Hermann taught me to dock, but later discussions with Peter Kolb really honed my understanding. Niu Huang inspired

my drive for docking automation, and his work is the genesis of chapter three. Sarah Boyce, Rafaela Ferreira, Oliv Eidam, and Magdalena Korczynska were early adopters (victims?) of both my automation tools and my crazy scoring schemes. I thank Sunil Koovakkat , Matthew Merski, and Elisabet Gregori Puigjane for great science discussions. For the memories and more, I thank Alan Graves, Yu Chen, Denise Teotico, Kerim Babaoglu, Brian Feng, Jerome Hert, Kristin Coan, Christian and Veena Laggner, Hao Fan, Gabriel Rocklin, Henry Lin, Sarah Barelier, Marcus Fischer, Nir London, Susan Mashiyama.

After I served six months as system administrator, Pascal Wassam and Teague Sterling carefully fixed everything I broke. My fantastic program administrator, Debbie Acoba-Idlebi, kept me walking the line towards graduation.

I thank my family, whose understanding and support helped me weather every tempest. My mom, Alice McLeod, was always there to pick me when I fell, and to freely give her honest opinion on almost anything. My little sister, Meredith Walker, has grown into a strong woman with my father's infectious positive attitude.

Dedicated to my father, Jim Lee Mysinger, whose will and determination even held back melanoma for twenty years, until the Eve of Christmas in 1999. When times are hard, I still hear his voice clearly in my head, "When the going gets tough, the tough get going."

The text of Chapter 1 is adapted with permission from:

Mysinger, M. M.; et al. "Rapid Context-Dependent Ligand Desolvation in Molecular Docking" *J Chem Inf Model* 50: 1561-73 (2010). http://dx.doi.org/10.1021/ci100214a

Copyright 2010 American Chemical Society. The supporting information for this paper has been included as Appendix A.1.

The text of Chapter 2 is adapted with permission from:

> Mysinger, M. M.[†]; Weiss, D. R.[†]; Ziarek, J. J.[†] et al. "Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4." *PNAS* 109: 5517-22 (2012). http://dx.doi.org/10.1073/pnas.1120431109

The supporting information for this paper has been included as Appendix A.2.

The text of Chapter 3 is adapted with permission from the Journal of Medicinal Chemistry,

as accepted for publication:

> Mysinger, M. M.; et al. "DUD-Enhanced – Better Ligands and Decoys for Better Benchmarking" *J Med Chem* (accepted)

Unpublished work copyright 2012 American Chemical Society. The supporting information

for this paper has been included as Appendix A.3.

Abstract

Ligand Desolvation in Molecular Docking (How, Why, and With What?)

Michael McLeod Mysinger

Ultimately, our bodies are biochemical factories of diabolical complexity. As scaffolds, reactors, engines, and signals, proteins are our essential building-blocks. Drugs, with their potential to alleviate symptoms or cure disease, are often small molecules that amplify or extinguish protein function in just the right way. Molecular docking attempts to understand and predict how those small molecule drugs interact with their protein targets inside us. In isolation, both ligand and protein are bathed in water – to bind one another, some water must necessarily depart. At its core, this dissertation is about how to account for desolvation of the ligand upon protein binding. To highlight why ligand desolvation is important, we discover new chemical probes for CXCR4, a protein target implicated in cancer and HIV. En route, we create the LogAUC metric and the DUD-E benchmarking dataset to better assess retrospective docking performance.

Our rapid context-dependent ligand desolvation scoring term relates the Generalized-Born effective Born radii for every ligand atom to a fractional desolvation, and then uses this fraction to scale an atom-by-atom decomposition of the full transfer free energy. In a test that fails with no desolvation, our method properly discriminates ligands from highly charges molecules. The method is also flexible, performing well whether the protein binding site is charged or neutral, open or closed.

We first retrospectively test ligand desolvation on the 40 original DUD targets, but discover many ways to improve that benchmark. So we construct DUD-E, an improved set with more diverse and biomedically relevant targets, totaling 102 proteins with 22,886 clustered ligands, each with 50 property-matched decoys. To ensure chemotype diversity we cluster the ligands by Bemis-Murcko frameworks. To improve decoys, we add net charge as an additional matched physico-chemical property, and only include the most dissimilar decoys by topology. To test our method prospectively, we screen both a homology model and then a crystal structure of CXCR4. Several of our novel scaffolds are potent and relatively small, with IC_{50} values as low as 306 nM, ligand efficiencies as high as 0.36, and substantial efficacy in blocking cellular chemotaxis.

Table of Contents

List of Tables

List of Figures

Introduction

My first task as a new lab member was simple, just test an implementation of ligand desolvation we had already been using several years on our new DUD benchmarking dataset spanning 40 protein targets[1]. After a few weeks learning the mechanics of molecular docking, the answer was clear, we were better off not using ligand desolvation at all. This was my one true chance to listen, to abandon the long slog of algorithm development for the easy and level path of new ligand discovery. Instead I stubbornly plowed ahead, following a tortuous path that touched the deepest dungeons of our docking toolchain.

But I have skipped a little too far ahead, as I have not yet described ligand desolvation and why it is important. The tiniest of molecules, such as aspirin or caffeine, can have profound effects on our biochemistry, easing our symptoms or even curing disease. These molecules often act by binding a target protein, such as cyclooxygenase-2 in the case of aspirin, altering its normal function. A computational tool that predicts how and when such tiny molecules bind their larger protein targets has enormous potential, to predict new molecules that perturb countless protein functions. These small molecules may just be useful to manipulate biochemical pathways, advancing our scientific understanding, or they might be optimized into pharmaceuticals, fundamentally altering our lives.

Molecular docking is a tool to solve this binding problem, ideally with the speed to search and then prioritize from among the millions of small molecules currently purchasable[2]. To achieve speed, a somewhat crude scoring function is used both to pick the best ligand binding orientation (pose) and to rank the small molecules for their predicted binding potential. The physics-based scoring function in DOCK evolved[3] to consist of two

terms representing shape-complementarity (van der Waals) and charge-charge interactions (electrostatics). While they capture the basic ligand-protein interaction energy, they neglect that both ligand and protein interact with a third party.

> Water is life's matter and matrix, mother and medium. There is no life without water.
>
> -- Albert Szent-Gyorgyi, Nobel Prize winner in Medicine

Water solvates both ligand and protein, and must depart to allow binding. For charged molecules this desolvation energy is large, opposing ligand binding. Yet the electrostatics energies are also large for charged molecules, often favoring binding. The small net difference determines whether a ligand will bind the protein tightly or prefer to remain alone in water. For small molecules of a similar size that bind the same protein site, the relative difference in receptor desolvation is small. So we focus on computing the desolvation of the ligands, both accurately and at docking speeds.

Without the counter-balance of ligand desolvation, electrostatics overpowers the docking score, such that the molecules with the highest charges get the best scores. To first approximation, particularly in enclosed binding sites, the move from bulk water to burial inside the protein is much like the transfer from a water phase to an organic one. So the full transfer free energy was used as an approximation to ligand desolvation[4, 5]. This indeed counter-balances electrostatics and removes the highly charged molecules, but goes too far, over-penalizing more open binding sites. This balance between the magnitude of the ligand desolvation term and electrostatics would turn out to be critical to making ligand desolvation perform well.

Meanwhile, my advisor Brian Shoichet recognized the shortcomings of full desolvation, and devised a solution. He would decompose the full transfer free energy on an atomic basis and scale each atom's contribution by its own context-dependent desolvation

fraction. After receiving tenure in 2002, he even had time to code this ligand desolvation method *himself*.

If you recall, I mentioned earlier that when I started this project the initial benchmarking tests showed that we got much better results if we just ignored ligand desolvation. After ruling out my inexperience with DOCK as the culprit, we were faced with three possible sources of ligand desolvation's failure – either the theory, the implementation, or the benchmarking set was flawed. So I checked each one, and *all three* were wrong. The theory was missing the ligand radius, and it seems like the integral boundary condition was fudged to achieve the right magnitude. The implementation, while exemplary Fortran 77 code, was not used during ligand minimization and was not available in any debugging tools. More critically, it was not interpolated, and so its high magnitude, stair-like jumps occasionally caused a small polar group (tail) to relegate the rest of a large molecule into an unfavorable position (wagging dog). A clever hack to mitigate this effect unfortunately confounded and befuddled me through months of early testing results. And the benchmarking set, well that rabbit hole eventually became chapter three of this thesis.

As all of the puzzle pieces began to click into place, a situation similar to Goldilocks and the Three Bears[6] emerged. Desolvation energies can strongly depend on the protein surface boundary, even more than the type of theory[7]. Context-dependent desolvation was still too strong, relative to electrostatics, when using the larger solvent-accessible surface of the protein. So we switched to the smallest surface (van der Waals) and then the next smallest (Connolly molecular surface), which were both too weak. Only by accounting for the solvent-excluded volume that occurs as ligand approaches protein were we able to adequately balance ligand desolvation and electrostatics. Technically this required calculating

and then integrating over millions of different surfaces, and to convince Brian to use the method, I had to speed it up by over 100 fold.

The final obstacle was the benchmarking set. Originally, ignoring desolvation was strongly preferred in charged binding sites. Like attracts like, so charged binding sites have charged ligands. Yet the DUD decoys were more neutral, as only 15% of all DUD decoys were charged versus 42% of the ligands[8]. So with no desolvation in the charged binding sites, the charged ligands got better, uncompensated scores than the more neutral decoys. This is the same performance bias that motivated property-matching of decoys to ligands in DUD, except we now needed to include net charge as an additional property. Rebuilding the decoys to be charge-matched, especially in an automated way, took at least a year of work by itself, but would also be the genesis of DUD-E in chapter three. Still to get reliable results, we rebuilt the docking inputs from scratch, careful to be consistent for both ligands and decoys. Many of my improvements to the docking toolchain on the ligand side became part of ZINC[2], while those on the receptor side bolstered our automated DOCKBlaster protocol[9].

I. Guide to the Chapters

I have adapted three published first or co-first author journal articles to form this thesis. I introduce each chapter by a short "gloss", which attempts to both summarize and to provide the larger context of that chapter.

The first chapter introduces our method and implementation of a fast context-dependent ligand desolvation scoring term. We show that, as expected, it discriminates against highly charge molecules and find versatile docking performance over the 40 tested DUD targets. But why did we develop the ligand desolvation method, if not to find new ligands for challenging protein targets? In the second chapter, we apply our ligand

desolvation method prospectively to discover potent and efficient small molecule inhibitors of CXCR4, a difficult protein-protein interface target strongly implicated in both cancer and HIV. Additionally, in parallel virtual screens to homology model and crystal structure, we show that CXCR4 was near the edge of our ability to find new ligands by homology, because it was far from known structures with scarce mutational data. But with what tool can new docking methods be tested in the future? Chapter three rebuilds the DUD benchmarking dataset from the ground up, while extending it to 102 targets with 22,886 ligands. We then use this DUD-E set to show that our physics-based ligand desolvation method solidly out-performs using no desolvation term. We finally test a thinner dielectric layer for electrostatics, returning full circle to the balance with ligand desolvation.

II. References

1. Huang, N; Shoichet, BK; Irwin, JJ. "Benchmarking Sets for Molecular Docking" *J Med Chem* **2006** 49: 6789-6801.

2. Irwin, JJ; Sterling, T; Mysinger, MM; Bolstad, ES; Coleman, RG. "Zinc - a Free Tool to Discover Chemistry for Biology" *J Chem Inf Model* **2012** doi: 10.1021/ci3001277.

3. Meng, EC; Shoichet, BK; Kuntz, ID. "Automated Docking with Grid-Based Energy Evaluation" *J Comput Chem* **1992** 13: 505-524.

4. Shoichet, BK; Leach, AR; Kuntz, ID. "Ligand Solvation in Molecular Docking" *Proteins* **1999** 34: 4-16.

5. Wei, BQ; Baase, WA; Weaver, LH; Matthews, BW; Shoichet, BK. "A Model Binding Site for Testing Scoring Functions in Molecular Docking" *J Mol Biol* **2002** 322: 339-355.

6. Hassall, J. *The Old Nursery Stories and Rhymes.* Blackie and Son: 1904.

7. Scarsi, M; Apostolakis, J; Caflisch, A. "Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions" *J Phys Chem A* **1997** 101: 8098-8106.

8. Irwin, JJ. "Community Benchmarks for Virtual Screening" *J Comput-Aided Mol Des* **2008** 22: 193-199.

9. Irwin, JJ; Shoichet, BK; Mysinger, MM; Huang, N; Colizzi, F; Wassam, P; Cao, Y. "Automated Docking Screens: A Feasibility Study" *J Med Chem* **2009** 52: 5712-5720.

III. List of My Related Publications

Mysinger, MM; Shoichet, BK. "Rapid Context-Dependent Ligand Desolvation in Molecular Docking" *J Chem Inf Model* **2010** 50: 1561-1573.

Mysinger, MM; Weiss, DR; Ziarek, JJ; Gravel, S; Doak, AK; Karpiak, J; Heveker, N; Shoichet, BK; Volkman, BF. "Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4" *Proc Natl Acad Sci USA* **2012** 109: 5517-5522.

Mysinger, MM; Carchia, M; Irwin, JJ; Shoichet, BK. "DUD-Enhanced – Better Ligands and Decoys for Better Benchmarking" *J Med Chem* **2012** (accepted).

Irwin, JJ; Shoichet, BK; Mysinger, MM; Huang, N; Colizzi, F; Wassam, P; Cao, Y. "Automated Docking Screens: A Feasibility Study" *J Med Chem* **2009** 52: 5712-5720.

Irwin, JJ; Sterling, T; Mysinger, MM; Bolstad, ES; Coleman, RG. "ZINC - A Free Tool to Discover Chemistry for Biology" *J Chem Inf Model* **2012** doi: 10.1021/ci3001277.

Gloss to Chapter 1

Through stubborn persistence and blind determination, we overcame many obstacles highlighted in the introduction, to eventually create our context-dependent ligand desolvation scoring term. The following chapter describes how we implemented and tested the resulting method. We first developed the LogAUC metric to quantify a docking algorithm's ability to find (enrich) ligands before non-ligands, weighting early ligand retrieval more favorably. As net charge is a key property in desolvation, we generated charge-matched decoys for the DUD benchmarking dataset covering 40 protein targets. Upon testing our ligand desolvation term, we found that by incorporating atom-by-atom context we created a versatile method that performs well in many different protein binding sites – be they charged or neutral, open or closed.

To compute ligand desolvation, we start with pre-calculated full transfer free energy, which emulates the ligand's movement from water to protein in the binding process. Because some parts of the ligand interact with protein more and others with water more, we scale each atom's transfer energy by its fractional desolvation. This fractional desolvation is pre-computed by numerical integration, based on Generalized-Born theory, by our sevsolv program. To achieve the right balance with electrostatics, we account for the solvent-excluded volume (SEV) that forms when a ligand atom gets so close to the protein that even a single water cannot fit between them. This scoring term is physics based, and can be mostly pre-calculated, enabling rapid context-dependent ligand desolvation to be added to any docking program.

We put the method through an extensive battery of tests, were it proved both versatile and high-performing. Our ligand desolvation term discriminated against highly charged molecules, as expected. Against challenging charge-matched decoys across 40 targets, it performs slightly better than using no desolvation term, on average. More importantly, the context-dependence helps it perform consistently across protein types. Pose fidelity tests show it is working adequately, and that it enriches ligands for the right reasons. Even against experimentally determined non-binders, it still strongly enriches known ligands.

Chapter 1:

Rapid Context-Dependent Ligand Desolvation

in Molecular Docking

Michael M. Mysinger[1], Brian K. Shoichet[1]*

1   Department of Pharmaceutical Chemistry, University of California San Francisco, 1700 4th

St, San Francisco CA 94143-2550

* Corresponding author (BKS)

The text of Chapter 1 is adapted with permission from:

The supporting information for this paper has been included as Appendix A.1.

1.1 Abstract

In structure-based screens for new ligands, a molecular docking algorithm must rapidly score many molecules in multiple configurations, accounting for both the ligand's interactions with receptor and its competing interactions with solvent. Here we explore a context-dependent ligand desolvation scoring term for molecular docking. We relate the Generalized-Born effective Born radii for every ligand atom to a fractional desolvation, and then use this fraction to scale an atom-by-atom decomposition of the full transfer free energy. The fractional desolvation is pre-computed on a scoring grid by numerically integrating over the volume of receptor proximal, weighted by distance. To test this method's performance, we dock ligands versus property-matched decoys over 40 DUD targets. Context-dependent desolvation better enriches ligands compared to both the raw full transfer free energy penalty and compared to ignoring desolvation altogether, though the improvement is modest. More compellingly, the new method improves docking performance across receptor types. Thus, whereas entirely ignoring desolvation works best for charged sites, and over-penalizing with full desolvation works well for neutral sites, the physically more correct context-dependent ligand desolvation is competitive across both types of targets. The method also reliably discriminates ligands from highly charged molecules, where ignoring desolvation performs poorly. Since this context-dependent ligand desolvation may be pre-calculated, it improves docking reliability with minimal cost to calculation time, and may be readily incorporated into any physics-based docking program.

1.2 Introduction

Molecular docking is widely used to computationally screen large chemical libraries for small molecules that complement receptors of known structure[1-4]. Complementarity and ranking are calculated by a scoring function that is either based on empirically-fit descriptors[5-9], knowledge-based potential functions[10-12], or physics-based terms[13-17]. Physics-based scoring functions borrow forcefield derived terms, such as van der Waals (vdW) and electrostatics to calculate the protein-ligand interaction energy[18, 19]. More problematic has been the representation of bulk water, which solvates the unbound ligand and protein, and differentially solvates the protein-ligand complex. Proper treatment of this net desolvation is essential, especially in polar or charged systems. This is because a large desolvation penalty opposes a large electrostatic interaction energy, while the much smaller net difference has a strong influence on binding affinity[20-22].

Desolvation is often subdivided into a polar electrostatic component and a non-polar component due to dispersion, hydrophobicity, and cavitation[23]. At a higher level of theory, when more computational power is available to analyze each ligand pose, molecular mechanics combined with Generalized-Born (GB) or Poisson-Boltzmann (PB) theory correlates to experimental solvation and occasionally even to binding affinities[24-29]. However, these methods remain at least five orders of magnitude slower than docking, which routinely screens more than 10^6 molecules, each in 10^3-10^5 configurations, highlighting the need for desolvation methods fast enough to be relevant to docking[30].

Early docking methods neglected desolvation entirely and thus favored large and highly charged molecules, since extra size and polarity increases the interaction energy without a compensating penalty for desolvation[31]. To remedy these problems without sacrificing speed, different approximations of desolvation have been investigated. A

common approach estimates the desolvation penalty as proportional to the solvent-accessible surface area scaled by atomic solvation parameters[11, 32-34]. Whereas surface area is often approximated to be proportional to the non-polar component of desolvation, it fails to capture the electrostatic component properly. An alternative to surface area is provided by volume-based Gaussian envelopes[35, 36], though these methods are only loosely related to electrostatics. In a recent effort to capture desolvation empirically, 1179 adjustable solvation parameters were fit to three different levels of potentials of mean force, where two levels were unique to every pair of 17 atom types[37]. Although well explored, both the variability in these atom-type approaches and the increasing number of parameters reflect the difficulty in representing the underlying physical terms. Another empirical approach approximates desolvation penalties based on displacement of ordered waters in the binding site[8, 38-40], though calculating the energetic cost or advantage of such displacement remains an area of active research.

The challenge of using physics-based desolvation in docking stems from the difficulties in calculating the energies accurately enough to balance what are often large magnitude terms, and doing so fast enough to be relevant. Thus other investigators have tried decomposing the electrostatics of ligand binding into components arising from partial ligand desolvation, partial receptor desolvation, and screened protein-ligand interactions. A faster implicit solvent model was recently generated by combining Lorentz-Debye-Sack theory with a first-shell approximation, using the fraction of solvent-accessible surface area to scale empirical covalent radii[41]. Though promising, this method has a non-linear relationship with what are thought to be more accurate Poisson-Boltzmann energies and its performance in docking screens remains to be fully explored. Majeux *et al.* estimate both ligand and receptor desolvation for fragments by combining the Coulomb field

approximation with a first-shell approximation, using empirical scaling to reduce the impact of those approximations[42]. Interestingly they also use grid-based numerical integration over volume elements, but for flexible ligands this has been slow as it requires a new ligand desolvation grid for each conformation.

In prior desolvation methods from our group[43] we modeled bound ligand as fully desolvated, approximating the ligand desolvation penalty as the full transfer free energy from high to low dielectric computed by the AMSOL program[44]. Whereas this removes the bias towards highly charged molecules, the magnitude of the desolvation penalty is often over-estimated, being ideal only when the ligand is completely surrounded by the protein so that it resembles being fully immersed in organic solvent. A preferable idea would be to scale the full desolvation contribution of each atom, proportional to how much it is embedded in the low-dielectric medium (i.e. protein). This scaling factor is offered by Generalized-Born theory, where each atom has an effective Born radius that encapsulates geometric, context-dependent information about that atoms relative desolvation[45]. We begin by relating the GB effective Born radii for every ligand atom to a fractional desolvation, and then use this fraction to scale an atom-by-atom decomposition of the full transfer free energy. To make this fast enough for docking, we pre-compute the fractional desolvation on a regular lattice grid for positions in the protein binding site. Correspondingly, we pre-calculate the atom-by-atom transfer free energy for each molecule in the docking library. This results in a rapid context-dependent ligand desolvation model for molecular docking.

The exact representation of the dielectric boundary can have more influence on the results than the difference between GB and PB[46], but its explicit incorporation into molecular docking would be prohibitively expensive[27]. We therefore investigated a treatment of the local solvent-excluded volume (SEV) that captures the first-shell effects of solvent

exclusion, yet can be pre-calculated. This context-dependent desolvation term approaches

the full transfer free energy as the ligand becomes fully engulfed in protein, and approaches

the exact GB energy in the limit of a single ligand atom. While these are attractive theoretical

properties, we needed to investigate if we can balance desolvation and interaction energy

accurately enough to allow for enrichment of ligands in docking screens. Thus we measured

the virtual screening performance of several ligand desolvation approximations on the 40

DUD benchmarking targets against their property-matched decoys, in an effort to reduce

background database bias[47]. To protect against database extrema, we also screened the

ligands against highly-charged molecules pulled from the free ZINC database[48]. To better

account for early enrichment when measuring virtual screening performance, we introduce

the LogAUC metric. We finally investigate screening performance in two model binding sites

against experimentally confirmed, as opposed to presumed, decoys[43, 49].

1.3 Methods

I. Fractional Ligand Desolvation Scoring Term

In GB theory[45], the effective Born radius α_i of a given atom can be expressed in

terms of the normal atomic radii a_i and a volume integral over all low dielectric space outside

the atom itself, where the contribution of each occupied volume element dV diminishes with

the distance r between it and atom i as follows:

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{4\pi} \int\limits_{in,\, r>a_i} \frac{1}{r^4} dV \qquad \textbf{Equation (1)}$$

A simple rearrangement allows us to express the fractional desolvation D_i in terms of the

volume integral

$$D_i = 1 - \frac{a_i}{\alpha_i} = \frac{a_i}{4\pi} \int\limits_{in,r>a_i} \frac{1}{r^4} dV \qquad \textbf{Equation (2)}$$

Checking boundary cases, when α_i approaches a_i then D_i goes to 0 indicating the atom is

completely surrounded by solvent, and conversely as α_i approaches ∞ then D_i goes to 1

indicating complete desolvation. We can numerically integrate to find the fractional

desolvation by constructing a three-dimensional grid and using the equation

$$D_i = \frac{a_i dV}{4\pi} \sum_k^{in,r>a_i} \frac{1}{r_{ik}^4} \qquad \textbf{Equation (3)}$$

where dV is the volume of one grid element and the summation runs over grid points inside

the low dielectric region (**Figure 1.1.A**).

To approximate ligand desolvation we previously assumed that the ligand becomes

fully desolvated in the protein binding site[31, 43], thus treating the desolvation penalty as the

full transfer free energy from high-dielectric solvent to low-dielectric protein, as computed

by AMSOL[44]. However, if we partition this full transfer free energy to the individual ligand

atoms, obtaining ΔG_i^{trans}, then we can multiply by the fractional desolvation D_i of each ligand

atom to estimate the total ligand desolvation penalty as

$$\Delta G_{desolv}^L = \sum_i^{ligand} \Delta G_i^{trans} D_i \qquad \textbf{Equation (4)}$$

This is an attractive way to account for ligand desolvation during molecular docking since

the atomic desolvation penalties may be pre-computed, as can the fractional desolvation at

any position in the binding site. During docking all that is required is multiplying the atomic

transfer free energies by the fractional desolvation interpolated from the pre-calculated

scoring grid.

II. Fractional Ligand Desolvation Implementation

When the ligand and protein low-dielectric volumes do not overlap substantially, we can decompose the fractional desolvation into ligand, receptor, and solvent-excluded volume (SEV) components. We can readily pre-calculate the receptor component and use a

Figure 1.1 Volume-Based Desolvation Implementation

A) Fractional desolvation numerical integration scheme: a_i = radius of ligand atom i, r_{ik} = distance from ligand atom i to volume element k.

B) Incorporating solvent-excluded volume: At close distance, a solvent-excluded volume (SEV) forms between protein (green) and ligand (yellow) where water probes (light blue) do not fit. We pre-calculate the atomic SEV region (grey) while neglecting the neighboring SEV region (darker blue).

C) Fractional desolvation plotted versus distance inside (+X) or outside (-X) an idealized solid low-dielectric slab: Atomic SEV method (cyan) plotted versus molecular surface method (red).

16
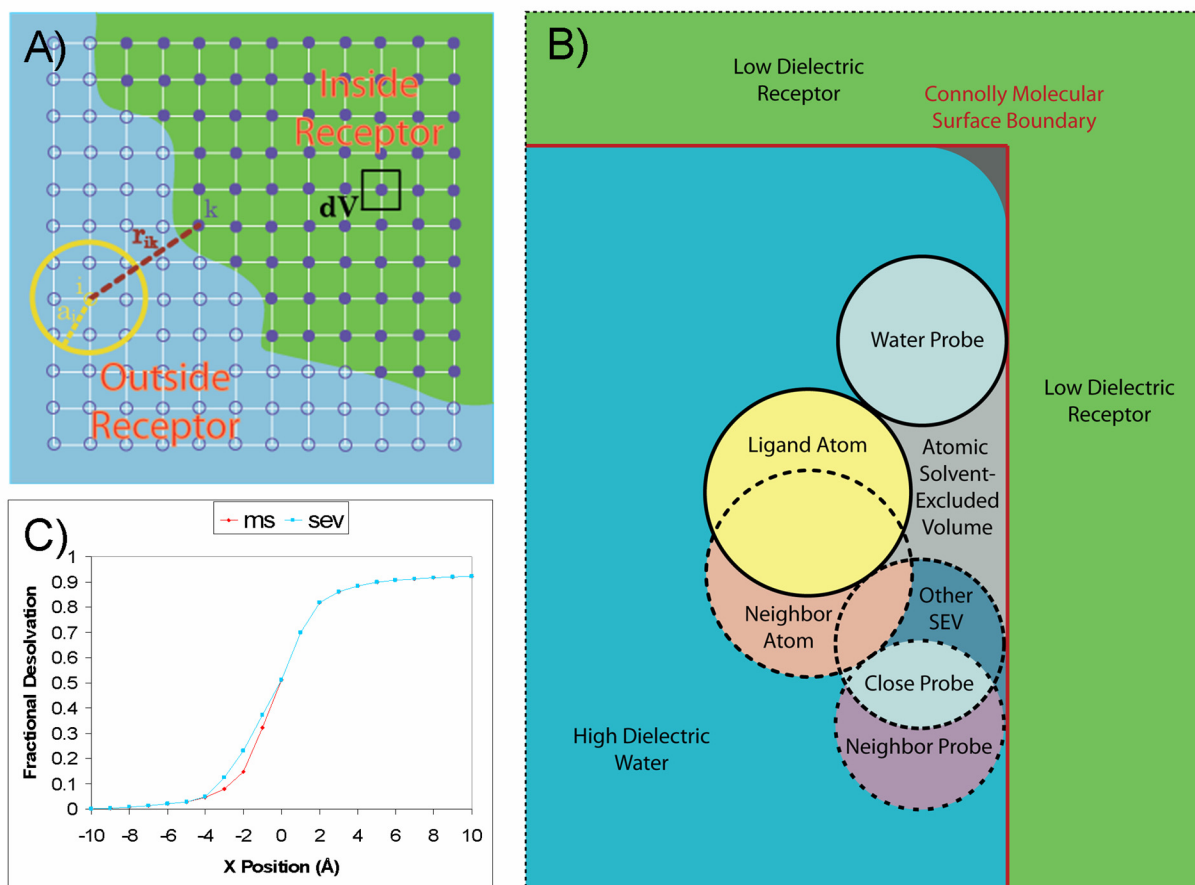
first-shell approximation to pre-calculate the SEV component. Admittedly, the portion of any ligand atoms' desolvation that depends on the position of *other* ligand atoms remains expensive to compute, because it varies between ligand conformations. When we multiply fractional desolvation by atomic transfer free energy (**Equation 4**), we therefore neglect the ligand component. This can lead to errors in some circumstances, but since this term is included in the AMSOL-derived[44] transfer free energies, expressing desolvation as a percentage of the initial transfer free energy partly compensates for this deficit (**Table A.1.6** gives an atom-by-atom estimate of this error for one challenging molecule). With these approximations, the fractional desolvation at every point depends only on the geometry of the receptor and can be pre-calculated.

We initially used the dielectric boundary of the receptor alone to compute the fractional desolvation grid. However, solvation energies can strongly depend on the dielectric boundary, even more than the exact type of implicit solvation model used[46]. The ideal dielectric boundary for ligand desolvation is the molecular surface of the protein-ligand complex, but calculating this on-the-fly at the time of docking would be prohibitive, so we investigated a faster method. Upon ligand binding, a solvent-excluded volume, or gap, forms between the ligand and receptor where even a single water molecule cannot fit (**Figure 1.1.B**). This SEV affects the fractional desolvation of the ligand, but generally depends on the pose of the ligand in the protein which is not known until the ligand is docked. To enable fractional desolvation pre-calculation, we assume for each ligand atom that we can independently account for the SEV due to it alone (grey region) while neglecting the SEV due to other atoms in the ligand (darker blue region). With this atomic SEV correction, our method reproduces the exact Generalized-Born energy in the limiting case of single ligand

atom. But even for molecules, neglecting distal excluded volume elements has only a modest effect since this term decays at $1/r^4$.

The numerical integration of the fractional desolvation is encoded in a DOCK accessory program sevsolv. Pre-computing the atomic SEV correction involves re-calculating the molecular surface for each ligand atom position in the grid and then numerically integrating using that new surface. The program computes the receptor molecular surface using the inkblot algorithm[50], which generates evenly spaced points on a sphere at the solvent-accessible surface (SAS) distance (vdW + probe radius) using the golden section spiral algorithm, before removing inaccessible surface points inside other receptor atoms. The grid inside the SAS is tentatively assigned to low-dielectric, and then a water probe is placed at each accessible SAS point and the region inside each probe is re-assigned back to high-dielectric. This effectively blots back to the molecular surface starting from the accessible SAS points. For the atomic SEV we keep track of how many receptor accessible SAS points are marking each grid point as solvent accessible, then for each grid point a ligand atom is placed there and a cubing algorithm finds all newly inaccessible receptor surface points. If none are found then this ligand atom position is either completely surrounded by solvent or completely buried in protein, and there is no atomic SEV correction. Otherwise, we find the ligand's accessible SAS points and increment the marks of the grid points within a probe radius. For the newly inaccessible receptor SAS points we decrement the marks of grid points within probe radius and make the grid point low-dielectric if no marks remain to cover it. Finally, we blot the region inside the ligand vdW as low-dielectric and compute the numerical integral over this new SEV surface. The current implementation takes under half an hour to generate the atomic SEV desolvation grid on a single 2008 CPU core for any of the 40 targets in DUD.

18

Here we use a 1.4 Å water probe radius and 1000 SAS points per atom, while the integration cutoff is 10 Å with a grid resolution of 0.5 Å. These choices were made as a compromise between speed and accuracy; the cost of the calculation rises as the third power of each of these terms. Since the cost of pre-calculating the desolvation grid is only paid once, before docking, should one want more accuracy it is available. For instance, for a probe atom buried in a 48 Å cube—a case where one might see maximal effects of truncation—the desolvation penalty changed by 3% in going from 0.5 Å to 0.33 Å resolution, and by an additional 1% to going to 0.167 Å resolution. Correspondingly, the penalty changed by 6% in going from a 10 Å to a 20 Å cutoff (**Figure A.1.4**). The input radii were set to 1.6, 1.65, 1.9, 1.9, 1.9, 0, and 1.6 Å for O, N, C, S, P, H, and other protein atoms respectively. Unless noted otherwise, the assumed Born radius for all ligand atoms is 1.4 Å, a value intermediate between hydrogens and heavy atoms, and was chosen to enable one fractional desolvation grid to be used for all ligand atoms during docking; other implementations may be imagined where different desolvation grids are pre-calculated for each ligand atom type.

Similarly, the atom-by-atom partitioning of the transfer free energy can be pre-calculated and stored in the ligand database prior to docking. Here we further assume that the atomic transfer free energies are similar between all ligand conformers. This assumption could have a substantial effect on overall solvation energies for highly flexible or charged molecules; this is an area of ongoing research. To compute atom-by-atom transfer free energies we perform two AMSOL calculations using the SM5.42R solvation model with the semi-empircal AM1 Hamiltonian[44], subtracting the results in water from those in hexadecane (dielectric = 2.06). From this procedure we obtain separate polar and non-polar atomic

transfer free energies, and the DOCK accessory mol2db stores them into the input flexibase[51] as found in ZINC[48].

We add this ligand desolvation scoring term to DOCK 3.5.54, resulting in the following scoring function:

$$E_{score} = E_{vdW} + E_{elec} + \Delta G_{desolv}^{L} \qquad \textbf{Equation (5)}$$

where the vdW term is based on the AMBER united-atom force field[52] and the electrostatics term is based on PB calculations by DELPHI[50], as described previously[13]. The fractional desolvation is tri-linearly interpolated from the pre-computed desolvation scoring grid for each ligand atom position and multiplied by the polar and non-polar desolvation terms read from the ligand flexibase.

III. DUD Database Docking

To reduce enrichment bias and focus on the screening performance of molecular docking itself, the DUD benchmarking set property-matches decoys to known ligands[47]. In using the 40 DUD targets to examine enrichment performance of ligand desolvation methods, we found that we had to property-match ligands and decoys on net formal charge, a term which was originally neglected. Without net charge, the original DUD decoys were much more neutral than the ligands (**Figure 1.2**). To overcome this deficit we recalculated DUD, matching the original ligands to new computational decoys. We property-matched on the original five physical terms plus net molecular charge, creating an automated procedure that may be applied to any user specified set of ligands (see **Chapter 3**).

The ligands and decoy sets of this charge-matched DUD database were built into input dockable flexibases by an updated ZINC[48] procedure. To remove input structure bias, the molecules were converted to smiles by Openeye's OEChem library[53], stereo-chemically

enumerated and built into 3-D structures by Molecular Networks' Corina[54], protonated and tautomerized by Schrödinger's Ligprep[55], and then run through two AMSOL[44] calculations to generate partial charges and atomic desolvation contributions as mentioned previously. Corina[54] was run again to generate aliphatic ring puckers, and each ring is selected in turn to serve as the starting rigid fragment for conformational enumeration by Openeye's Omega[56]. The conformational ensemble and atomic desolvation were combined by mol2db[51] to generate the input docking hierarchies.

Input parameters for docking and grid generation were as described previously[47], except as noted. We preferentially used the semi-automated preparation of the receptor for the 13 targets where it was available. The bump limit was set to allow up to one steric clash, with a DISTMAP grid resolution of 0.333 Å. We performed 50 steps of simplex minimization for the best initial pose. In the electrostatic potential map calculations using DELPHI[50], the exterior and interior dielectrics were set to 78.5 and 2.06 respectively, while the salt concentration was 0.145 M and the ion exclusion radius was 2 Å.

IV. LogAUC Virtual Screening Metric

Virtual screening performance is typically evaluated using enrichment or receiver operator characteristic (ROC) plots[57], which measure the prioritization of annotated ligands versus known or presumed decoys among the highest ranking molecules. Often one wishes to over-weight molecules that rank among the very top of the "hit-list", as these are the ones most likely to be tested for activity in an actual prospective experiment. Several approaches have been introduced to do so[58, 59], here we adopt a variation of the standard ROC plot that, rather than plotting percentage of decoys found vs. percentage of annotated ligands, plots the $\log_{10}$ of the decoys found on the x-axis, which preferentially weights early enrichment.

The area under the curve (AUC) is a well-regarded metric to summarize the overall performance of a virtual screening campaign as a single number[57]. While ROC AUC can be formulated in alternate ways, it can mechanically be constructed by integrating under the ROC curve and interpreted as the fraction of the area under the ROC curve over the area under the best possible ROC curve, which happens to be 1. By analogy, in the case of our semilog ROC plot, we can construct the same fraction of the area under the current log curve over the area under the perfect log curve, and define that fraction as the LogAUC. The lone wrinkle is that the area under the perfect log curve is, in general, infinite. However, if we are practical and limit our focus to a region of log space we can actually measure, say above a certain threshold λ, then the perfect log area is finite in that region.

Formally, we define LogAUC$_\lambda$, where the log area computations run from λ to 1.0, and in this paper we refer to LogAUC$_{0.001}$ as simply LogAUC. For integrating the area under the log curve we use an analytical formula derived by log transforming an individual trapezoidal segment. The y-intercept (b) of the line for a given trapezoidal segment is:

$$b_i = y_{i+1} - x_{i+1}\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right)$$ **Equation (6)**

LogAUC$_\lambda$ is then calculated as follows:

$$LogAUC_\lambda = \frac{\displaystyle\sum_i^{where\ x_i \geq \lambda}(\frac{y_{i+1} - y_i}{\log_e 10}) + b_i\left(\log_{10} x_{i+1} - \log_{10} x_i\right)}{\log_{10}\left(1.0/\lambda\right)}$$ **Equation (7)**

From parallel reasoning based on semilog ROC plots, Clark and Webster-Clark construct the pROC AUC metric[58], which is similar to the numerator of LogAUC except that the area integration is over horizontal bars instead of vertical trapezoids. The advantage of constructing LogAUC as a fraction over the ideal area is that the choice of base for the

logarithm is irrelevant, because changing base simply results in a constant that cancels between numerator and denominator. Also, by explicitly defining the area of interest using λ and integrating vertically, we are able to avoid the singularity at $x_i=0$ encountered in pROC. More importantly the fixed integration area means we can compare LogAUC values across databases of different sizes and across targets with different ratios of actives to inactives. While the recent NSQ_AUC metric[60] also recognizes the need to normalize based on a perfect ROC curve, the multiplicative random curve based rescaling makes the metric harder to rationalize. In contrast LogAUC$_\lambda$ is simply the percentage of the total area below a semilog ROC curve plotted from λ to 1, which intuitively feels like linear AUC, except that it's derived from the semilog plot. The random line on a semilog ROC plot (dashed line, **Figure 4**) occupies only a sliver of the total area, and indeed its LogAUC is just 14.5%. In order to more easily see the performance above random enrichment, we report the "adjusted LogAUC" as the LogAUC minus this random LogAUC, so that positive values mean enrichments better than random and the maximum enrichment occurs at 85.5% adjusted LogAUC.

V. Charge Outliers

As noted previously[31], a substantial problem of neglecting desolvation is that highly charged molecules pollute the top hit lists of polar targets. Since the revised DUD is property-matched on net charge, the enrichment of a given target's ligands versus its own decoys fails to interrogate the ability of the desolvation penalty to penalize these highly charged hit list polluters. To address this issue, we pulled 500 random molecules from ZINC with each net charge of -4, -3, -2, +2, +3, and +4 to create a charge outlier database of 3000 molecules. We then compared the enrichment of the ligands for each target against this

background database of random highly charged molecules, to see how competitive these random charge outliers were under the various scoring functions.

VI. Pose Fidelity

To check how ligand desolvation affected docked ligand poses, we investigated pose fidelity across the 114 protein-ligand complexes in the DOCK5 benchmark[61]. The automated DOCKBlaster[62] procedure was used to prepare all calculations beginning with just the PDB code for 84 out of 114 targets, while the ligand three-letter ligand identifier was needed to build 16 additional targets. For the following targets manual intervention was needed, typically to fix ligand or cofactor parameters: 1FLR, 1IMB, 1AOE, 4COX, 1ETT, 3CPA. The updated ligand flexibase building procedure described above in the DUD Database Docking section was used to generate input ligand databases, except that Schrödinger's Epik[63] replaced Ligprep[55] for protomer and tautomer generation.

VII. Model Binding Site Docking

To test enrichment performance versus true experimental decoys, rather than the presumed DUD decoys, we docked against two model cavity sites where non-binders have been experimentally confirmed. The hydrophobic pocket introduced into T4 lysozyme by the Leu99→Ala substitution (L99A) accepts small apolar ligands. Previously, 73 ligands and 64 experimental decoys have been determined for this cavity[30, 43, 64-67]. In contrast, the cavity introduced into cytochrome C peroxidase by the Trp191→Gly substitution (W191G) accepts charged ligands. Previously, 61 ligands and 26 experimental decoys have been determined for this charged cavity[30, 49, 68, 69]. We analyzed ROC curves for the ligands against either the experimental decoys alone or the experimental decoys seeded into a fragment-like subset of ZINC filtered only by molecular weight (30-250 Da).

1.4 Results

Here we compare the performance of three ligand desolvation approximations: "none", "full", and "sev". "None" lacks any scoring penalty for ligand desolvation, serving as a crude lower bound; while "full" uses the entire transfer free energy from water to hexadecane as the ligand desolvation penalty, serving as an upper bound. Our ligand desolvation term takes the per-atom contributions from the full desolvation penalty and scales them by the fractional desolvation each ligand atom experiences in Generalized-Born theory. The "sev" method computes fractional desolvation while further accounting for the solvent-excluded volume (gap) that arises between a single ligand atom and the receptor at close distances.

To check that the fractional desolvation implementation is reasonable, we first looked at a low-dielectric slab infinite in the yz-plane covering half of all space with the dielectric boundary occurring at x=0 Å. Plotting the fractional desolvation in this idealized system using just the receptor's molecular surface (ms) or the atomic SEV method (sev) shows several key features (**Figure 1.1.C**). The fractional desolvation approaches zero far away from the slab, but approaches a value slightly less than 1 (0.92) deep inside the slab. As complete desolvation can only be achieved when buried in an infinite slab[70], the lower value is due to cutting off the numerical integral at 10 Å (**Figure A.1.4.B**). When straddling the dielectric wall at x=0 Å, the fractional desolvation is close to the envisioned one-half (0.51). Also, the "ms" and "sev" results are identical except for the region between 0 and 4 Å away from the slab, where the solvent-excluded volume between ligand and slab increases the fractional desolvation, as expected.

Next we used the 40 diverse protein targets in the DUD benchmark[47] to examine the overall virtual screening performance of our context-dependent ligand desolvation terms.

25

Figure 1.2 DUD Charge Distribution

Percentage of new and original DUD molecules that have each given net charge. Over 40% of new and original ligands are charged, but only 15% of the original decoys are, while the new decoys closely match the charge distribution of the ligands.
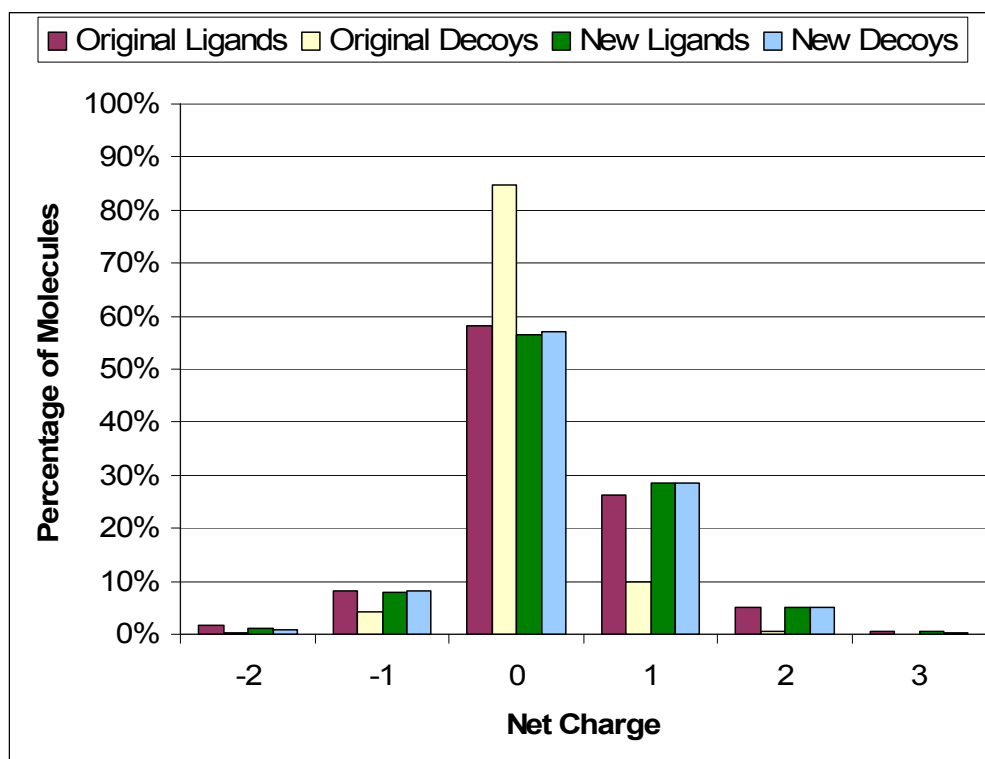
DUD combines known ligands for each of the targets with decoys designed to be a challenging test of the predictive value added by docking. To remove enrichment biases due to simple physical properties, the decoys are property-matched to the ligands; and to ensure they are not actual binders, the decoys are selected to be topologically dissimilar from the ligands. Despite this physical property matching, we found it necessary to eliminate an additional property bias that is critical when looking at ligand desolvation, that of net molecular charge. When net charge was not property-matched, as in the original DUD, the resulting enrichments were artificially inflated for targets with highly charged ligands using no ligand desolvation, and conversely inflated for targets with neutral ligands using full

desolvation. We therefore recalculated the DUD decoys, retaining the same ligands, but adding net-charge as an additional matched property. While only 15% of the original DUD decoys were charged, around 40% of the new decoys are now charged, and the entire charge distribution of the new decoys now matches the ligands closely (**Figure 1.2**). This updated DUD database is available free of charge at http://dud.docking.org.

With new DUD decoys in hand, we can compare the virtual screening performance of the three different approximations to ligand desolvation. After docking, we compute ROC curves for the DUD ligands versus their own decoys, using the adjusted LogAUC to summarize the performance of each curve. LogAUC equally weights the area under the curve in the decade between 0.1% and 1% with the other two decades with upper bounds of 10% and 100%. This results in a metric with the attractive features of AUC, but biased towards early enrichment, and compares favorably to other early enrichment metrics (see Methods).

The results for the no desolvation ("none") and full desolvation ("full") are highly target dependent, whereas the context dependent desolvation shows competitive enrichment across most targets (**Figure 1.3, Table 1.1**). In one set of targets (highlighted blue) where no desolvation outperforms full desolvation, which typically have charged or open binding sites, context-dependent ligand desolvation generally tracks no desolvation. For instance in poly-ADP-ribose polymerase with its very open binding pocket, no desolvation gives 30.5% adjusted LogAUC, outperforming full desolvation at 16.4%, but context-dependent "sev" solvation falls just behind no desolvation at 28.2% adjusted LogAUC. In another set of targets (highlighted green) where full desolvation outperforms no desolvation, which typically have neutral or enclosed binding sites, context-dependent ligand desolvation adapts

to track the better performing full desolvation. For instance in HIV protease with its mostly

enclosed binding pocket, no desolvation performs worst at 5.4% adjusted LogAUC, full

Figure 1.3 DUD Enrichment Comparisons

Comparison of adjusted LogAUC against DUD decoys over the 40 DUD targets. The docking calculations are identical except that "none" contains no ligand desolvation term, "full" uses the full ligand transfer free energy from water to organic solvent as the desolvation term, and "sev" uses context-dependent ligand desolvation while accounting for the local solvent-excluded volume as a ligand atom approaches the receptor.
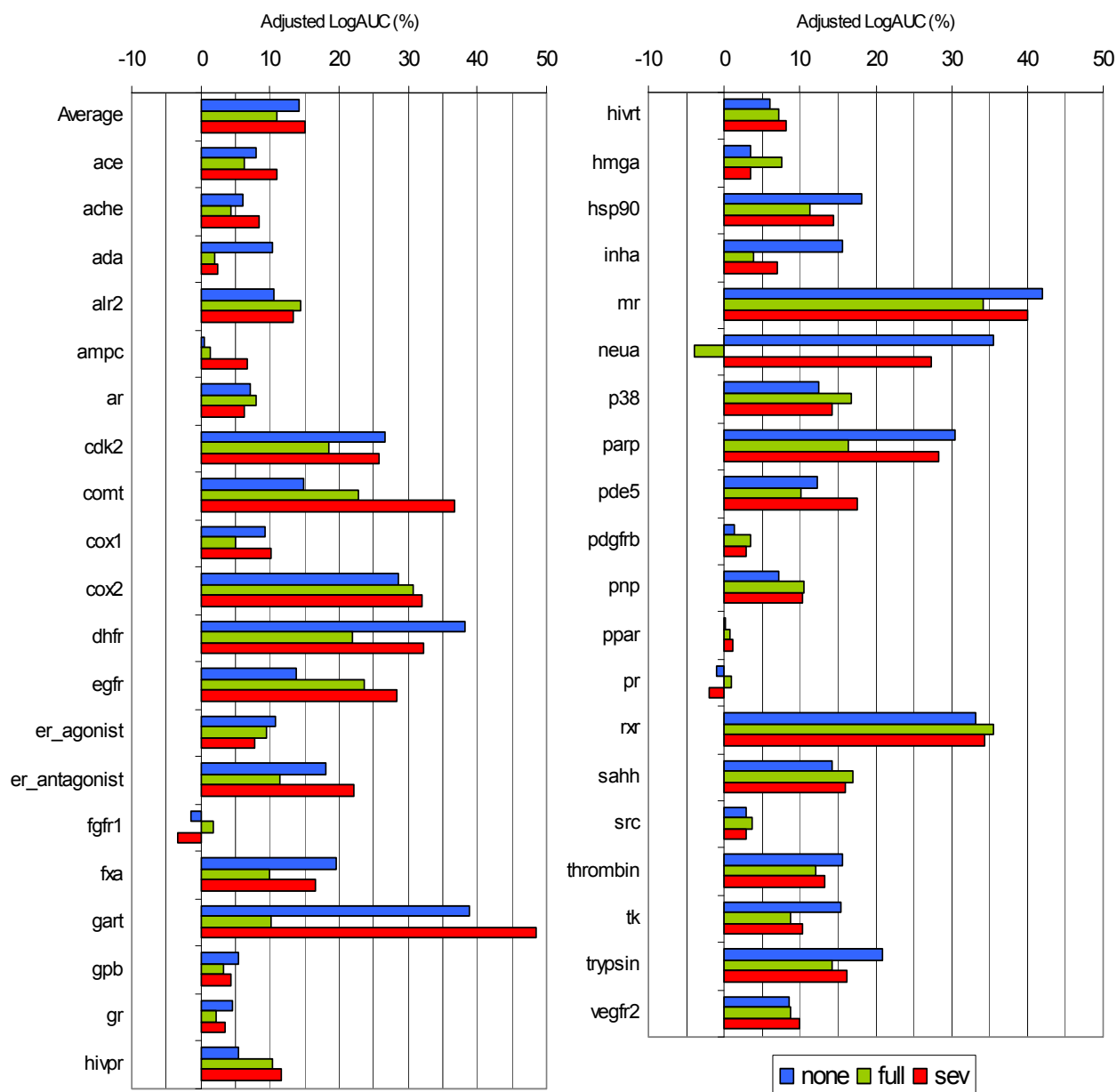
desolvation performs better at 10.3%, while "sev" performs best at 11.6%. While no and full

desolvation show considerable variation in enrichment from target-to-target, often switching

from best to worst, context-dependent desolvation is more consistent across targets.

For some targets both no and full desolvation perform similarly while "sev"

desolvation outperforms both (highlighted red). Other targets are uninformative, either due

to poor performance across the board (highlighted purple), or due to similar performance

regardless of the ligand desolvation method (highlighted black). Averaged over all 40 targets,

full desolvation performs worst with an average of 10.9% adjusted LogAUC while no

desolvation averages a much better 14.3% and the atomic solvent-excluded volume method

performs best with 15.0% adjusted LogAUC. We also include a table with the comparison of

adjusted LogAUC to traditional AUC, plus ROC-based enrichment factors at 1 and 10%

(**Table A.1.1**).

Context-dependent ligand desolvation tracks no desolvation across the entire curve

in those targets where no desolvation performs well (**Figure 1.4**), such as cyclin-dependent

kinase 2 (panel A) and factor Xa (B), and tracks full desolvation across the entire curve in

systems where it performs well, such as epidermal growth factor receptor (C), consistent

with its ability to perform well in different receptor contexts. While context-dependent

desolvation does not always match the best curve, it is usually able to at least interpolate

between no and full ligand desolvation, as demonstrated in enoyl ACP reductase (D). In fact,

context-dependent desolvation is particularly adept at improving early enrichment, as shown

in cycloxygenase-2 (E) and dihydrofolate reductase (F). Semilog ROC plots of all 40 targets

are also included (**Figure A.1.3**). We further analyzed how the ranks of the ligands and

decoys change upon switching from no desolvation to "sev" desolvation (**Figure A.1.5**). For

some targets, the difference between ranking with and without desolvation correlates highly

Table 1.1 Enrichments Against Matched Decoys over the 40 DUD Targets

		Ligand Desolvation Type		
Adjusted LogAUCs	**Abbreviation**	**none**	**full**	**sev**
Target	*Average*	*14.3*	*10.9*	*15.0*
angiotensin-converting enzyme	ace[c]	7.9	6.4	11.1
acetylcholinesterase	ache[c]	6.1	4.3	8.5
adenosine deaminase	ada[e]	10.4	2.0	2.3
aldose reductase	alr2[b]	10.7	14.4	13.4
AmpC beta-lactamase	ampc[e]	0.5	1.3	6.7
androgen receptor	ar[d]	7.1	7.9	6.3
cyclin-dependent kinase 2	cdk2[a]	26.7	18.5	25.7
catechol O-methyltransferase	comt[b]	14.9	22.8	36.8
cycloxygenase-1	cox1[c]	9.2	5.0	10.1
cycloxygenase-2	cox2[d]	28.6	30.6	32.1
dihydrofolate reductase	dhfr[a]	38.3	21.9	32.1
epidermal growth factor receptor	egfr[b]	13.8	23.7	28.3
estrogen receptor agonists	er_agonist[d]	10.8	9.4	7.7
estrogen receptor antagonists	er_antagonist[a]	18.0	11.5	22.1
fibroblast growth factor receptor kinase	fgfr1[e]	-1.5	1.8	-3.4
factor Xa	fxa[a]	19.6	9.9	16.5
glycinamide ribonucleotide transformylase	gart[a]	38.8	10.1	48.6
glycogen phosphorylase β	gpb[e]	5.4	3.3	4.3
glucocorticoid receptor	gr[e]	4.6	2.2	3.5
HIV protease	hivpr[b]	5.4	10.3	11.6
HIV reverse transcriptase	hivrt[d]	6.0	7.1	8.1
hydroxymethylglutaryl-CoA reductase	hmga[e]	3.5	7.5	3.5
heat shock protein 90	hsp90[a]	18.1	11.4	14.5
enoyl ACP reductase	inha[a]	15.7	3.9	7.1
mineralocorticoid receptor	mr[a]	42.0	34.1	40.0
neuraminidase	neua[a]	35.5	-4.0	27.3
p38 mitogen activated protein	p38[b]	12.6	16.7	14.2
poly-ADP-ribose polymerase	parp[a]	30.5	16.4	28.2
phosphodiesterase 5	pde5[c]	12.4	10.2	17.6
platelet derived growth factor receptor kinase	pdgfrb[e]	1.4	3.6	3.0
purine nucleoside phosphorylase	pnp[b]	7.1	10.5	10.3
peroxisome proliferator activated receptor γ	ppar[e]	0.2	0.8	1.2
progesterone receptor	pr[e]	-0.9	0.9	-2.0
retinoic X receptor α	rxr[d]	33.2	35.6	34.3
S-adenosyl-homocysteine hydrolase	sahh[d]	14.2	17.0	16.1
tyrosine kinase SRC	src[e]	2.9	3.6	2.9
thrombin	thrombin[d]	15.7	12.1	13.2
thymidine kinase	tk[a]	15.4	8.7	10.2
trypsin	trypsin[a]	20.9	14.2	16.2
vascular endothelial growth factor receptor	vegfr2[d]	8.6	8.9	9.8

[a]No desolvation performs much better than full. [b]Full desolvation performs better than none. [c]No and full desolvation are similar but "sev" outperforms both. [d]All desolvation models perform similarly. [e]All desolvation models perform poorly.
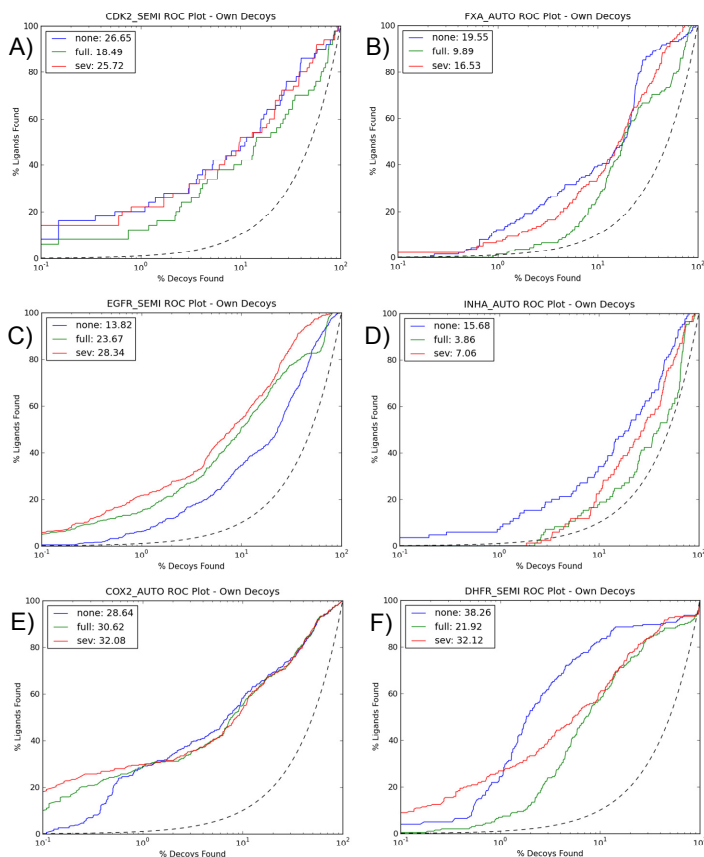
Figure 1.4 Selected DUD Enrichment Plots

Individual semilog ROC plots highlight the adaptability of "sev" context-dependent desolvation, which properly tracks no desolvation in both panels A) cdk2 and B) fxa and properly tracks full desolvation in panel C) egfr. In the cases where it does not track the best curve, it typically takes intermediate values as shown in panel D) inha. Early enrichment is often improved the most, as shown in panels E) cox2 and F) dhfr.

for both ligands and decoys, while for other targets there is so much scatter as to defy generalization. However, where there is scatter in ligand ranks this is mirrored by scatter in the decoy ranks.

Looking at the screening performance against property-matched decoys removes background database biases, allowing us to truly compare ligand desolvation methods. However, it fails to capture how they respond to extreme molecules, of the sort one might find when running large database screens. Ligand desolvation is particularly sensitive to differences in net charge, so we constructed a set of charge outliers having net charges from -4 to -2 and from +2 to +4, choosing 500 molecules of each type randomly from the ZINC database[48]. We docked the charge outliers to the 40 DUD targets, computed ROC curves against the charge outliers, and summarized them by their adjusted LogAUC (**Figure 1.5**).

Figure 1.5 Enrichment versus Random Charged Outliers

Enrichment of the DUD ligands compared to a background of highly charged molecules selected randomly from ZINC. The smaller the ligand desolvation penalty, "full">"sev">"none", the less we discriminate against the random charge outliers and thus get worse enrichments.
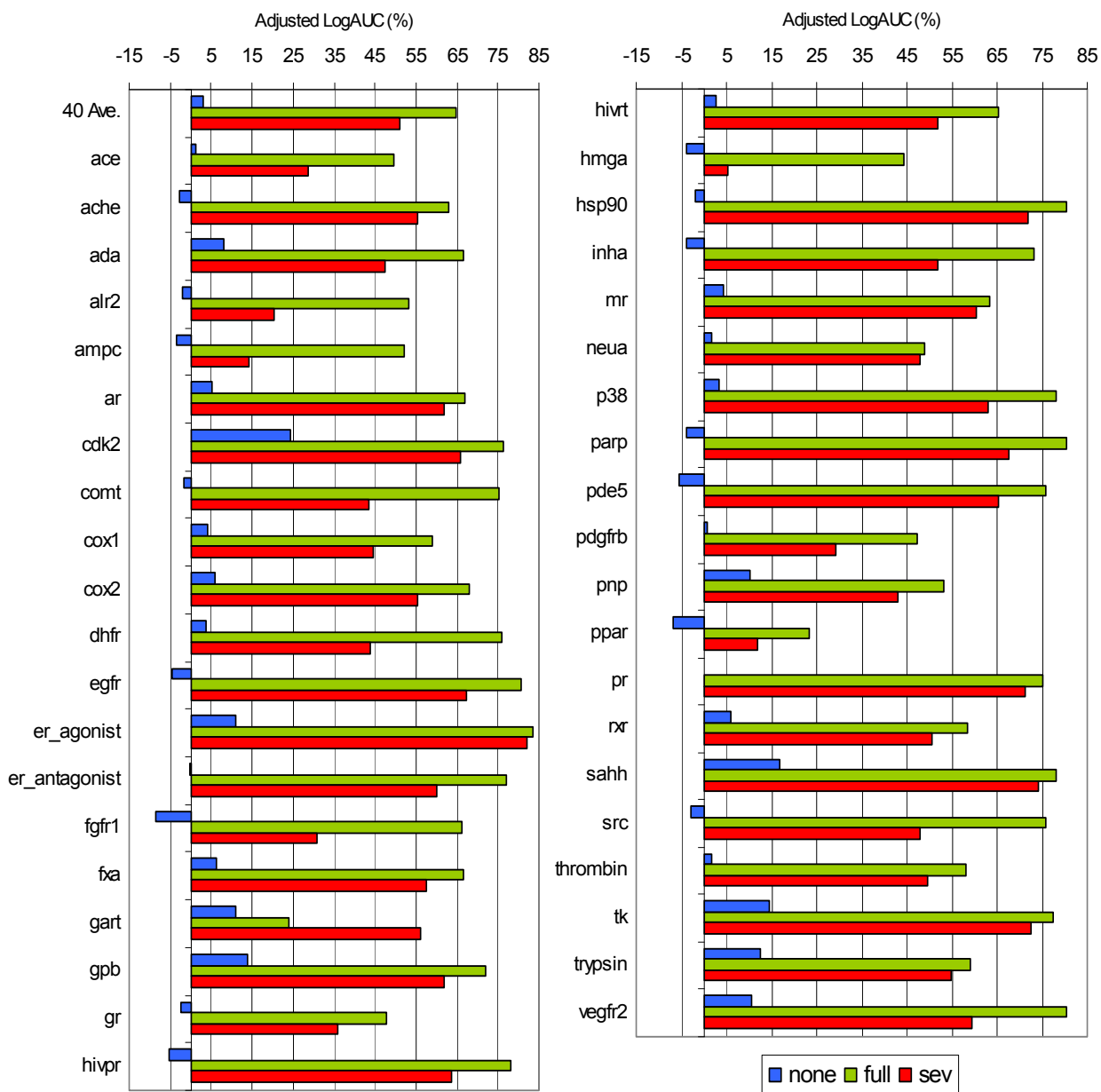
In sharp contrast to the performance against DUD decoys, where no desolvation performed competitively for many targets, its performance against charge outliers is very poor. Unfettered by a desolvation term, no desolvation often enriches the highly charged molecules more than the annotated ligands because the electrostatic interaction energy grows stronger as the ligands become more highly charged. By the same token, full desolvation performs well against charged outliers, much better than it does against DUD decoys, because it over-weights the ligand desolvation penalty. Context-dependent desolvation significantly enriches true ligands over these highly charged outliers, often tracking the superior discriminatory power of full desolvation, without over-weighting the ligand desolvation penalty.

To examine how well the ligand desolvation approximations affect docked ligand pose, we used the 114 crystal-structures in the DOCK5 benchmark. While we were able to begin docking with all 114 structures, one target (1HSL) consistently failed during sphere matching and was discarded. The best scoring pose was below the usual threshold of 2 Å RMSD in 61 (none), 61 (full), and 54 (sev) of the remaining targets. In contrast, the gap from full desolvation to "sev" in number of successful predictions at thresholds of 1.5 Å and 2.5 Å is 1 and 4 targets instead of 7. In most targets the poses are visually similar regardless of the ligand desolvation approximation used (**Figure 1.6**). In the protein tyrosine phosphatase 1B (PTP1B) structure 1PTV (panel A), all three desolvation approximations lead to docked poses that closely match the crystallographic pose. However, the phosphate moiety remains singly protonated with "sev" and "full" desolvation. In another PTB1B structure 1C83 (B) and the acetyl cholinesterase structure 1ACJ (C), "sev" desolvation torques polar groups to more exposed positions, causing the crystallographic agreement to improve. In a few cases such as 1PDZ (D), "sev" desolvation spins the molecule out of alignment with the crystal
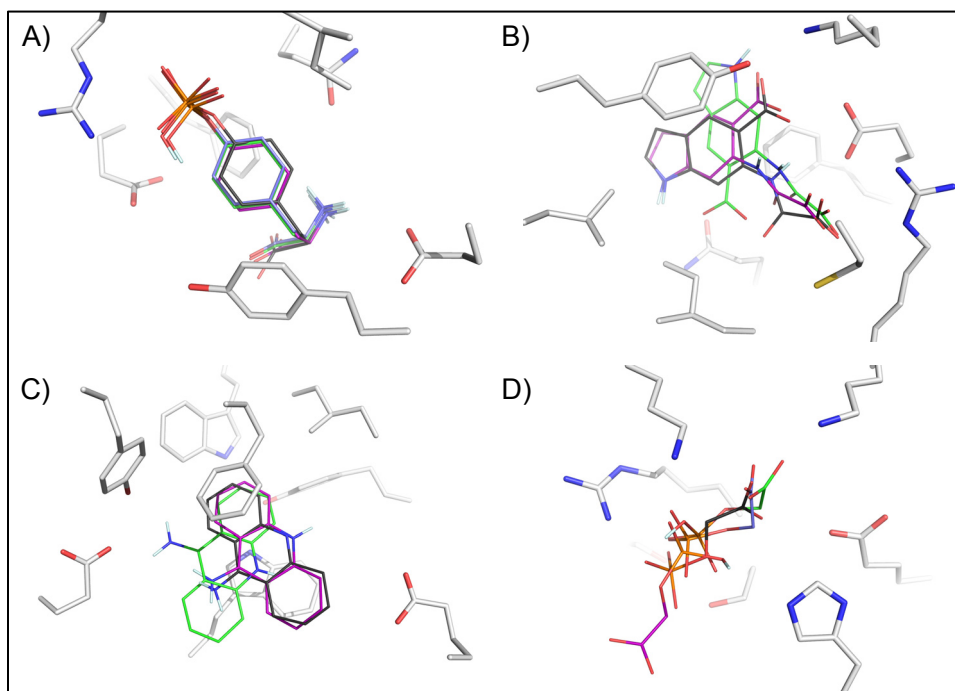
Figure 1.6

Pose Fidelity

Over most targets all three desolvation methods give similar poses, as in A) 1PTV. Occasionally context-dependent "sev" desolvation improved poses, as in B) 1C83 and C) 1ACJ. For a few cases "sev" desolvation reduces pose fidelity, as in D) 1PDZ. The carbon atoms are colored as follows: receptor (white), crystallographic ligand (dark gray), no desolvation (blue), full desolvation (green), "sev" desolvation (purple).

pose. Overall, pose fidelity with context-dependent ligand desolvation is similar to full or no desolvation.

Though efforts are made during DUD decoy generation to ensure they are topologically dissimilar from the ligands, there is no guarantee that the decoys do not, in fact, bind. To examine the screening performance against validated decoys, we employed two artificial binding cavities, a small hydrophobic pocket created in T4 lysozyme by the Leu99→Ala substitution (L99A), and a charged binding pocket created in cytochrome C peroxidase by the Trp191→Gly substitution (W191G). In the enclosed hydrophobic pocket of L99A, we would expect full desolvation to be a good approximation (**Figure 1.7.A**), and indeed it gives the best LogAUC at 18.6 %, with "sev" tracking it at 15.5%, and no desolvation at -1.5%. In the enclosed but charged pocket of W191G (**Figure 1.7.B**), "sev" and full desolvation tie for the best LogAUC at 48.8%, with no desolvation lagging at 8.1%

adjusted LogAUC. Compared to the high enrichments we typically see against larger databases[30, 49], overall enrichment against these experimental decoys is poor; but note that these are especially challenging decoys since most of them were experimentally tested due to strong scores in earlier docking screens.

Figure 1.7 ROC for Experimental Decoys in Model Binding Sites

ROC curves of known ligands versus experimental decoys for A) the small hydrophopic cavity of L99A T4 lysozyme and B) the negatively charged cavity in W191G cyctochrome C peroxidase.
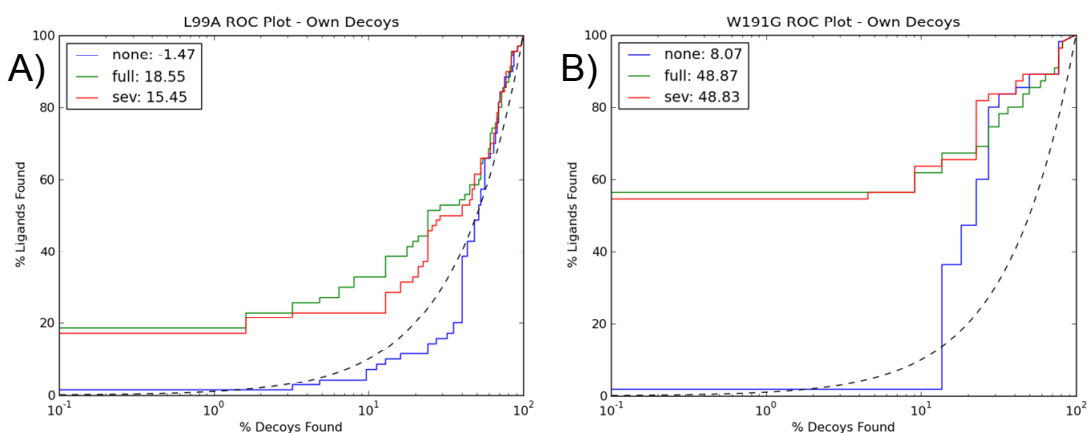
1.5 Discussion

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 so[1-4] 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 imagine that the new model may perform *worse* than the old by upsetting this prior cancellation of errors. Thus it is comforting that in moving to a more realistic model of ligand desolvation the results against the DUD benchmark do not deteriorate. More convincingly, context-dependent desolvation confers improved stability to target-by-target and ligand-by-ligand variability. It is almost an aphorism in the field that different docking methods are better

suited to different targets, and this behavior may be seen when comparing no and full desolvation (**Figure 1.3**), where no desolvation often better enriches targets with charged ligands and more open binding sites such as DHFR, GART, and PARP, while full desolvation often better enriches targets with neutral ligands and more enclosed binding sites such as EGFR, HIVPR, and P38. In striking contrast, context-dependent desolvation, though still varying in performance from target-to-target, always gives either the best or second-best enrichments and does not swing wildly from best to worst. Even though no desolvation outperforms full desolvation against matched decoys, against charged outliers no desolvation performs only marginally better than random. Conversely, full desolvation has extraordinarily high enrichments against charged outliers, while context-dependent desolvation approaches full desolvation in its ability to discriminate against them (**Figure 1.5**). This stability to variations in ligand preferences and binding site contexts reflects the better physical model in the context-dependent ligand desolvation penalties.

Even with this improved stability, it would be comforting if "sev" desolvation gave substantially better enrichment than ignoring desolvation entirely; currently its overall improvement is marginally significant but certainly not substantial. Inspection of particular cases suggest that desolvation leads to worse enrichments largely in targets that bind more positive ligands. This made us wonder about the charge distribution in these ligands, which had significant charge localized on the hydrogen atoms. We therefore explored using more reasonable radii for hydrogen and heavy atoms in our desolvation calculations, moving from treating all atoms as having a radius of 1.4 Å, to modeling hydrogens as size 1.0 and all heavy atoms as 1.8 Å. Encouragingly, enrichment improved substantially for the full DUD set, especially against targets recognizing positively charged ligands (**Figure A.1.2**). The largest improvements occurred, for instance, in DHFR, INHA, and NEUA, whose ligands all have

key positively charged moieties. In overall enrichment, the adjusted LogAUC gap between no desolvation, at 14.3%, and "sev" desolvation, at 15.0%, widens from 0.7 to 1.5% in the process of giving the hydrogen atoms their own desolvation grid, resulting in an overall average enrichment of 15.8% LogAUC. While this still leaves us with an overall performance improvement that remains small, it is comforting that as we move to better physical models the performance consistently improves without any fitting whatsoever. This need not be the case—we rely so heavily on cancellation of errors in docking, that one might easily imagine models that are physically more correct but that actually reduce performance by our standard enrichment metrics. Further improvement of our physical models, for instance by including receptor terms in desolvation and internal energy terms in the ligand, thus seems not only scholarly but pragmatic.

Three results in this study may interest the specialist. First, we introduce a virtual screening metric, LogAUC, which tackles the "early enrichment" problem by computing the percentage of the ideal area that lies under the semilog enrichment curve. LogAUC shares many desirable characteristics with ROC AUC: as it is easily interpreted, robust, and independent of similar extensive variables[57]. Second, we introduce the atomic solvent-excluded volume (sev) method to deal with the often neglected low-dielectric region that forms in the gaps between ligand and protein (**Figure 1.1.B**). This still allows pre-calculation of fractional desolvation while improving the desolvation penalty magnitude at no computational cost during docking, with only a moderate (~30 minute) pre-calculation cost per protein. Third, we provide a revised version of the DUD database which corrects some problems with the initial release, such as the inclusion of "decoys" that were in fact ligands, and adds a sixth physical property term against which to balance ligands and decoys — net charge. We also provide the charged outlier sets, which can be challenging for any method

that does not consider desolvation, explicitly or implicitly, and are meant to mimic molecules one will encounter in a large unbiased library screen, such as ZINC. The updated version of DUD is freely available at http://dud.docking.org.

Certain caveats merit airing. We multiply the fractional desolvation by the entire transfer free energy, and this is formally correct only for the self-energy terms of polar desolvation. Simple multiplication is just an approximation for both the Coulombic term and non-polar desolvation. We additionally ignore the fractional desolvation due to the other ligand atoms. Exact treatment of these desolvation terms is possible, but as it would add substantially to the on-the-fly cost, it may be better left to subsequent re-scoring. Along the same lines, we use atomic desolvation energies derived from a single ligand conformation to represent all conformations docked. Although this is reasonable for many molecules, there will be others where the solvation energy changes substantially from conformation-to-conformation. Whereas the calculation and storage costs of atomic solvation energies of every conformation of every ligand seemed daunting enough to ignore for this study, this term can be pre-calculated and so may merit further investigation. Finally, it should be clear that we only consider ligand desolvation here, and not its logical complement protein desolvation. Doing so will undoubtedly affect docking performance substantially, but this is a term that may be added without affecting the calculated ligand solvation term.

Notwithstanding these caveats, an important result of this work is the incorporation of a physically reasonable model of ligand desolvation with negligible cost to docking run time. Thus, the total time to screen 134,000 flexible molecules of the new DUD set against all 40 targets was 2446 cpu hours without correcting for desolvation, while the total time to prosecute this screen with "sev" desolvation was 3862 cpu hours. The method thus incorporates critical aspects of higher-level solvation theory, while maintaining the speed

advantages of docking. This is achieved by pre-calculating most terms before docking, and as such this method should be applicable to any physics-based docking approach.

1.6 Acknowledgements

I. Supporting Information

In Table A.1.1, we compare adjusted LogAUC to traditional AUC. In Figure A.1.2, we examine the effect of using separate desolvation grids for hydrogen and heavy atoms. In Figure A.1.3, we give all 40 ROC plots against property-matched DUD decoys. In Figure A.1.4, we show the effect of changing the grid spacing and integration cutoff. In Figure A.1.5, we compare the ligand and decoys ranks as "sev" desolvation is turned off or on. In Table A.1.6, we estimate the error in the self-energy term due to neglecting the ligand component of fractional desolvation. This information is included in Appendix A.1.

1.7 References

1. Jorgensen, WL. "The Many Roles of Computation in Drug Discovery" *Science* **2004** 303: 1813-1818.
2. Alvarez, JC. "High-Throughput Docking as a Source of Novel Drug Leads" *Curr Opin Chem Biol* **2004** 8: 365-370.
3. Shoichet, BK; McGovern, SL; Wei, B; Irwin, JJ. "Lead Discovery Using Molecular Docking" *Curr Opin Chem Biol* **2002** 6: 439-446.
4. Cavasotto, CN; Orry, AJ. "Ligand Docking and Structure-Based Virtual Screening in Drug Discovery" *Curr Top Med Chem* **2007** 7: 1006-1014.
5. Eldridge, MD; Murray, CW; Auton, TR; Paolini, GV; Mee, RP. "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* **1997** 11: 425-445.

6. Jones, G; Willett, P; Glen, RC; Leach, AR; Taylor, R. "Development and Validation of a Genetic Algorithm for Flexible Docking" *J Mol Biol* **1997** 267: 727-748.

7. Rarey, M; Kramer, B; Lengauer, T; Klebe, G. "A Fast Flexible Docking Method Using an Incremental Construction Algorithm" *J Mol Biol* **1996** 261: 470-489.

8. Friesner, RA; Banks, JL; Murphy, RB; Halgren, TA; Klicic, JJ; Mainz, DT; Repasky, MP; Knoll, EH; Shelley, M; Perry, JK; Shaw, DE; Francis, P; Shenkin, PS. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy" *J Med Chem* **2004** 47: 1739-1749.

9. Jain, AN. "Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine" *J Med Chem* **2003** 46: 499-511.

10. Muegge, I; Martin, YC. "A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach" *J Med Chem* **1999** 42: 791-804.

11. Gohlke, H; Hendlich, M; Klebe, G. "Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions" *J Mol Biol* **2000** 295: 337-356.

12. Velec, HF; Gohlke, H; Klebe, G. "Drugscore(Csd)-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of near-Native Ligand Poses and Better Affinity Prediction" *J Med Chem* **2005** 48: 6296-6303.

13. Meng, EC; Shoichet, BK; Kuntz, ID. "Automated Docking with Grid-Based Energy Evaluation" *J Comput Chem* **1992** 13: 505-524.

14. Morris, GM; Goodsell, DS; Halliday, RS; Huey, R; Hart, WE; Belew, RK; Olson, AJ. "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function" *J Comput Chem* **1998** 19: 1639-1662.

15. Abagyan, R; Totrov, M; Kuznetsov, D. "Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation" *J Comput Chem* **1994** 15: 488-506.

16. McMartin, C; Bohacek, RS. "Qxp: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design" *J Comput-Aided Mol Des* **1997** 11: 333-344.

17. Dominguez, C; Boelens, R; Bonvin, AM. "Haddock: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information" *J Am Chem Soc* **2003** 125: 1731-1737.

18. Moitessier, N; Englebienne, P; Lee, D; Lawandi, J; Corbeil, CR. "Towards the Development of Universal, Fast and Highly Accurate Docking/Scoring Methods: A Long Way to Go" *Br J Pharmacol* **2008** 153: S7-S26.

19. Brooijmans, N; Kuntz, ID. "Molecular Recognition and Docking Algorithms" *Annu Rev Biophys Biomol Struct* **2003** 32: 335-373.

20. Shoichet, BK. "No Free Energy Lunch" *Nat Biotechnol* **2007** 25: 1109-1110.

21. Chang, CE; Gilson, MK. "Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm" *J Am Chem Soc* **2004** 126: 13156-13164.

22. Chen, W; Chang, CE; Gilson, MK. "Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design" *Biophys J* **2004** 87: 3035-3049.

23. Sitkoff, D; Sharp, KA; Honig, B. "Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models" *J Phys Chem* **1994** 98: 1978-1988.

24. Guimaraes, CR; Cardozo, M. "Mm-Gb/Sa Rescoring of Docking Poses in Structure-Based Lead Optimization" *J Chem Inf Model* **2008** 48: 958-970.

25. Kalyanaraman, C; Bernacki, K; Jacobson, MP. "Virtual Screening against Highly Charged Active Sites: Identifying Substrates of Alpha-Beta Barrel Enzymes" *Biochemistry* **2005** 44: 2059-2071.

26. Zou, XQ; Sun, YX; Kuntz, ID. "Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model" *J Am Chem Soc* **1999** 121: 8033-8043.

27. Liu, HY; Kuntz, ID; Zou, XQ. "Pairwise Gb/Sa Scoring Function for Structure-Based Drug Design" *J Phys Chem B* **2004** 108: 5453-5462.

28. Wang, J; Kang, X; Kuntz, ID; Kollman, PA. "Hierarchical Database Screenings for Hiv-1 Reverse Transcriptase Using a Pharmacophore Model, Rigid Docking, Solvation Docking, and Mm-Pb/Sa" *J Med Chem* **2005** 48: 2432-2444.

29. Grant, JA; Pickup, BT; Nicholls, A. "A Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods" *J Comput Chem* **2001** 22: 608-640.

30. Graves, AP; Shivakumar, DM; Boyce, SE; Jacobson, MP; Case, DA; Shoichet, BK. "Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing" *J Mol Biol* **2008** 377: 914-934.

31. Shoichet, BK; Leach, AR; Kuntz, ID. "Ligand Solvation in Molecular Docking" *Proteins* **1999** 34: 4-16.

32. Wesson, L; Eisenberg, D. "Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution" *Protein Sci* **1992** 1: 227-235.

33. van Dijk, AD; Bonvin, AM. "Solvated Docking: Introducing Water into the Modelling of Biomolecular Complexes" *Bioinformatics* **2006** 22: 2340-2347.

34. Fernandez-Recio, J; Totrov, M; Abagyan, R. "Icm-Disco Docking by Global Energy Optimization with Fully Flexible Side-Chains" *Proteins* **2003** 52: 113-117.

35. Stouten, PFW; Frommel, C; Nakamura, H; Sander, C. "An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations" *Mol Simul* **1993** 10: 97-&.

36. Morris, GM; Huey, R; Lindstrom, W; Sanner, MF; Belew, RK; Goodsell, DS; Olson, AJ. "Autodock4 and Autodocktools4: Automated Docking with Selective Receptor Flexibility" *J Comput Chem* **2009** 30: 2785-2791.

37. Cerutti, DS; Jain, T; McCammon, JA. "Cirse: A Solvation Energy Estimator Compatible with Flexible Protein Docking and Design Applications" *Protein Sci* **2006** 15: 1579-1596.

38. Schnecke, V; Kuhn, LA. "Virtual Screening with Solvation and Ligand-Induced Complementarity" *Perspect Drug Discov* **2000** 20: 171-190.

39. Verdonk, ML; Chessari, G; Cole, JC; Hartshorn, MJ; Murray, CW; Nissink, JW; Taylor, RD; Taylor, R. "Modeling Water Molecules in Protein-Ligand Docking Using Gold" *J Med Chem* **2005** 48: 6504-6515.

40. Friesner, RA; Murphy, RB; Repasky, MP; Frye, LL; Greenwood, JR; Halgren, TA; Sanschagrin, PC; Mainz, DT. "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes" *J Med Chem* **2006** 49: 6177-6196.

41. Morreale, A; Gil-Redondo, R; Ortiz, AR. "A New Implicit Solvent Model for Protein-Ligand Docking" *Proteins* **2007** 67: 606-616.

42. Majeux, N; Scarsi, M; Caflisch, A. "Efficient Electrostatic Solvation Model for Protein-Fragment Docking" *Proteins* **2001** 42: 256-268.

43. Wei, BQ; Baase, WA; Weaver, LH; Matthews, BW; Shoichet, BK. "A Model Binding Site for Testing Scoring Functions in Molecular Docking" *J Mol Biol* **2002** 322: 339-355.

44. Li, JB; Zhu, TH; Hawkins, GD; Winget, P; Liotard, DA; Cramer, CJ; Truhlar, DG. "Extension of the Platform of Applicability of the Sm5.42r Universal Solvation Model" *Theor Chem Acc* **1999** 103: 9-63.

45. Bashford, D; Case, DA. "Generalized Born Models of Macromolecular Solvation Effects" *Annu Rev Phys Chem* **2000** 51: 129-152.

46. Scarsi, M; Apostolakis, J; Caflisch, A. "Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions" *J Phys Chem A* **1997** 101: 8098-8106.

47. Huang, N; Shoichet, BK; Irwin, JJ. "Benchmarking Sets for Molecular Docking" *J Med Chem* **2006** 49: 6789-6801.

48. Irwin, JJ; Shoichet, BK. "Zinc--a Free Database of Commercially Available Compounds for Virtual Screening" *J Chem Inf Model* **2005** 45: 177-182.

49. Brenk, R; Vetter, SW; Boyce, SE; Goodin, DB; Shoichet, BK. "Probing Molecular Docking in a Charged Model Binding Site" *J Mol Biol* **2006** 357: 1449-1470.

50. Gilson, MK; Sharp, K; Honig, B. "Calculating Electrostatic Interactions in Biomolecules: Method and Error Assessment" *J. Comput. Chem* **1987** 9: 327-335.

51. Lorber, DM; Shoichet, BK. "Hierarchical Docking of Databases of Multiple Ligand Conformations" *Curr Top Med Chem* **2005** 5: 739-749.

52. Weiner, SJ; Kollman, PA; Case, DA; Singh, UC; Ghio, C; Alagona, G; Profeta, S; Weiner, P. "A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins" *J Am Chem Soc* **1984** 106: 765-784.

53. *Oechem Tk*, 1.7.2.4; Openeye Scientific Software: Santa Fe, NM, 2009.

54. Sadowski, J; Gasteiger, J; Klebe, G. "Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures" *J Chem Inf Comput Sci* **1994** 34: 1000-1008.

55. *Ligprep*, 23110; Schrödinger: New York, NY, 2009.

56. *Omega*, 2.3.2; Openeye Scientific Software: Santa Fe, NM, 2008.

57. Nicholls, A. "What Do We Know and When Do We Know It?" *J Comput-Aided Mol Des* **2008** 22: 239-255.

58. Clark, RD; Webster-Clark, DJ. "Managing Bias in Roc Curves" *J Comput-Aided Mol Des* **2008** 22: 141-146.

59. Truchon, JF; Bayly, CI. "Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem" *J Chem Inf Model* **2007** 47: 488-508.

60. Katritch, V; Rueda, M; Lam, PC; Yeager, M; Abagyan, R. "Gpcr 3d Homology Models for Ligand Screening: Lessons Learned from Blind Predictions of Adenosine A2a Receptor Complex" *Proteins* **2010** 78: 197-211.

61. Moustakas, DT; Lang, PT; Pegg, S; Pettersen, E; Kuntz, ID; Brooijmans, N; Rizzo, RC. "Development and Validation of a Modular, Extensible Docking Program: Dock 5" *J Comput-Aided Mol Des* **2006** 20: 601-619.

62. Irwin, JJ; Shoichet, BK; Mysinger, MM; Huang, N; Colizzi, F; Wassam, P; Cao, Y. "Automated Docking Screens: A Feasibility Study" *J Med Chem* **2009** 52: 5712-5720.

63. Shelley, JC; Cholleti, A; Frye, LL; Greenwood, JR; Timlin, MR; Uchimaya, M. "Epik: A Software Program for Pk(a) Prediction and Protonation State Generation for Drug-Like Molecules" *J Comput-Aided Mol Des* **2007** 21: 681-691.

64. Eriksson, AE; Baase, WA; Wozniak, JA; Matthews, BW. "A Cavity-Containing Mutant of T4 Lysozyme Is Stabilized by Buried Benzene" *Nature* **1992** 355: 371-373.

65. Morton, A; Baase, WA; Matthews, BW. "Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme" *Biochemistry* **1995** 34: 8564-8575.

66. Su, AI; Lorber, DM; Weston, GS; Baase, WA; Matthews, BW; Shoichet, BK. "Docking Molecules by Families to Increase the Diversity of Hits in Database Screens: Computational Strategy and Experimental Evaluation" *Proteins* **2001** 42: 279-293.

67. Graves, AP; Brenk, R; Shoichet, BK. "Decoys for Docking" *J Med Chem* **2005** 48: 3714-3728.

68. Fitzgerald, MM; Musah, RA; McRee, DE; Goodin, DB. "A Ligand-Gated, Hinged Loop Rearrangement Opens a Channel to a Buried Artificial Protein Cavity" *Nat Struct Biol* **1996** 3: 626-631.

69. Musah, RA; Jensen, GM; Bunte, SW; Rosenfeld, RJ; Goodin, DB. "Artificial Protein Cavities as Specific Ligand-Binding Templates: Characterization of an Engineered Heterocyclic Cation-Binding Site That Preserves the Evolved Specificity of the Parent Protein" *J Mol Biol* **2002** 315: 845-857.

70. Gilson, MK; Honig, B. "The Inclusion of Electrostatic Hydration Energies in Molecular Mechanics Calculations" *J Comput-Aided Mol Des* **1991** 5: 5-20.

Gloss to Chapter 2

Why did we create our context-dependent ligand desolvation method, if not to use it to find new ligands? In this chapter, we put it to the ultimate test, to find new inhibitors for the highly charged and solvent-exposed binding site of the chemokine receptor CXCR4. The natural chemokine ligand of CXCR4 is itself a protein; making the binding site we are trying to inhibit a notoriously difficult protein-protein interface[1]. Despite the challenge, by using our context-dependent ligand desolvation method and a new thin dielectric boundary in the electrostatics calculations, we achieved a great 17% hit rate in our prospective virtual screen against a CXCR4 crystal structure.

CXCR4 belongs to a family of proteins, called G-protein coupled receptors (GPCRs), that are the protein targets of at least 30% of all drugs[2]. However, protein structures for this family are still rare and difficult to obtain. If we could use the few structures available to predict the structures of the rest of the family, then we could potentially use molecular docking to find new inhibitors for many of these important targets. For this study only four GPCR template structures were available, and all were very far in sequence homology from CXCR4. Nevertheless, we wanted to generate a homology model before crystal structure release, and use it to virtually screen for new CXCR4 inhibitors. Then, after crystal structure release, we would repeat the virtual screen and compare the prospective ligand discovery rates.

This was a challenging project, not only because the homology modeling was difficult, but because docking to find inhibitors for a protein-protein interface was itself

uncertain. Indeed, without the modern ligand desolvation corrections I introduced, our lab found zero lead-like hits out of 37 attempts in CTX-M β-lactamase[3], a charged and open cavity reminiscent of CXCR4. On at least three occasions Brian almost terminated the project, even calling for a vote at group meeting to determine the projects fate. But perseverance paid off, as we confirmed 4 hits out of 23 molecules (17%) predicted from the crystal structure, while we confirmed only 1 hit out of 24 molecules (4%) predicted from the homology model. In a similar experiment involving dual virtual screens in dopamine D3 receptor[4], a GPCR much closer to template structures, the homology model performed just as well as the crystal structure for prospective ligand discovery. Contrasting these two targets and four campaigns illuminates the limits of homology models, and thus the areas of the GPCR landscape that may be amenable to structure-based ligand discovery.

Most exciting was that several of the novel scaffolds were potent and relatively small, with IC_{50} values as low as 306 nM, ligand efficiencies as high as 0.36, and substantial efficacy in blocking cellular chemotaxis. The potency and efficiency of these molecules has few precedents among protein-protein interface inhibitors, and supports structure-based efforts to discover new leads for chemokine GPCRs.

I. References

1. Wells, JA; McClendon, CL. "Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces" Nature 2007 450: 1001-1009.

2. Overington, JP; Al-Lazikani, B; Hopkins, AL. "How Many Drug Targets Are There?" Nat Rev Drug Discov 2006 5: 993-996.

3. Chen, Y; Shoichet, BK. "Molecular Docking and Ligand Specificity in Fragment-Based Inhibitor Discovery" Nat Chem Bio 2009 5: 358-364.

4. Carlsson, J; Coleman, RG; Setola, V; Irwin, JJ; Fan, H; Schlessinger, A; Sali, A; Roth, BL; Shoichet, BK. "Ligand Discovery from a Dopamine D3 Receptor Homology Model and Crystal Structure" Nat Chem Biol 2011 7: 769-778.

Chapter 2:

Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4

Michael M. Mysinger[1†], Dahlia R. Weiss[1†], Joshua J. Ziarek[2†], Stéphanie Gravel[3,4], Allison K. Doak[1], Joel Karpiak[1], Nikolaus Heveker[3,4], Brian K. Shoichet[1]*, Brian F. Volkman[2]*

1   Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158-2550

2   Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI 53226

3   Centre de Recherche, CHU Sainte-Justine, Montréal (QC), Canada H3T-1C5

†Co-first authors

*Corresponding authors (BFV & BKS)

2.1 Abstract

G-protein coupled receptors (GPCRs) are key signaling molecules and are intensely studied. Whereas small-molecule recognizing GPCRs have been successfully targeted for drug discovery, protein-recognizing GPCRs, such as the chemokine receptors, have few drugs or even good small molecule reagents. This reflects both the difficulties that attend protein-protein interface inhibitor discovery, and the lack of structures for these targets. Imminent structure determination of chemokine receptor CXCR4 motivated docking screens for new ligands against a homology model and subsequently the crystal structure. Over 3 million molecules were docked against the model and then against the crystal structure; 24 and 23 high-scoring compounds from the respective screens were tested experimentally. Docking against the model yielded only one antagonist, which resembled known ligands and lacked specificity, while the crystal structure docking yielded four that were chemically novel and apparently specific. Intriguingly, several of the novel scaffolds were potent and relatively small, with IC_{50} values as low as 306 nM, ligand efficiencies as high as 0.36, and substantial efficacy in blocking cellular chemotaxis. The potency and efficiency of these molecules has few precedents among protein-protein interface inhibitors, and supports structure-based efforts to discover new leads for chemokine GPCRs.

2.2 Introduction

G-protein coupled receptors (GPCRs) play a central role in many normal physiological pathways and altered diseased states, and are the targets of about 30% of marketed drugs[1]. Ligand discovery against small-molecule GPCRs such as the bioamine receptors has been particularly productive, as have structure-based screens against their crystal structures[2-5]. Targeting larger-molecule recognizing GPCRs has been more difficult.

47

Whereas multiple reagents are available for lipid and peptidergic GPCRs, their molecular weights are substantially higher than those typical for bioamine receptors and they are less ligand efficient. This reflects the challenges faced in ligand discovery against peptide-protein and lipid-protein interfaces. These difficulties are even more acute against chemokine GPCRs, which recognize folded proteins around 100 amino-acids in length and are thus protein-protein interface (PPI) targets[6]. Though there are several example drugs in this class, such as maraviroc, plerixafor, and vorapaxar, finding organic molecules with good affinity and the physical properties of oral drugs is notoriously difficult for PPI targets, as reflected in the high molecular weight and hydrophobicity of the few PPI drugs[7].

A public competition to predict ligand complexes with the structure of C-X-C chemokine receptor 4 (CXCR4) inspired us to bring structure-based discovery to bear against this key member of the chemokine family[8]. CXCR4 natively recognizes the CXCL12 chemokine, an 8 kDa protein. Like many other PPI targets, CXCR4 plays a key signaling role: it is constitutively expressed in many organs and is implicated in chemotactic roles as diverse as lymphopoiesis, myelopoeisis, embryogenesis, angiogenesis, cardiogenesis, neuron migration and cerebral development[9, 10]. The receptor is involved in disease states such as myocardial infarction/reperfusion injury[11], WHIM syndrome[12], HIV infection[13], and the growth and development of more than 20 different types of cancer[14]. Despite the intense interest, only a few potent and selective small molecule antagonists have been discovered for CXCR4[15-17]. New ligand chemotypes, which a structure-based approach enables, might provide leads to perturb the critical biology for which CXCR4 is responsible.

Notwithstanding intense effort, experimental structures of GPCRs remain scarce and so homology models are often used for GPCR ligand discovery[18-21]. Such models potentially enable structure-based discovery against many more targets than have been experimentally

determined, but their reliability has never been tested prospectively. From a technical perspective, the ability to compare a discovery campaign against a homology model to one against the subsequently released crystal structure might illuminate model viability in an unbiased and wholly prospective way.

Thus we had two broad questions that we hoped to address in this study. First, can we discover biologically useful ligands for CXCR4 using a structure-based approach? Second, how does a prospective docking screen against a homology model of the receptor compare to that against the crystal structure? The first question reflects the intense biological interest in this target and its problematic status as a PPI, with all the challenges those present for ligand discovery. The second question might inform which parts of the class A GPCR family are good candidates for homology-based drug discovery. A prospective homology-model driven program against the D3 receptor was recently shown to be as effective as one against its crystal structure[22], but transmembrane sequence identity between D3 receptor and its nearest structural template is 42%. CXCR4 has at best 25% transmembrane sequence identity to the nearest template structure. If we can only rely on models with sequence identities as high as D3 receptor, then only about 10% of GPCRs might be modeled given current structural coverage; however, if 25% sequence identity suffices then over 70% are viable for structure-based ligand discovery. Experimental tests of docking hits against both model and crystal structure will potentially discover new ligands for CXCR4, and may also inform the general usefulness of distant GPCR homology models.

2.3 Results

I. Homology Model Construction

The effort began with calculating homology models for CXCR4. The GPCR Dock 2010 Assessment[8] challenged us to predict the orientation of the small molecule IT1t prior to release of the first CXCR4 structures[23]. We followed a strategy (**Figure A.2.1**) that used enrichment of known ligands to guide model selection, as pioneered in earlier studies[19]. Initial homology models were refined for sequence alignment of CXCR4 to the four crystallographic templates then available, β_1 and β_2 adrenergic receptors, adenosine A_{2A} receptor and rhodopsin (**Figure A.2.2**). To expand backbone diversity, we used low-frequency elastic normal modes to perturb template backbones[24]. We calculated 576 and 510 homology models from the crystallographic templates and the perturbed structures, respectively. We docked known ChEMBL04 ligands[25] and property-matched decoys to each model then measured the retrospective enrichment using adjusted LogAUC[26]. Enrichment is a widely used metric in docking, reflecting the ranking of known ligands selected from a database of decoy molecules, compared to what would be expected at random. This can either be expressed as overall enrichment over random, or, as we do here, a log-weighted enrichment to emphasize the highest ranking molecules, which are the most likely to be selected for testing. In the adjusted LogAUC metric, a value of 0% represents completely random selection. To select models, we also used the rank of the co-crystal ligand IT1t (though its bound structure was still unknown), the number of ligands interacting with the critical residue $E^{7.39}$ (Ballesteros-Weinstein numbering[27]), and the complementarity between docked ligands and modeled binding-site. Five top models and their corresponding ligand IT1t binding positions were ultimately submitted to the competition.

Before release of the crystal structure we continued to develop homology models for prospective ligand discovery. In the final iteration we built 2044 homology models without ECL2, docking each to 60 known ligands and 2456 property-matched decoys. Overall, 36 billion ligand orientations and 55 billion conformations were sampled, so over 64 trillion complexes completed within 176 cpu-days or 9 hours of wall clock time on our cluster. We selected one model based on the criteria mentioned above, upon which we generated 1000 ECL2 variants. We selected a single loop model with ligand enrichment of 22% LogAUC for prospective screening.

II. Homology Model Virtual Screen

To predict new CXCR4 ligands from the prospective homology model, we used DOCK 3.6 to virtually screen the lead-like subset of ZINC[28], i.e. molecules with molecular weights below 350, logP less than 3.5, and 7 or less rotatable bonds. Each of the 3.3 million molecules in ZINC was sampled in an average of 11,000 orientations and 2,700 conformations, or 41 trillion complexes sampled overall. Each complex was scored for complementarity based on van der Waals (using a modified AMBER potential function) and electrostatic interaction energies (using potentials calculated with DelPhi), corrected for ligand desolvation[26]. The full screen took 372 cpu-days, or 18 hours of wall clock time on our cluster.
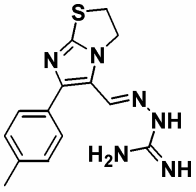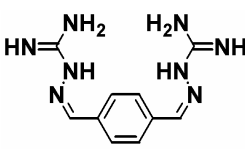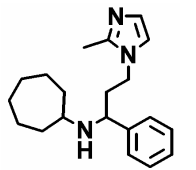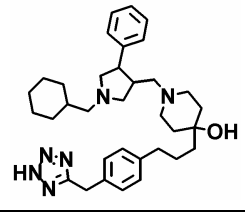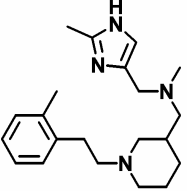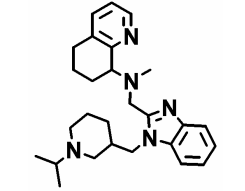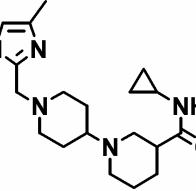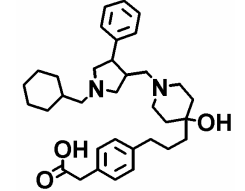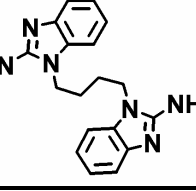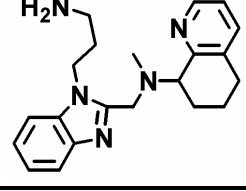
We then selected molecules for experimental testing. Commonly, this takes the form of visually inspecting the top ranking 500 molecules or up to 1% of the docking-ranked database. It is well-known that docking scoring functions are approximate and incomplete, but less discussed are problems with compound representation in database libraries (e.g., incorrect ionization states, overly strained conformations, and simple lack of availability

from vendors). More generally, what makes a good lead molecule reflects a plurality of not only orthogonal but sometimes opposed criteria. For instance larger, hydrophobic molecules will often bind tighter and score better, but biological efficacy and solubility often favors smaller, less hydrophobic molecules. Whereas we prefer molecules that engage all of their functional groups with the protein, we also are looking for the formation of key, "warhead" interactions with tightly defined geometric criteria. These and other extra-thermodynamic criteria have not been reduced to a single function—and given their opposed nature it might be difficult to do so—but may be rapidly evaluated by the eye of the trained investigator. In CXCR4, molecules were rejected in rough order of importance due to: 1) wrong ionization state, 2) unavailability, 3) high internal energy, 4) unsatisfied polar interactions and 5) low hit diversity. Molecules were prioritized for key salt-bridges to $E^{7.39}$ and at least one other anionic residue, plus a complementary fit to the binding-site.

Prior to release of the crystal structure we purchased 24 high-ranking molecules for testing, all in the top 1800 (0.05%) of 3.3 million molecules docked. One of these inhibited CXCL12 induced calcium flux in cell culture, with an IC_{50} of 107 μM, a hit rate of 4% (Compound **1**, **Table 2.1**, **Figure 2.1.A**). Compound **1** ranked 1725th and fit deep within the modeled binding-site, forming salt bridges to $E^{7.39}$ and $D^{6.58}$ in the putative docked complex (**Figure 2.2.A**). To measure chemical similarity to known CXCR4 ligands in ChEMBL09[25], we represented compound **1** by a two-dimensional topological fingerprint, ECFP4, and compared the bits (features) using the Tanimoto coefficient (T_c), as is widely done in the field[29]. Despite a T_c of 0.36, indicating marginal novelty, the molecule was a combination of two previously observed chemotypes, so we did not consider it particularly novel. Moreover, specificity counterscreens suggested that compound **1** also inhibited the related chemokine receptor CCR2.

Table 2.1 CXCR4 Parallel Virtual Screen Hits

Compound **1** discovered from homology model docking screen. Compounds **2-5** discovered from crystal structure docking screen.

#	Structure	Calcium flux IC$_{50}$ [uM]	Binding IC$_{50}$ [uM]	LE	Crystal Rank	Model Rank	T$_c$[a]	Closest[b]
1		107	N/A	0.26	37	1725	0.36	
2		76	N/A	0.24	418	5380[c]	0.23	
3		57	0.31	0.36	489	5800[c]	0.24	
4		77	25.1	0.25	137	30898[c]	0.23	
5		55	13.9	0.28	499	50121[c]	0.32	

[a] Tanimoto similarity to the most similar CXCR4 small molecule ligand in ChEMBL09 database

[b] Most similar CXCR4 small molecule ligand in ChEMBL09 database

[c] Ranks reported are not filtered for broken molecules

III. Crystal Structure Virtual Screen

With the crystal structure released, we again screened the lead-like subset of ZINC, now composed of 4.2 million molecules[28]. Docking statistics were similar, with each molecule sampled in an average of 10,200 orientations and 2,100 conformations, or 87 trillion complexes sampled overall (Supporting Methods). From among the top 0.03% of the docking hit list, we purchased 23 molecules for testing. Compounds **2-5** (17% hit rate) substantially inhibited CXCL12-induced calcium flux in cell culture, with IC_{50} values ranging from 55-77 μM (**Table 2.1**, **Figure 2.1.B-E**). In the docked poses, all four inhibitors formed salt bridges to $E^{7.39}$ and $D^{2.63}$ (**Figure 2.2.C-F**). Three formed salt-bridges through an

Figure 2.1 Dose-Response Curves for Inhibitors 1-5

(A-E) Percent calcium flux calculated as maximum minus minimum fluorescence as a percent of baseline (n=6). (F) [I^{125}]-CXCL12 radioligand displacement by compound **3**.
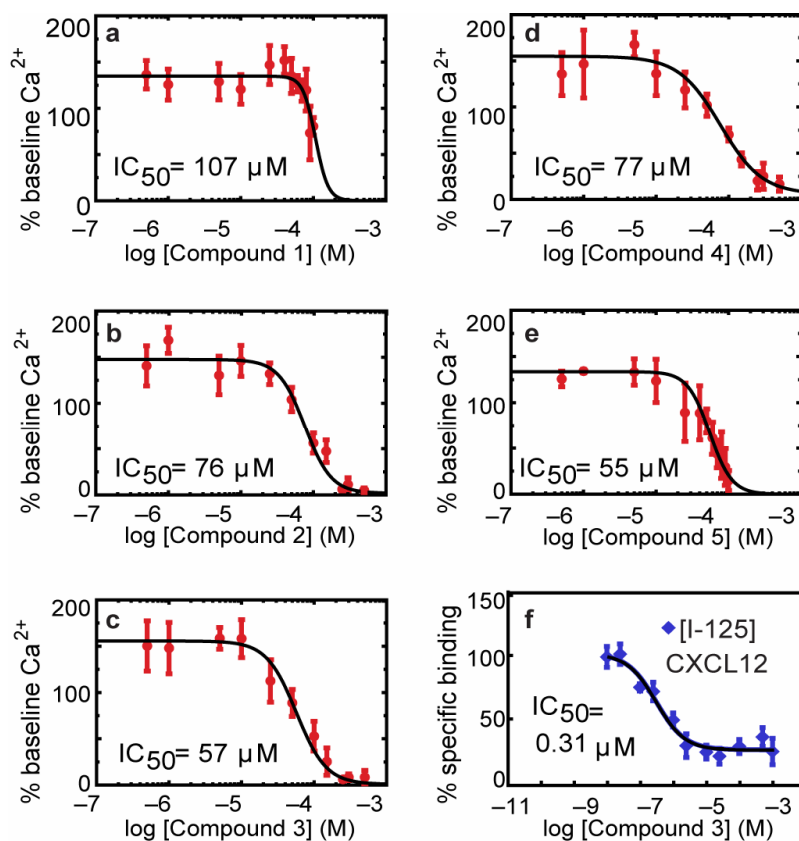
Figure 2.2 Docking Modes of Discovered Inhibitors

(A) The homology model with compound **1** (blue) in the docking mode from the homology model. (B) The crystal structure with compound **1** (yellow) docked. (C) Compound **2** (yellow) docked to crystal structure and the co-crystal ligand pose of small molecule 1T1t (blue lines). (D-F) Compounds **3**-**5** docked to crystal structure.
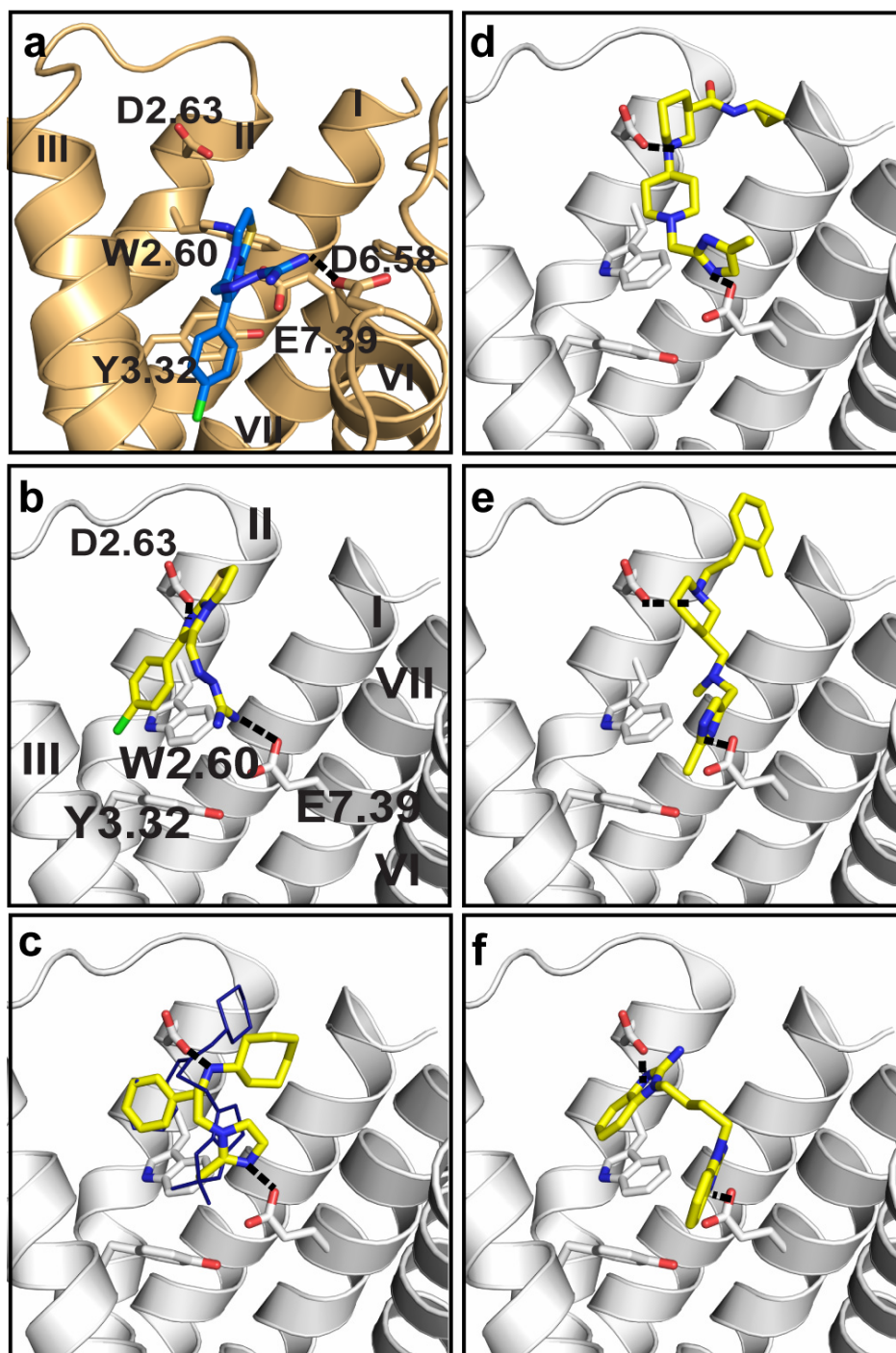
unprecedented imidazole functional group. Again comparing chemical similarity to known

CXCR4 ligands, T_c values of 0.23-0.32 supports their novelty (**Table 2.1**). While all four

inhibitors are also topologically distinct from one another, the similar docked poses suggest

that compound **3** and **4** fall into the same structural class. All of the new compounds have

molecular weights of 300-350 and calculated logP values of 0.5-3.5, placing them within the

lead-like range[30].

IV. Biological Activity

For biological relevance, compounds must not only inhibit calcium flux, but must

also inhibit lymphocyte migration. All five ligands inhibited human THP-1 monocyte

migration induced by CXCL12 in cell culture (**Figure 2.3, Table A.2.1**), with compounds **1**

and **2** almost completely inhibiting chemotaxis at 100 μM.

Small molecules may perturb chemokine signaling without competitive displacement

of the large chemokine protein, for instance binding under it in the TM part of the site[31, 32].

Figure 2.3 Inhibition of CXCL12 Induced Chemotaxis in THP-1 Cells.

(A) Following incubation with compound **1**-**5** the number of cells that migrated into the lower
chamber was counted. (B) Schematic of the Transwell chemotaxis chamber.
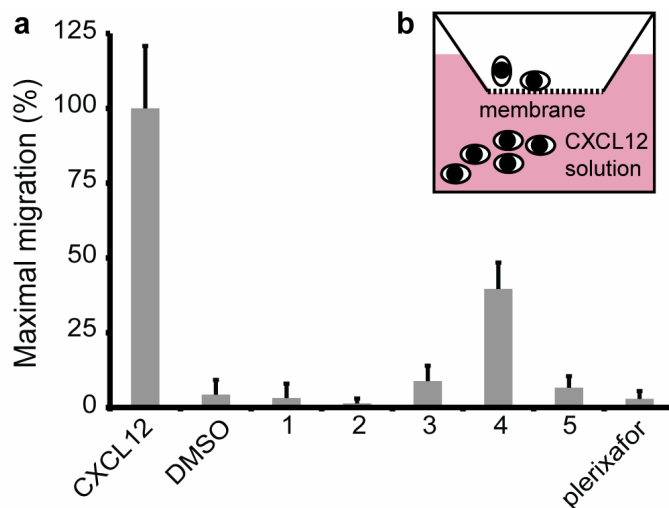
Four of the five ligands modulated the binding of radiolabeled CXCL12 (**Figure 2.1.F,**

Figure A.2.3). Compounds **3** and **5** disrupted CXCL12 binding with an IC_{50} of 306 nM and

14 µM, respectively. Compound **1** reduced binding with an IC_{50} of 224 µM, but had a steep

dose-response curve (aggregation counter screening, below). Compound **4,** though

efficacious as a signaling antagonist, actually increased CXCL12 binding, while compound **2**

did not modulate binding at all. These observations are consistent with the often allosteric

binding of small molecules to chemokine receptors.

Ligand efficiency (LE) corrects binding energies for size, dividing free energy of

binding expressed as RT $\log(IC_{50})$ (in kcal/mol) by heavy atom count. While LE is

notoriously low (poor) for PPI inhibitors, compound **3** has a LE of 0.36, putting it in the

range of that for oral drugs. The four other compounds have LE values of 0.24-0.28 (**Table

2.1**), more typical of PPI inhibitors.

V. Model Analysis

The release of the CXCR4 crystal structure allowed us to compare the prospective

model against the experimental structure. In the prospectively screened model, ligand IT1t

docks deep in the TM bundle, similar to ligand binding in the structural templates. In the

CXCR4 crystal structure, IT1t binds higher in the site (**Figure A.2.4**), resulting in a poor

RMSD of 9.5 Å between observed and predicted ligand position (**Table A.2.2**). Overall

binding-site agreement was better, with a transmembrane binding-site heavy-atom RMSD of

2.3 Å. Retrospective enrichment of known ligands was also high at 21% LogAUC (30%

LogAUC before modeling ECL2), comparing favorably with the crystal structure at 28%

LogAUC (**Figure A.2.5.A**). This placed it at the top of the enrichment distribution amongst

all the models we built, despite having an average binding-site RMSD to the crystal structure (**Figure A.2.6**).

We were interested to compare our predicted model to those with higher fidelity to the in the GPCR Dock 2010 Assessment[8], with a view to evaluating their usefulness for docking screens. We investigated the two top scoring models in the assessment, both of which predicted the IT1t ligand pose better than our model had, and computed their retrospective enrichment of ChEMBL04 ligands. These two models, VU-5[33] and COH-1[34] led to ligand enrichments of 5% and 6% LogAUC, respectively, using our docking method (**Table A.2.2, Figure A.2.5.A**). We also docked the five newly discovered ligands and over 3 million lead-like molecules in ZINC (**Table A.2.3, Figure A.2.5.B**); consistent with their modest retrospective enrichments of ChEMBL ligands, neither model ranked any of the new ligands well. Docked against the VU-5 model, the best scoring new ligand was compound **5** with a rank of 3282, while none of the other four ligands ranked better than 312,000. Similarly, against the COH-1 the best ranked ligand was compound **1**, which ranked 19,977, while no other ligand ranked above 63,000. These docking results take nothing away from the success of these models in the competition, but support the idea that even the field's best models, at this level of sequence identity, may struggle to achieve a structural fidelity high enough to support new ligand discovery.

Posing this question another way, we wondered if we ourselves had explored a model closer to the crystal structure, from among the several thousand calculated, that may have performed better in the docking. We retrospectively used ligand RMSD, known ligand enrichment, and binding site RMSD to select the most accurate model from among 2044 loopless models we originally sampled. The selected model had a much improved ligand RMSD of 2.9 Å (compared to the 9.5 Å we had originally predicted, **Table A.2.2**). Despite

enriching known ligands well (LogAUC of 22%, **Figure A.2.5.A**), its performance was substantially below the 30% LogAUC found for the model that was ultimately used in the prospective docking. Indeed, when we docked the five new ligands against what was structurally the best of our sampled models, their rankings were mediocre: the top scoring of these molecules was compound **2**, which ranked 15,344 of over 3 Million ZINC molecules docked, while three other new ligands ranked below 22,200. Meanwhile, for the truly prospective homology model, despite its poor predicting of the geometry of the crystallographic ligand IT1t, the top scoring new ligand in the docking was compound **1**, which ranked 2803 of the 3.3 Million ZINC molecules docked, while compounds **2, 3, 4** and **5** ranked 50121, 30898, 5380, and 5800, respectively. This suggests that a combination of geometric but also docking ranking criteria are appropriate in selecting models to be used for docking prediction, as has been suggested by others. Overall, it supports the idea that a certain minimum of sequence identity is required to be able to calculate a high fidelity model that can reliably select novel ligands, a point to which we will return.

VI. Aggregation Counter Screen

Colloidal aggregation constitutes perhaps the greatest source of false positives in screens against soluble proteins, but has not previously been observed for membrane-bound receptors. All of our ligands had Hill slopes in the calcium flux assay of 1.5 or higher, which is often associated with a colloidal-mechanism of inhibition. Though such a slope could be accounted for classically, by binding to a dimeric form of the receptor, as adopted in the crystal structure, all were counter-screened for non-specific inhibition due to aggregation[35, 36]. To test for aggregation under our exact assay conditions, spin-down precipitation was used to remove putative colloids[37]. After spin-down, the supernatant activity of compounds **1** to **5**

was unaffected, while the activity of compound **6** (**Table A.2.4**) was sharply attenuated. Due to the high Hill slope of 5 and inhibition of the counter-screen enzyme cruzain at 200 μM (see SI), we cannot completely discount the aggregation of compound **1** at high concentration, though at its IC_{50} value of 107 μM it seems well-behaved. Conversely, compound **6** (**Figure A.2.7**) inhibits CXCR4 via a colloidal aggregation mechanism, which constitutes the first description of aggregation-based activity against membrane-bound receptors; this mechanism may merit future vigilance in GPCR screening efforts. The behavior of all four other antagonists was consistent with well-behaved, classical binding to CXCR4.

2.4 Discussion

Two key results emerge from this study. First, five CXCR4 inhibitors, in three new chemotypes, were discovered; they are all substantially smaller than most known CXCR4 ligands giving them relatively favorable ligand efficiencies. Indeed the best of them, compound **3**, a 306 nM antagonist of the receptor, has a ligand efficiency (LE) of 0.36; its good physical properties put it well-within the lead-like range of compounds that might be optimized as tools and bioactive molecules. All five ligands inhibit CXCR4-mediated chemotaxis in cell culture. The four inhibitors derived from the x-ray screen are specific for CXCR4 versus CCR2, a close homolog, and so may hold potential as reagents to modulate HIV infection, metastasis and inflammation. Second, we compare a blind prospective virtual screen against a GPCR homology model to both a subsequent screen against the crystal structure, and to a twin study against dopamine receptor D3 (DRD3)[22]. The CXCR4 homology model had a hit rate of 4% , with a single antagonist of modest novelty and specificity; the crystal structure screen had a hit rate of 17%, with at least three of the four

ligands being novel and all four being specific. Conversely, docking against a DRD3 homology model discovered as many ligands as docking against the crystal structure. Contrasting these two targets and four campaigns illuminates the areas of the GPCR landscape that may be amenable to structure-based ligand discovery.

Protein-protein interfaces (PPI) are notoriously difficult to modulate with "drug-like" organic molecules. Few PPI inhibitors possess a LE greater than 0.23[7]; to achieve a reasonable affinity they are large and often hydrophobic, requiring extensive optimization to deliver them into biological milieus. In some ways CXCR4 is typical of PPI sites: at 20 Å across and 20 Å deep, its orthosteric site is much larger and more solvent exposed than those in biogenic amine GPCRs, for example, and bears a high net charge with at least five anionic residues. Conversely, CXCR4, though large, sports well-defined subsites where mixtures of charge and hydrophobic complementarity might be exploited by small molecules. Indeed the docking poses of the new inhibitors exploit such a subsite, also occupied by the co-crystal ligand, defined by $E^{7.39}$ and $D^{2.63}$. This may explain the unusually high LE (0.36) of compound **3**, which is far above that expected for most PPI inhibitors, and indeed for the one approved CXCR4 drug, plerixafor (LE of 0.25). Antagonists for chemokine receptors are PPI inihibitors with typical LE of 0.2, although LE as high as 0.43 (repertaxin, a CXCR1 inhbitor) have been reported[38]; the conserved nature of the chemokine receptor interface[32], and the high LE of compound **3**, suggest structure-based campaigns against additional chemokine receptors may also result in reagents with pharmaceutically relevant properties. Indeed maraviroc, a recently introduced HIV drug that targets the chemokine CCR5 receptor, has a ligand efficiency of 0.33 placing it within the range expected of most oral drugs; it is one of only a few PPI drugs on the market[7].

From a technical perspective, it is interesting to ask why the CXCR4 homology model performed so much worse than either the dopamine D3 homology model or the crystal structure of CXCR4. One important contribution was ligand-bias in the database. A great advantage of docking against the dopamine receptor was the bias toward biogenic amine mimetics in even an "unbiased" library such as ZINC, which simply catalogs commercially available molecules. Thus, there were not only many dopaminergic-like molecules in ZINC to find, but also many analogs of these hits in ZINC, allowing an SAR by catalog campaign that drove affinity from 1.6 μM for an original docking hit to 81 nM for an optimized lead. This bias was also observed in docking screens against the adenosine A2a receptor[4, 39] and the β2 adrenergic receptor[3]. Several lines of evidence suggest that the bias towards CXCR4-like ligands was much reduced in ZINC: there are relatively few molecules that share the same size and charge properties as known ligands, and in contrast to DRD3 we found very few analogs in the database even for our newly discovered lead-like ligands.

A larger contribution to the weakness of the prospective CXCR4 model screen was clearly accuracy of the model itself. The DRD3 model, with 42% sequence identity to its template, closely resembles the crystal structure (binding-site RMSD of 1.65 Å), and the large number of known ligands and ample mutational data helped to correctly predict the co-crystal ligand pose (also 1.65 Å RMSD). This level of accuracy was sufficient to attain a 23% hit rate in a virtual screen of the DRD3 model. While the relatively poor 4% hit rate of the prospective CXCR4 homology model may simply reflect our ineptness, our models were competitive with the field, predicting the CXCR4-IT1t conformation better than all but a few models submitted to the public competition[8]. To further assess the overall state of CXCR4 homology modeling, we performed docking screens against the two most accurate competition CXCR4 models, assessing how well they ranked the five new ligands discovered

herein. Against these external models, retrospective enrichment of previously known CXCR4 ligands was poor, as was ranking of our five new hits. The relatively poor outcome of screening CXCR4 homology models versus screening a DRD3 homology model likely reflects the reduced accuracy that can be achieved at 25% vs 42% sequence identity to structural templates, even with the best homology models the field now offers.

While developing more accurate homology models is not a goal of this study, the lesson we draw is that for GPCRs sharing 42% or better sequence identity with a structurally-determined template, and with sufficient mutant studies to predict ligand binding, accurate models may be within the reach of general approaches. For those targets with lower sequence identities, certainly in the 18 to 25% range that characterized the templates available to us, homology models accurate enough for predictive ligand discovery may be out of reach, even with domain expertise. Putting aside issues of disease-relevance and experimental pragmatism, which will naturally dominate, one might imagine an additional prioritization axis for future GPCR crystal structures that considers the number of new targets they enable to be reliably modeled.

Those technical points should not obscure the key, biological result from this study: the ability to discover new chemical matter, with favorable physical properties, for this critical protein-protein interface. Given the difficulties for which these PPI targets are notorious, and a lack of favorable bias in the docking library, we were uncertain as to whether even the CXCR4 *crystal structure* would lead to new ligands. Instead, the 17% hit rate observed was substantial, certainly much higher than we have experienced with soluble enzymes[40, 41] and three novel chemotypes emerged. The ligand with the highest affinity by radioligand displacement, compound **3**, had an IC_{50} of 306 nM and a ligand efficiency of 0.36. This is within the range of favorable leads for drug discovery, well above the 0.23

ligand efficiency expected for most PPI inhibitors[7]. All of the new inhibitors were active in cell culture, inhibiting CXCR4-mediated chemotaxis, the primary cellular endpoint for a chemokine receptor ligand, consistent with their promise as leads. More generally, this study suggests that structures of chemokine receptors will provide pragmatic templates for probe and drug discovery; such molecules are much needed for biological understanding and for treating devastating diseases in cancer, virology, and inflammation.

2.5 Methods

I. Homology Modeling and Docking

Homology modeling and docking proceeded as described (Results, Supporting Methods). The large solvent exposed CXCR4 binding-site presents a challenge for docking: to compensate, we used a new procedure to balance electrostatics with rapid context-dependent ligand desolvation[26]. We filled the CXCR4 pocket with a single layer of low-dielectric spheres, excluding any spheres displaced from the surface to perturb the bulk-dielectric minimally, while allowing the ligand to interact strongly with charged groups throughout the binding-site. The large binding cavity also presented a challenge for exhaustive ligand sampling. To compensate, we divided the binding pocket into three partially overlapping sub-sites for sampling and docked separately against each (this reduces number of orientations by about 3^4)[42]. A single scoring grid was used to represent the entire site.

II. Calcium Flux-based Assays

THP-1 monocytes were resuspended in assay buffer containing FLIPR Calcium4 dye. Compounds were added at 100 μM (single point) or the indicated concentrations (dose-

response). After a 20 s baseline measurement, CXCL12 was added at 30 nM and resulting calcium response was measured for an additional 50 s. CCL2, a chemokine that targets CCR2 (a distinct receptor of THP-1 cells) was added as a control for compound specificity during the single point compound screening. Approximately 60 s after CXCL12 addition, 6 nM CCL2 was added to each well and calcium mobilization was measured for an additional 40 s. Percent calcium flux for each agonist was calculated from the maximum fluorescence minus the minimum fluorescence as a percent of baseline. A two-tailed student's t-test between either the 30 nM CXCL12 control or the 6 nM CCL2 control, and the compound of interest was used to identify statistically significant inhibitory compounds (**Figure A.2.8.a-b**). For significant CXCR4 inhibitors, the assay was repeated in a dose-response format.

III. Chemotaxis and Viability

Chemotaxis experiments were performed in THP-1 cells as described (Supporting Methods). CXCL12 ligand (30 nM) and respective compounds (100 μM) was added to the lower chamber. Percent maximal migration was calculated as the number of migrated cells with compound divided by number that migrated to CXCL12 alone.

IV. Radioligand Binding

Binding studies were performed on pre-B leukemia REH cells as described (Supporting Methods). The competition binding assays were carried out using 50 pM $[I^{125}]$-CXCL12 as a tracer.

V. Counter-Screens for Aggregation

In spin-down counter screens for aggregation, compounds were centrifuged at 16,000×G for 20 min. Supernatant was removed and used for calcium flux experiments as above. Cruzain inhibition assays were performed as reported[36].

VI. Compound Sources

Compounds were obtained from the National Cancer Institute and commercial suppliers. All active compounds were tested for purity by LC/MS at UCSF and were judged pure by peak height and identity (Supporting Methods).

2.6 Acknowledgements

I. Author Contributions

BKS, NH and BFV designed the study. MMM and DRW performed homology modeling and docking. JJZ performed all calcium flux and chemotaxis assays. SG performed radioligand binding assays. AKD performed aggregation counter screening. JK performed

loop modeling. All authors contributed to the writing of the manuscript. The authors declare

no competing financial interest.

2.7 References

1. Overington, JP; Al-Lazikani, B; Hopkins, AL. "How Many Drug Targets Are There?" Nat Rev Drug Discov 2006 5: 993-996.

2. Katritch, V; Jaakola, VP; Lane, JR; Lin, J; Ijzerman, AP; Yeager, M; Kufareva, I; Stevens, RC; Abagyan, R. "Structure-Based Discovery of Novel Chemotypes for Adenosine a(2a) Receptor Antagonists" J Med Chem 2010 53: 1799-1809.

3. Kolb, P; Rosenbaum, DM; Irwin, JJ; Fung, JJ; Kobilka, BK; Shoichet, BK. "Structure-Based Discovery of Beta2-Adrenergic Receptor Ligands" Proc Natl Acad Sci USA 2009 106: 6843-6848.

4. Carlsson, J; Yoo, L; Gao, ZG; Irwin, JJ; Shoichet, BK; Jacobson, KA. "Structure-Based Discovery of A2a Adenosine Receptor Ligands" J Med Chem 2010 53: 3748-3755.

5. Sabio, M; Jones, K; Topiol, S. "Use of the X-Ray Structure of the Beta2-Adrenergic Receptor for Drug Discovery. Part 2: Identification of Active Compounds" Bioorg Med Chem Lett 2008 18: 5391-5395.

6. Veldkamp, CT; Ziarek, JJ; Peterson, FC; Chen, Y; Volkman, BF. "Targeting Sdf-1/Cxcl12 with a Ligand That Prevents Activation of Cxcr4 through Structure-Based Drug Design" J Am Chem Soc 2010 132: 7242-7243.

7. Wells, JA; McClendon, CL. "Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces" Nature 2007 450: 1001-1009.

8. Kufareva, I; Rueda, M; Katritch, V; Participants; Stevens, RC; Abagyan, R. "Status of Gpcr Modeling and Docking as Reflected by Community-Wide Gpcr Dock 2010 Assessment" Structure 2011 19: 1108-1126.

9. Zou, YR; Kottmann, AH; Kuroda, M; Taniuchi, I; Littman, DR. "Function of the Chemokine Receptor Cxcr4 in Haematopoiesis and in Cerebellar Development" Nature 1998 393: 595-599.

10. Ma, Q; Jones, D; Borghesani, PR; Segal, RA; Nagasawa, T; Kishimoto, T; Bronson, RT; Springer, TA. "Impaired B-Lymphopoiesis, Myelopoiesis, and Derailed Cerebellar Neuron Migration in Cxcr4- and Sdf-1-Deficient Mice" Proc Natl Acad Sci USA 1998 95: 9448-9453.

11. Hu, X; Dai, S; Wu, WJ; Tan, W; Zhu, X; Mu, J; Guo, Y; Bolli, R; Rokosh, G. "Stromal Cell Derived Factor-1 Alpha Confers Protection against Myocardial Ischemia/Reperfusion Injury: Role of the Cardiac Stromal Cell Derived Factor-1 Alpha Cxcr4 Axis" Circulation 2007 116: 654-663.

12. Hernandez, PA; Gorlin, RJ; Lukens, JN; Taniuchi, S; Bohinjec, J; Francois, F; Klotman, ME; Diaz, GA. "Mutations in the Chemokine Receptor Gene Cxcr4 Are Associated with Whim Syndrome, a Combined Immunodeficiency Disease" Nat Genet 2003 34: 70-74.

13. Endres, MJ; Clapham, PR; Marsh, M; Ahuja, M; Turner, JD; McKnight, A; Thomas, JF; Stoebenau-Haggarty, B; Choe, S; Vance, PJ; Wells, TN; Power, CA; Sutterwala, SS; Doms, RW; Landau, NR; Hoxie, JA. "Cd4-Independent Infection by Hiv-2 Is Mediated by Fusin/Cxcr4" Cell 1996 87: 745-756.

14. Burger, JA; Kipps, TJ. "Cxcr4: A Key Receptor in the Crosstalk between Tumor Cells and Their Microenvironment" Blood 2006 107: 1761-1767.

15. Thoma, G; Streiff, MB; Kovarik, J; Glickman, F; Wagner, T; Beerli, C; Zerwes, HG. "Orally Bioavailable Isothioureas Block Function of the Chemokine Receptor Cxcr4 in Vitro and in Vivo" J Med Chem 2008 51: 7915-7920.

16. Zhu, A; Zhan, W; Liang, Z; Yoon, Y; Yang, H; Grossniklaus, HE; Xu, J; Rojas, M; Lockwood, M; Snyder, JP; Liotta, DC; Shim, H. "Dipyrimidine Amines: A Novel Class of Chemokine Receptor Type 4 Antagonists with High Specificity" J Med Chem 2010 53: 8556-8568.

17. De Clercq, E. "Recent Advances on the Use of the Cxcr4 Antagonist Plerixafor (Amd3100, Mozobil) and Potential of Other Cxcr4 Antagonists as Stem Cell Mobilizers" Pharmacol Ther 2010 128: 509-518.

18. Kellenberger, E; Springael, JY; Parmentier, M; Hachet-Haas, M; Galzi, JL; Rognan, D. "Identification of Nonpeptide Ccr5 Receptor Agonists by Structure-Based Virtual Screening" J Med Chem 2007 50: 1294-1303.

19. Evers, A; Klebe, G. "Successful Virtual Screening for a Submicromolar Antagonist of the Neurokinin-1 Receptor Based on a Ligand-Supported Homology Model" J Med Chem 2004 47: 5381-5392.

20. Cavasotto, CN; Orry, AJ; Murgolo, NJ; Czarniecki, MF; Kocsi, SA; Hawes, BE; O'Neill, KA; Hine, H; Burton, MS; Voigt, JH; Abagyan, RA; Bayne, ML; Monsma, FJ, Jr. "Discovery of Novel Chemotypes to a G-Protein-Coupled Receptor through Ligand-Steered Homology Modeling and Structure-Based Virtual Screening" J Med Chem 2008 51: 581-588.

21. Tikhonova, IG; Sum, CS; Neumann, S; Engel, S; Raaka, BM; Costanzi, S; Gershengorn, MC. "Discovery of Novel Agonists and Antagonists of the Free Fatty Acid Receptor 1 (Ffar1) Using Virtual Screening" J Med Chem 2008 51: 625-633.

22. Carlsson, J; Coleman, RG; Setola, V; Irwin, JJ; Fan, H; Schlessinger, A; Sali, A; Roth, BL; Shoichet, BK. "Ligand Discovery from a Dopamine D3 Receptor Homology Model and Crystal Structure." Nat Chem Biol 2011 7: 769-778.

23. Wu, B; Chien, EY; Mol, CD; Fenalti, G; Liu, W; Katritch, V; Abagyan, R; Brooun, A; Wells, P; Bi, FC; Hamel, DJ; Kuhn, P; Handel, TM; Cherezov, V; Stevens, RC. "Structures of the Cxcr4 Chemokine Gpcr with Small-Molecule and Cyclic Peptide Antagonists" Science 2011 330: 1066-1071.

24. Yang, Q; Sharp, KA. "Building Alternate Protein Structures Using the Elastic Network Model" Proteins 2009 74: 682-700.

25. Gaulton, A; Bellis, LJ; Bento, AP; Chambers, J; Davies, M; Hersey, A; Light, Y; McGlinchey, S; Michalovich, D; Al-Lazikani, B; Overington, JP. "Chembl: A Large-Scale Bioactivity Database for Drug Discovery" Nucleic Acids Res 2011.

26. Mysinger, MM; Shoichet, BK. "Rapid Context-Dependent Ligand Desolvation in Molecular Docking" J Chem Inf Model 2010 50: 1561-1573.

27. Wong, RS; Bodart, V; Metz, M; Labrecque, J; Bridger, G; Fricker, SP. "Comparison of the Potential Multiple Binding Modes of Bicyclam, Monocylam, and Noncyclam Small-Molecule Cxc Chemokine Receptor 4 Inhibitors" Mol Pharmacol 2008 74: 1485-1495.

28. Irwin, JJ; Shoichet, BK. "Zinc--a Free Database of Commercially Available Compounds for Virtual Screening" J Chem Inf Model 2005 45: 177-182.

29. Hert, J; Willett, P; Wilton, DJ; Acklin, P; Azzaoui, K; Jacoby, E; Schuffenhauer, A. "Comparison of Topological Descriptors for Similarity-Based Virtual Screening Using Multiple Bioactive Reference Structures" Org Biomol Chem 2004 2: 3256-3266.

30. Ursu, O; Rayan, A; Goldblum, A; Oprea, TI. "Understanding Drug-Likeness" Wiley Interdisciplinary Reviews-Computational Molecular Science 2011 1: 760-781.

31. Vaidehi, N; Schlyer, S; Trabanino, RJ; Floriano, WB; Abrol, R; Sharma, S; Kochanny, M; Koovakat, S; Dunning, L; Liang, M; Fox, JM; de Mendonca, FL; Pease, JE; Goddard, WA, 3rd; Horuk, R. "Predictions of Ccr1 Chemokine Receptor Structure and Bx 471 Antagonist Binding Followed by Experimental Validation" J Biol Chem 2006 281: 27613-27620.

32. Allegretti, M; Bertin, R; Bizzarri, C; Beccari, A; Mantovani, A; Locati, M. "Allosteric Inhibitors of Chemoattractant Receptors: Opportunities and Pitfalls" Trends Pharmacol Sci 2008 29: 280-286.

33. Roumen, L; Sanders, MPA; Vroling, B; De Esch, IJP; De Vlieg, J; Leurs, R; Klomp, JPG; Nabuurs, SB; De Graaf, C. In Silico Veritas: The Pitfalls and Challenges of Predicting Gpcr-Ligand Interactions. In 2011; Vol. 4, pp 1196-1215.

34. Lam, AR; Bhattacharya, S; Patel, K; Hall, SE; Mao, A; Vaidehi, N. "Importance of Receptor Flexibility in Binding of Cyclam Compounds to the Chemokine Receptor Cxcr4" J Chem Inf Model 2011 51: 139-147.

35. McGovern, SL; Helfand, BT; Feng, B; Shoichet, BK. "A Specific Mechanism of Nonspecific Inhibition" J Med Chem 2003 46: 4265-4272.

36. Doak, AK; Wille, H; Prusiner, SB; Shoichet, BK. "Colloid Formation by Drugs in Simulated Intestinal Fluid" J Med Chem 2010 53: 4259-4265.

37. Coan, KE; Shoichet, BK. "Stoichiometry and Physical Chemistry of Promiscuous Aggregate-Based Inhibitors" J Am Chem Soc 2008 130: 9606-9612.

38. Allegretti, M; Bertini, R; Cesta, MC; Bizzarri, C; Di Bitondo, R; Di Cioccio, V; Galliera, E; Berdini, V; Topai, A; Zampella, G; Russo, V; Di Bello, N; Nano, G; Nicolini, L; Locati, M; Fantucci, P; Florio, S; Colotta, F. "2-Arylpropionic Cxc Chemokine Receptor 1 (Cxcr1) Ligands as Novel Noncompetitive Cxcl8 Inhibitors" J Med Chem 2005 48: 4312-4331.

39. Katritch, V; Rueda, M; Lam, PC; Yeager, M; Abagyan, R. "Gpcr 3d Homology Models for Ligand Screening: Lessons Learned from Blind Predictions of Adenosine A2a Receptor Complex" Proteins 2010 78: 197-211.

40. Powers, RA; Morandi, F; Shoichet, BK. "Structure-Based Discovery of a Novel, Noncovalent Inhibitor of Ampc Beta-Lactamase" Structure 2002 10: 1013-1023.

41. Chen, Y; Shoichet, BK. "Molecular Docking and Ligand Specificity in Fragment-Based Inhibitor Discovery" Nat Chem Bio 2009 5: 358-364.

42. Shoichet, BK; Bodian, DL; Kuntz, ID. "Molecular Docking Using Shape Descriptors" J Comput Chem 1992 13: 380-397.

Gloss to Chapter 3

By this point, our rapid context-dependent ligand desolvation term had given prospective hit rates of 17% and 25% in virtual screens to CXCR4 and DRD3, and looked promsing in several ongoing ligand discovery projects. Yet, in chapter one the desolvation method only did better than using no desolvation by a seemingly modest 0.7% average LogAUC, over the 40 DUD targets. Yet shockingly, I had performed many docking experiments to that charge-matched DUD set, and a 0.7% average gain was one of the largest perturbations I observed. Thus we started to believe we could build a better benchmarking set, with which we could better test future docking methods. Additionally, in making the charge-matched decoys for chapter one, I had automated, improved, and stabilized many stages of our docking toolchain, which would give me a head start at constructing this newer, larger, more ambitious benchmarking set.

The Directoy of Useful Decoys (DUD)[1], despite the acronym, had been hugely successful; so it was fun that an enhanced version became DUD-E, with homage to the movie "The Big Lebowski". More and more diverse targets were included, to a total of 102. Ligands were improved by adding more, each with measured affinity, and then clustering them to improve diversity. Decoys were improved by adding net charge to propery-matching, by adapting to chemical space during property-matching, and by using the most dissimilar decoys to reduce the number of decoys that might actually bind.

We decomposed the changes in performance between DUD-E and DUD, to find that the DUD-E style decoys increase overall enrichment due to fewer false decoys, the new

ligands reduce enrichment due to diversity, and the new target preparations increase enrichment due to a large scale docking campaign that selected better crystal structures. We again examined our context-dependent ligand desolvation term, finding that in DUD-E it handily outperforms no desolvation by 3.8% average LogAUC over the 102 targets. To help the balance with electrostatics, we automate the thin electrostatic dielectric layer that we manually constructed for CXCR4 in chapter two. We find that thin layers improve average LogAUC by 1.0% on the 96 targets with automatically generated dielectric layers. This suggests that over a more comprehensive set of targets, and what we argue is a better set of ligands and decoys, the advantage of more physically correct treatments becomes more pronounced.

While we expect this DUD-E release (http://dude.docking.org) to be used to test docking methods well into the future, it was also designed to be easy to refine and extend. In addition to ready to download canonical sets, all the data and many scripts used in the sets' construction can also be downloaded. A decoy server (http://decoys.docking.org) is provided that uses the same algorithm which made DUD-E decoys, and takes any user supplied list of ligands. This enables challenging retrospective performance tests on any target of interest.

I. References

1. Huang, N; Shoichet, BK; Irwin, JJ. "Benchmarking Sets for Molecular Docking" J Med Chem 2006 49: 6789-6801.

Chapter 3:

DUD-Enhanced – Better Ligands and Decoys for Better Benchmarking

Michael M. Mysinger[1], Michael Carchia[1], John. J. Irwin[1]*, Brian K. Shoichet[1]*

1   Department of Pharmaceutical Chemistry, University of California, San Francisco

* Corresponding authors (BKS & JJI)

The supporting information for this paper has been included as Appendix A.3.

3.1 Abstract

A key metric to assess molecular docking remains ligand enrichment against challenging decoys. Whereas the directory of useful decoys (DUD) benchmarking set has been widely used, clear areas for optimization have emerged. Here we describe an improved benchmarking set that includes more diverse and biomedically relevant targets, such as GPCRs and ion channels, totaling 102 proteins with 22,886 clustered ligands drawn from ChEMBL, each with 50 property-matched decoys drawn from ZINC. To ensure chemotype diversity we cluster the ligands of each target by Bemis-Murcko frameworks. To improve the decoys, we add net charge as an additional matched physico-chemical property, and only include those decoys that are most dissimilar, by topology, from the ligands. Further, we link every decoy with its particular ligand to enable detailed set decomposition. An online automated tool (http://decoys.docking.org) generates these improved matched decoys for any given ligand set. We test this dataset by docking ligands and decoys against all 102 targets, using the results to improve the balance between ligand desolvation and electrostatics in DOCK 3.6. The full DUD-E benchmarking set is freely available at http://dude.docking.org.

3.2 Introduction

While molecular docking screens routinely leverage protein structure to discover new ligands[1-4], quantitative assessment of their performance remains problematic[5]. Though prospective assessment of docking performance is irreplaceable[6, 7], it is both time consuming and expensive. Since a general correlation between docking scores and affinities is beyond current methods[8, 9], the field relies on ligand enrichment in docking hit lists to evaluate retrospective performance[10-14]. "Enrichment" measures how known ligands rank versus a

background of decoy molecules, and so depends not only on the nature of the ligands, but also on the background decoys. Thus to compare docking enrichments, a benchmarking set of ligands and decoys is needed.

The original Directory of Useful Decoys (DUD) was designed to meet this benchmarking need while controlling for decoy bias on enrichment[15, 16]. Given a random drug-like set of decoys, Verdonk *et. al.* showed that targets which bind high molecular weight ligands naturally get higher enrichments, due to correlation between larger molecules and better docking scores[17]. In contrast, actual ligand binding affinities correlate with molecular size only for very small molecules[18]. Unable to separate the true correlations of simple molecular properties that aid prospective ligand discovery from the artificial correlations that arise from biases, it is informative to ask what value molecular docking adds beyond these properties. To this end, DUD decoys are matched to the physical chemistry of ligands on a target-by-target basis: by the properties of molecular weight, calculated logP, number of rotatable bonds, hydrogen bond donors and acceptors. To fulfill their role as negative controls, decoys should not actually bind, so DUD used 2-D similarity fingerprints to minimize the topological similarity between decoys and ligands. In short, DUD decoys were chosen to resemble ligands physically and so be challenging for docking, but at the same time be topologically dissimilar to minimize the likelihood of actual binding.

Through intense use[19-26], weaknesses in the original DUD set have appeared in both the ligands and decoys. Good and Oprea noted that a handful of chemotypes dominate many ligand sets, allowing high ranks for one scaffold to cause good overall enrichment[27]. One way to circumvent this problem is using chemotype retrieval metrics[28], but another is to remove the "analog bias" from the database by clustering on ligand scaffolds. After clustering the 40 targets, Good's subset of DUD contains only 13 targets with over 15

ligands, indicating a need for more targets with more ligands. Another important goal is to increase target diversity, for example by adding membrane domain proteins, none of which are represented in DUD.

As there were weaknesses in the DUD ligands, this was also true of the decoys. Several investigators[29-31] observed that despite property matching on logP, net formal charge is still imbalanced in DUD; 42% of all ligands are charged versus only 15% of decoys. Property matching of decoys to ligands could also be tightened by choosing decoys more embedded in ligand property-space[32, 33]. Despite a 2-D chemical dissimilarity filter to prevent decoys from being active, some original DUD decoys still appear to bind, and these false decoys artificially reduce docking enrichment[32]. Addressing both false decoys and decoy property embedding, Vogel *et. al.* released DEKOIS for the original 40 DUD targets. Gatica and Cavasotto generated ligand and decoy sets for 147 G Protein-Coupled Receptors (GPCRs) while adding net charge to property matching[34]. Very recently, a python GUI application was announced to generate property-matched decoys[35]. By ignoring synthetic feasibility, Wallach and Lilien generate virtual decoy sets for the original DUD targets with tighter property-matching[33]. Instead of generating computational decoys, the MUV set selects decoys for 17 targets that were negative in public high throughput screens[36]. Instead of generating decoys at all, REPROVIS-DB assembles ligand and database data from earlier successful virtual screens which are deemed reproducible[37].

Here we describe a new version of DUD that addresses these liabilities and develops new functionality. By drawing on ChEMBL09[38], each DUD-Enhanced (DUD-E) ligand has a measured affinity supported by a literature reference. Though ligands are now typically clustered by Bemis-Murcko frameworks[39] to reduce chemotype bias, there are still on average 224 ligands per target. The target list is expanded from 40 to 102, favoring targets

with many ligands and multiple[40] structures. The additions include several drug relevant membrane proteins: five GPCRs, two ion channels, and two cytochrome P450s. Meanwhile, false decoys are reduced by more stringent filtering of topological dissimilarity. Where possible, measured experimental decoys are included. Finally, we consider how DUD-E performs as a benchmark versus the original DUD, and explore its use as a tool for evaluating and optimizing molecular docking.

3.3 Results

The ideal target for a benchmarking set would be well studied, with many measured ligand affinities and multiple, diverse co-crystal ligand structures. To this end, the enhanced DUD database (DUD-E) is largely based on the intersection of ChEMBL[38], for ligand annotations and affinities, and the RCSB PDB[40], for structures. As we sought targets to enlarge the set, the 40 original DUD targets were first priority, 38 of which we included. Platlet-derived growth factor receptor β was dropped, as it was a homology model. Estrogen receptor α (ESR1) is a single target in DUD-E, whereas it was split into agonist and antagonists previously. To enlarge the benchmarking set, we used three main criteria. First, we favored new target classes with pharmacological precedence. Second, we sought targets with many ligands and crystal structures, as they likely reflect a combination of target relevance and ease of study. Third, we preferred targets that could modestly enrich known ligands using fully automated docking, as these may be both easy to prepare and amenable to docking. Conversely, targets with mostly covalent ligands were de-prioritized.

DUD-E targets are defined by their UniProt[41] gene prefix, with data from each species being combined into a single dataset. While ChEMBL annotates ligands to a particular UniProt accession code, the ligand overlap between orthologous targets is

surprisingly small. For example, among 1555 unique ligands with affinities below 1 μM for the human dopamine D3 receptor and 744 ligands for the rat ortholog, only 85 ligands are in both sets. These two orthologs share 97% trans-membrane sequence identity (79% overall), so this low overlap suggests to us that ChEMBL ligand annotations are sparse, and do not typically reflect species specificity. Therefore we pooled the data for all species, defining a DUD-E target as a UniProt gene prefix (such as DRD3), and not the full gene_species pair (such as DRD3_HUMAN or P35462).

The 102 targets span diverse protein categories, including 26 kinases, 15 proteases, 11 nuclear receptors, 5 GPCRs, 2 ion channels, 2 cytochrome P450s, 36 other enzymes, and 5 miscellaneous proteins (**Figure 3.1**). Altogether 66,695 raw ligands, defined as those with annotated affinities better than 1 μM to their target, molecular weights less than 600, and fewer than 20 rotatable bonds were extracted from ChEMBL09 (or the AmpC β-lactamase

Figure 3.1 DUD-E Target Classification

Number of the 102 targets that belong to eight broad protein categories.
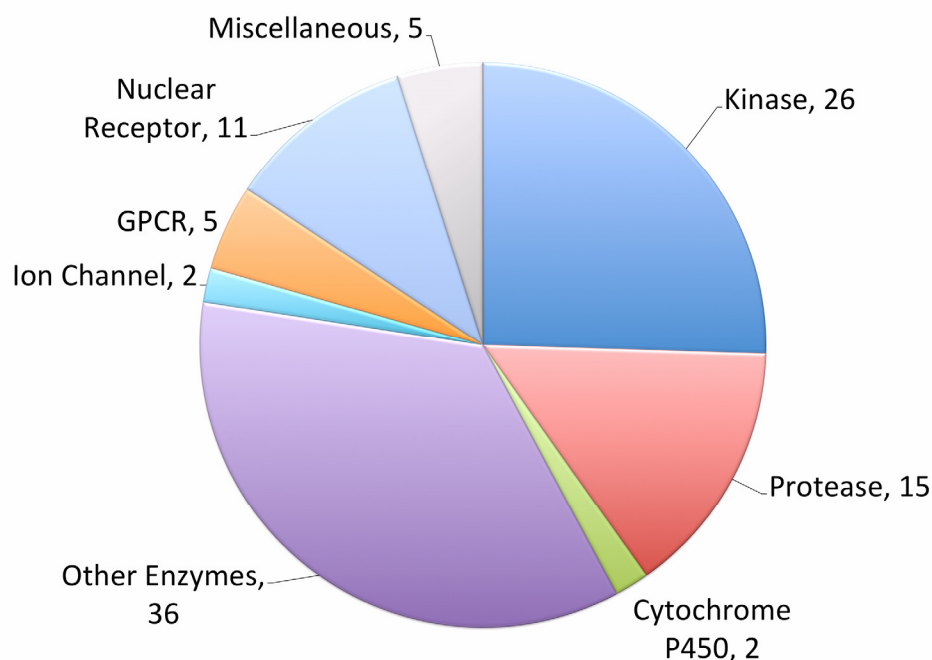
literature) (**Table 3.1**). That is an average of 654 ligands per target with a minimum of 40 and a maximum of 3,090. Though negative binding is rarely reported, we also found 9,219 experimental decoys (i.e. no measurable affinity up to 30 µM), with a maximum of 1,070 for cyclo-oxygenase-1 (PGH1).

Table 3.1 Characteristics of DUD-E

	Total	ChEMBL	Manual	
# Targets	102	101	1	
	Total	Average	Minimum	Maximum
# Raw Ligands	66,695	653.9	40	3,090
# Clustered Ligands	22,886	224.4	40	592
# Experimental Decoys	9,219	90.4	1	1070
# Clustered Ligands Unique Charge States	28,377	278.2	46	1030
# Computational Decoys	1,411,214	13,835	2,300	51,500

With targets selected, we chose a single X-ray structure to represent each target in docking studies (**Table 3.2**, **Table A.3.1**). To find the structure most amenable to docking, we used an automated docking campaign to screen 3690 PDB structures against their clustered ligands and property-matched decoys (see below). Preference was given to higher resolution, to higher automated enrichment, and to the human ortholog. We avoided mutant structures, unresolved active site loops, extraneous bound peptides, or structures too constrained for many of that target's ligands. Where we had domain knowledge, the most representative structure was preferred, for example a DFG-in structure for kinases or an antagonist structure for estrogen receptor α (ESR1). For 57 out of 102 targets, a DOCK Blaster[42] prepared structure was used for DUD-E, directly from the automated tool chain. Another 45 targets required manual intervention, most due to simple errors in automated

Table 3.2 Overview of Representative Targets

Class	Id	Description	Total Ligands	Clustered Ligands	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
Cytochrome P450	CP2C9	Cytochrome P450 2C9	145	120	7,450	1R9O	7	3	60
	CP3A4	Cytochrome P450 3A4	302	170	11,800	3NXU	7	2	63
GPCR	AA2AR	Adenosine A2a receptor	3057	482	31,550	3EML	28	22	83
	ADRB1	Beta-1 adrenergic receptor	648	247	15,850	2VT4	19	11	76
	CXCR4	C-X-C chemokine receptor type 4	40	40	3,406	3ODU	36	18	90
Ion Channel	GRIA2	Glutamate receptor ionotropic, AMPA 2	476	158	11,845	3KGC	23	23	71
	GRIK1	Glutamate receptor ionotropic kainate 1	136	101	6,550	1VSO	35	27	86
Kinase	AKT1	Serine/threonine-protein kinase AKT	585	293	16,450	3CQW	27	29	72
	MK10	c-Jun N-terminal kinase 3	199	104	6,600	2ZDT	24	11	82
	MK14	MAP kinase p38 alpha	2205	578	35,850	2QD9	17	10	74
Miscel-laneous	KIF11	Kinesin-like protein 1	272	116	6,850	3CJO	34	35	77
	XIAP	Inhibitor of apoptosis protein 3	100	100	5,150	3HL5	52	55	88
Nuclear Receptor	ESR1	Estrogen receptor alpha	1297	383	20,685	1SJ0	18	15	67
	MCR	Mineralocorticoid receptor	201	94	5,150	2AA2	-4	2	36
	THB	Thyroid hormone receptor beta-1	246	103	7,450	1Q4X	36	38	79
	PPARD	Peroxisome proliferator-activated receptor delta	699	240	12,250	2ZNP	32	20	89
Other Enzymes	FNTA	Protein farnesyltransferase type I alpha	1430	592	51,500	3E37	16	7	76
	HDAC8	Histone deacetylase 8	309	170	10,450	3F07	29	24	80
	HIVINT	HIV type 1 integrase	167	100	6,650	3NF7	8	2	64
	KITH	Thymidine kinase	57	57	2,850	2B8T	15	0	80
	PARP1	Poly [ADP-ribose] polymerase-1	1031	508	30,050	3L3M	25	21	79
	PUR2	GAR transformylase	50	50	2,700	1NJS	51	50	92
Protease	DPP4	Dipeptidyl peptidase IV	1939	533	40,950	2I78	41	41	87
	FA10	Coagulation factor X	3090	537	28,325	3KL6	39	36	87
	LKHA4	Leukotriene A4 hydrolase	343	171	9,450	3CHP	18	4	82
	MMP13	Matrix metallo-proteinase 13	1632	572	37,200	830C	12	5	71

preparation (e.g. incomplete metal atom preparation, missing cofactors, or non-standard amino acids). A select few needed expert intervention to arrive at modest enrichment, such as adding crystallographic waters, changing histidine protonation, flipping ambiguous side-chains such as asparagine, or increasing a local dipole moment on a specific residue (a technique we often use prospectively to improve polar complementarity[43, 44]). In five targets we incorporated prior docking preparations used for prospective ligand discovery: adenosine A_{2A} receptor (AA2AR)[44], β_1 adrenergic receptor (ADRB1), AmpC β-lactamase (AMPC), C-X-C chemokine receptor type 4 (CXCR4)[3], and dopamine D_3 receptor (DRD3)[45].

To increase scaffold diversity and to make smaller, more manageable ligand sets, we clustered the raw ChEMBL ligands by their Bemis-Murcko frameworks[39]. These frameworks include ring systems of the molecule and connecting linkers, minus any side fragments. For example, the 7th largest Murcko cluster in kinesin-like protein 1 (KIF11) has seven ligands, all close analogs (**Figure 3.2.A**). If at least 100 frameworks are present, then we include only the highest affinity ligand from each framework. If fewer are available, we raise the number of ligands selected from each framework until we obtain more than 100 molecules, trading diversity for quantity. Returning to kinesin-like protein 1, we extracted only 70 Murcko frameworks (**Figure 3.2.B**). Out of 276 raw ligands, the five largest Murcko clusters contain 146 ligands (53%). Selecting the 2 or 3 highest affinity ligands from each framework results in 98 and 118 ligands, respectively, so we stop at 3 ligands per framework. In the process we still manage to remove 158 lower affinity compounds from highly redundant clusters. In a few targets more than 600 ligands remain even after clustering, so we reduce the affinity threshold below 1 μM, in the sequence (300, 100, 30, 10, and 3 nM), until fewer than 600 frameworks are found. For example, in adenosine A_{2A} receptor there are 3096 raw ligands

resulting in 1099 frameworks at 1 μM, but we can reduce the number of frameworks to 483

using a 30 nM affinity threshold (**Figure 3.2.C**).

Figure 3.2 Ligand Clustering

A) The 7th largest Murcko cluster of kinesin-like protein 1 (KIF11), showing both the scaffold (left) and all seven member ligands. B) Number of ligands in each of the 70 KIF11 Murcko frameworks. We remove lower affinity compounds over-represented clusters (above the green line), while retaining 100 ligands. C) Number of adenosine A_{2A} receptor (AA2AR) Murcko clusters is plotted against affinity threshold. Fewer than 600 clusters are present using a 30 nM affinity threshold.
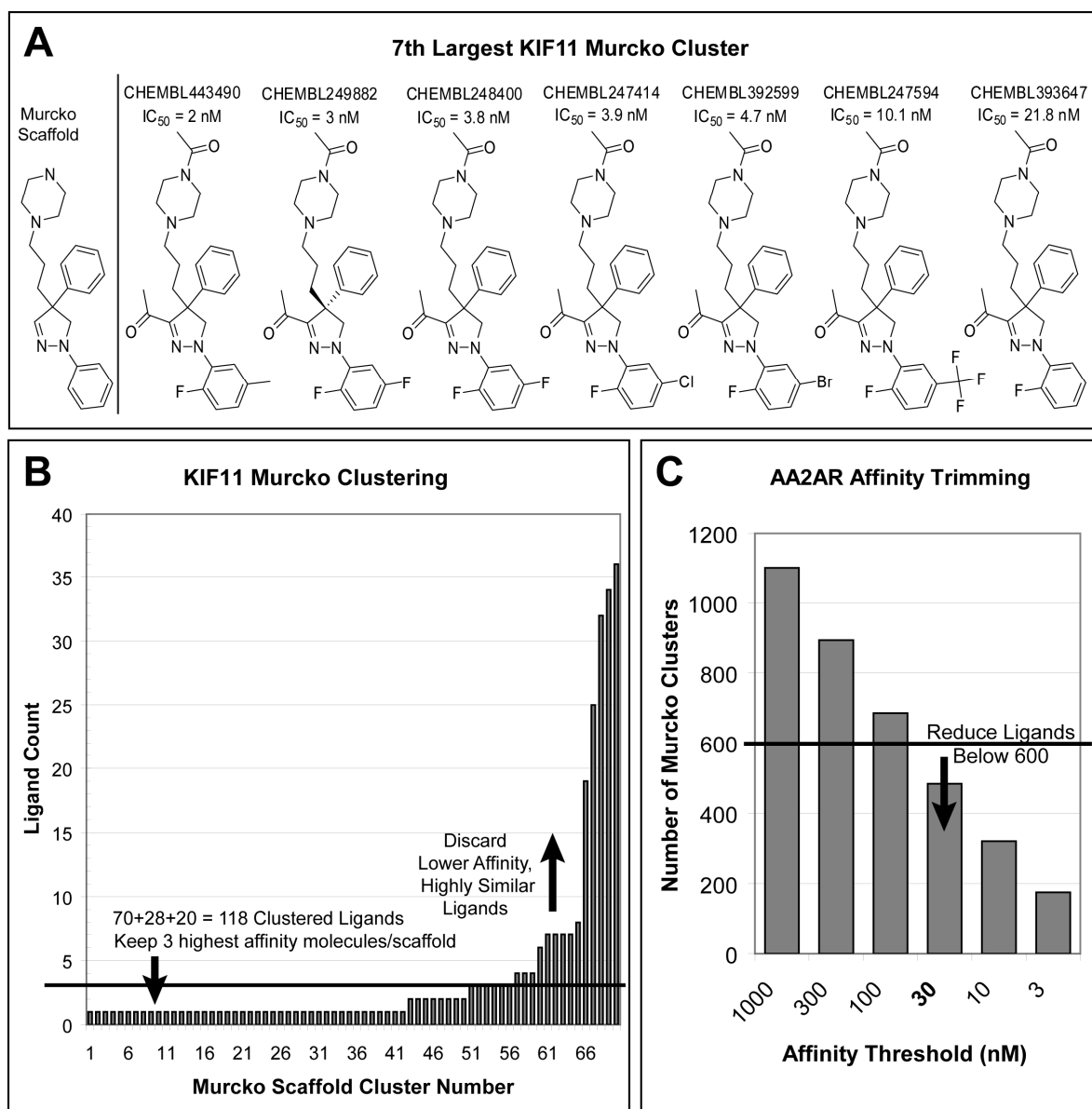
To examine the effect of clustering on docking enrichments, we docked the three targets with the highest and lowest fractions of clustered to raw ligands, of those with enough ligands to pick one ligand per Murcko cluster. To measure docking performance we use LogAUC, an aggregate metric that gives early enrichment more weight. As described previously[31], LogAUC is completely analogous to AUC, but in the transformed space after you have zoomed in on early enrichment by taking the semilog of the x-axis. In tryptase β1 (TRYB1), the target with the highest clustered fraction, clustering substantially decreases the LogAUC by 6%, whereas in the other five targets clustering increases the LogAUC (**Table A.3.2**). The mean absolute deviation over the six targets is 3.7% LogAUC, but in all cases the raw and clustered ROC curves have similar shapes (data not shown). Overall, we believe the clustered sets provide a better measure of docking performance with lower docking effort, and will be used in the remainder of this work.

A key problem with the original DUD decoys was that they sometimes closely resembled the ligands – occasionally even being confirmed as binders. Enforcing 2-D topological dissimilarity between decoys and ligands should eliminate this problem in principle, but in practice critical ligand binding "warheads" often remain in the decoy set, e.g. amidine groups in factor Xa (FA10). By identifying these warheads in three targets (**Figure 3.3.A**), we investigated how to eliminate false decoys. In the original DUD, CACTVS fingerprints were used to select decoys with Tanimoto coefficients (T_c) to ligands below 0.9, which is roughly similar to using Daylight fingerprints with T_c below 0.7[15]. In recent work[31] we used Daylight fingerprints with a more restrictive $T_c < 0.5$. Using this filter on the enhanced DUD ligand sets, we still see 39%, 53%, and 96% of possible warhead bearing molecules passing through in factor Xa (FA10), glycinamide ribonucleotide transformylase (PUR2), and thymidine kinase (KITH), respectively (**Figure 3.3.B**). Using Daylight with T_c

< 0.325, we reduce FA10 warheads below 1%, but still see 14% and 34% in PUR2 and

KITH. Clearly different targets and even different ligands require different absolute

thresholds. To circumvent this, we remove a percentage of the most similar decoys for each

ligand, sorted by maximum T_c to any ligand. This allows the effective absolute threshold to

Figure 3.3 Decoy Generation

A) Three key "warhead" groups from factor Xa (FA10), glycinamide ribonucleotide transformylase (PUR2), and thymidine kinase (KITH). B) Fraction of warheads remaining is plotted against the dissimilarity method. The dissimilarity methods consist of a fingerprint (Daylight or ECFP4) and either a hard cutoff or a fraction of the most dissimilar decoys to be retained. C) Property distributions of estrogen receptor α (ESR1) for both the 383 ligands (blue) and the 20,685 property-matched decoys (red).
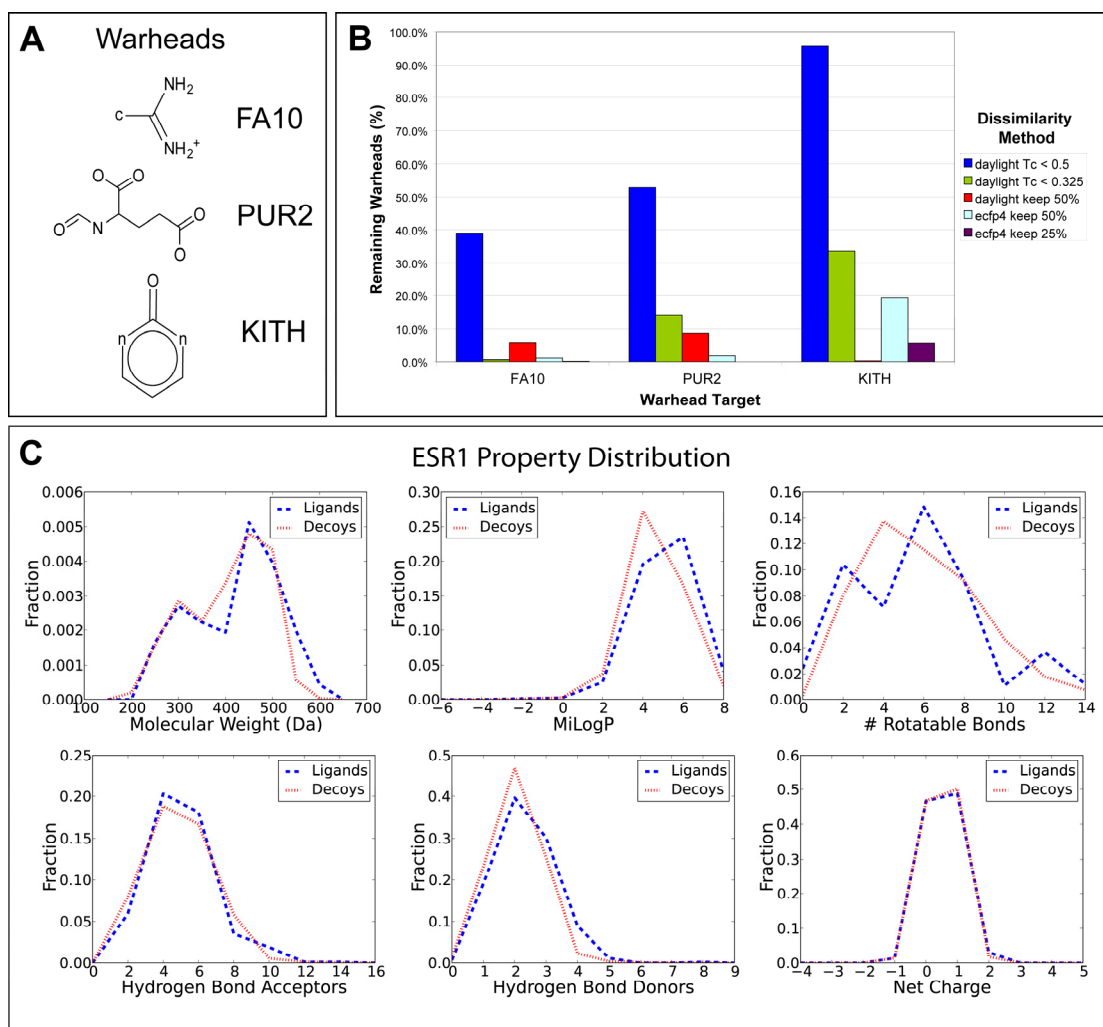
vary. Removing 50% of the decoys with Daylight was better in KITH, while removing 50% with ECFP4 was better in FA10 and PUR2. The final procedure of using ECFP4 fingerprints and removing 75% of the decoys, results in 0.2%, 0%, and 5.8% of warheads remaining, substantially reducing the number of false decoys. Having refined the decoy dissimilarity procedure on three targets where we could define a warhead, we then applied it to all generated decoys.

In addition to reducing false decoys, the DUD-E decoy generation procedure is extensively revised. Each decoy derives from a particular ligand, where decoy property ranges around the ligands properties adjust to seven possible widths. This adapts to local chemical space around each ligand, allowing more closely matched decoys. Also, net charge is added to the property matching, as it is a critical in electrostatics and desolvation. The improved property-matching can be seen in the property histograms for estrogen receptor α (ESR1) (**Figure 3.3.C**), as well as in the averages and standard deviations for all the targets (**Table A.3.3**). Using ZINC[46] for the potential decoy pool makes the decoys purchasable, enabling experimental testing for actual binding to the target. As a result of this work, this enhanced decoy procedure has been fully automated, and is available online to generate DUD-E style decoys for any user supplied list of input ligands at http://decoys.docking.org.

The original DUD paper[15] showed that a property-matched decoy set is more challenging for docking than a random collection of molecules. Therefore, we compared enrichments using property-matched decoys to those using a random drug-like background, which consisted of all ChEMBL12 ligands with affinities better than 10 μM. Switching from a drug-like background to DUD-E property-matched decoys does reduce average enrichment over the 102 targets, from 26.8% to 24.4% LogAUC (**Table A.3.4**). Yet for three targets, the property-matched sets unexpectedly led to much better enrichment, by

more than 15% LogAUC. In both glutamate receptor ionotropic kainate 1 (GRIK1) and purine nucleoside phosphorylase (PNPH), the ligands have low molecular weights (**Table A.3.3**) and thus score poorly against the generally larger ChEMBL12 molecules, just as Verdonk[17] suggests. In urokinase-type plasminogen activator (UROK), the top of the drug-like docking hit list is dominated by decoys with amidine "warheads". Since these are likely binders, the increased property-matched enrichment results from fewer false decoys. Indeed, the 2.4% LogAUC reduction that occurs upon switching to property-matched decoys arises from these two competing factors: property matching the decoys reduces enrichment, and reduction of false decoys increases enrichment.

Table 3.3 Decomposition of Enrichment Changes between DUD and DUD-E

Incremental Change	All Original	New Style Decoys	Switch Ligands	Switch Target Preparation
Decoys	DUD	DUD-E	DUD-E	DUD-E
Ligands	DUD	DUD	DUD-E	DUD-E
Receptor Preparation	DUD	DUD	DUD	DUD-E
Average LogAUC*	14.8	19.7	16.4	22.8

*Over the 37 common targets (details in **Table A.3.5**)

Overall, enrichment as measured by average LogAUC is 1.5 fold higher in DUD-E compared to the original DUD. To understand this, we first isolated the change due to the revised decoy generation procedure. Using the original DUD ligands and target preparations, but switching from original decoys to these revised decoys substantially increases the average enrichment over the 37 directly comparable targets from 14.8% to 19.7% LogAUC (**Table 3.3**, **Table A.3.5**). With the new adaptive property-matching procedure incorporating net

charge, the revised decoys might have been expected to lower enrichment, but instead we see an overall increase. Inspecting the docking hit lists, we observe a dramatic decrease in high scoring decoys that resemble ligands to a degree that they might actually bind. Indeed, all three targets with identifiable warheads that we used to tune the dissimilarity procedure show large increases in enrichment: FA10 increases from 13% to 28% LogAUC, PUR2 from 40% to 62% LogAUC, and KITH from 1% to 32% LogAUC. If we now isolate the switch from original ligands and revised decoys to both DUD-E ligands and decoys, we see a moderate decrease in average enrichment from 19.7% to 16.4% LogAUC. We attribute this decrease to the larger, more diverse clustered ligand lists in DUD-E. Lastly, switching the target preparation itself substantially increases enrichment from 16.4% to 22.8% LogAUC, which demonstrates the power of the large-scale docking campaign used to select representative, dockable structures. The overall effect is to dramatically increase average enrichment from 14.8% LogAUC in original DUD to 22.8% LogAUC in the fully revised DUD-E.

A central motivation for any benchmarking set is to test, at least retrospectively, new methods. We wanted to explore how our recent context-dependent ligand desolvation method[31] behaved against the DUD-E benchmark. We therefore used it to re-examine the utility of solvent-excluded volume (SEV) ligand desolvation versus using no desolvation term (None), or using the full transfer free energy from water to hexadecane (Full). In our initial study of these terms on the 40 original DUD targets, SEV improved upon None by just 0.7% average LogAUC. Conversely, over the 102 DUD-E targets, SEV substantially outperforms None by 3.8% LogAUC on average, with average LogAUC values of 20.6, 14.3, and 24.4% for None, Full, and SEV desolvation methods, respectively (**Figure 3.4, Table A.3.4**). Despite these average trends, ROC curves on individual targets can vary significantly

Figure 3.4 Desolvation Enrichment Comparisons over DUD-E

Docking results over DUD-E as measured by LogAUC. "None" has no ligand desolvation term, "SEV" uses solvent-excluded volume ligand desolvation, "Thin" employs a thin low-dielectric layer in the electrostatic calculations.
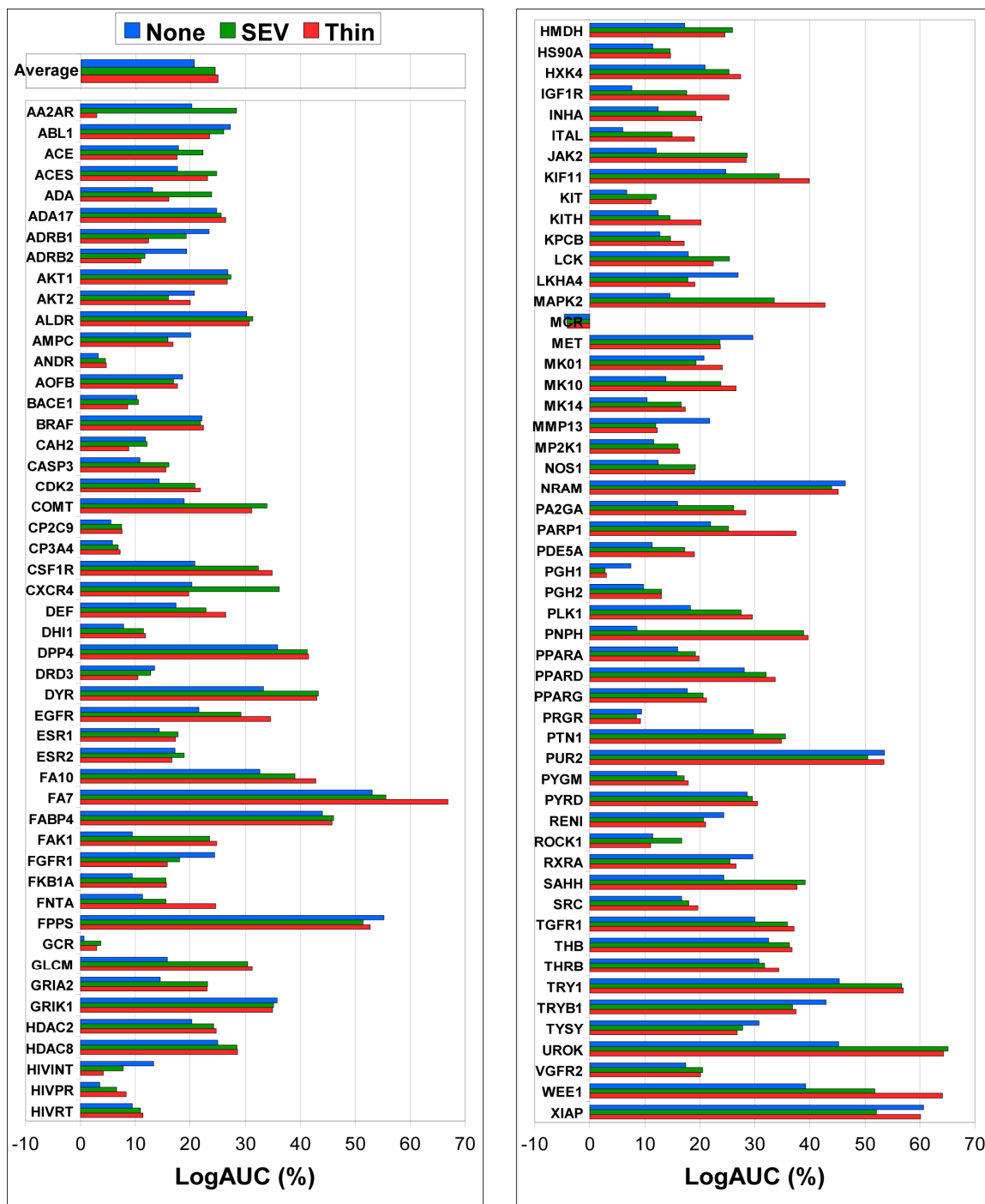
among the various methods (**Figure 3.5**). As in the original desolvation analysis, some targets are more amenable to full desolvation, such as catechol O-methyltransferase (COMT) and purine nucleoside phosphorylase (PNPH), while others are more amenable to no desolvation, such as factor X (FA10) and glycinamide ribonucleotide transformylase (PUR2). Against the DUD-E benchmark, SEV desolvation not only outperforms the other methods, but performs well in both types of targets. This suggests that over a more comprehensive set of targets, and what we argue is a better set of ligands and decoys, the advantage of the more physically correct SEV ligand desolvation treatment becomes more pronounced.

Electrostatic interaction with the protein is a large term that opposes ligand desolvation, with their relative balance being critical for binding. Since we do not know the binding pose of putative ligands prior to docking, we need to approximate the region of low dielectric the ligand might occupy to pre-compute electrostatic grids. Previously, we used the negative image of the receptor (computed by SPHGEN) to construct this low dielectric region, but manual tweaking was often required. In the large open binding pocket of CXCR4, we observed that using a thin layer of low-dielectric around just the edge of the protein allowed ligands to interact with it, while reducing the bulk dielectric perturbation at the center of its large binding pocket[3]. Here we explore using an automated thin dielectric layer strategy across the entire DUD-E set. Visually, these new automated thinner dielectric layers are more physically realistic, even in the rare case when they are effectively thicker than the previous layers (due to a water probe being able to penetrate the prior layer). With these thin low-dielectric layers (Thin), the average LogAUC over the 102 targets improves from 24.4% to 24.9% (**Figure 3.4, Table A.3.4**). Six targets use manually prepared dielectric layers (AA2A2, ADRB1, AMPC, CDK2, CXCR4, and DRD3), and thus do not directly reflect the difference between automated dielectric layers. Excluding those six enlarges the

Figure 3.5 Representative ROC Plots

ROC plots using no desolvation (None), solvent-excluded volume ligand desolvation (SEV), the thin low-dielectric layer (Thin), or a drug-like background that consists of all ChEMBL12 ligands with affinities better than 10 μM (Drug-like). The black dotted line represents the results expected from docking ligands randomly. LogAUC percentages are reported in the legend text.
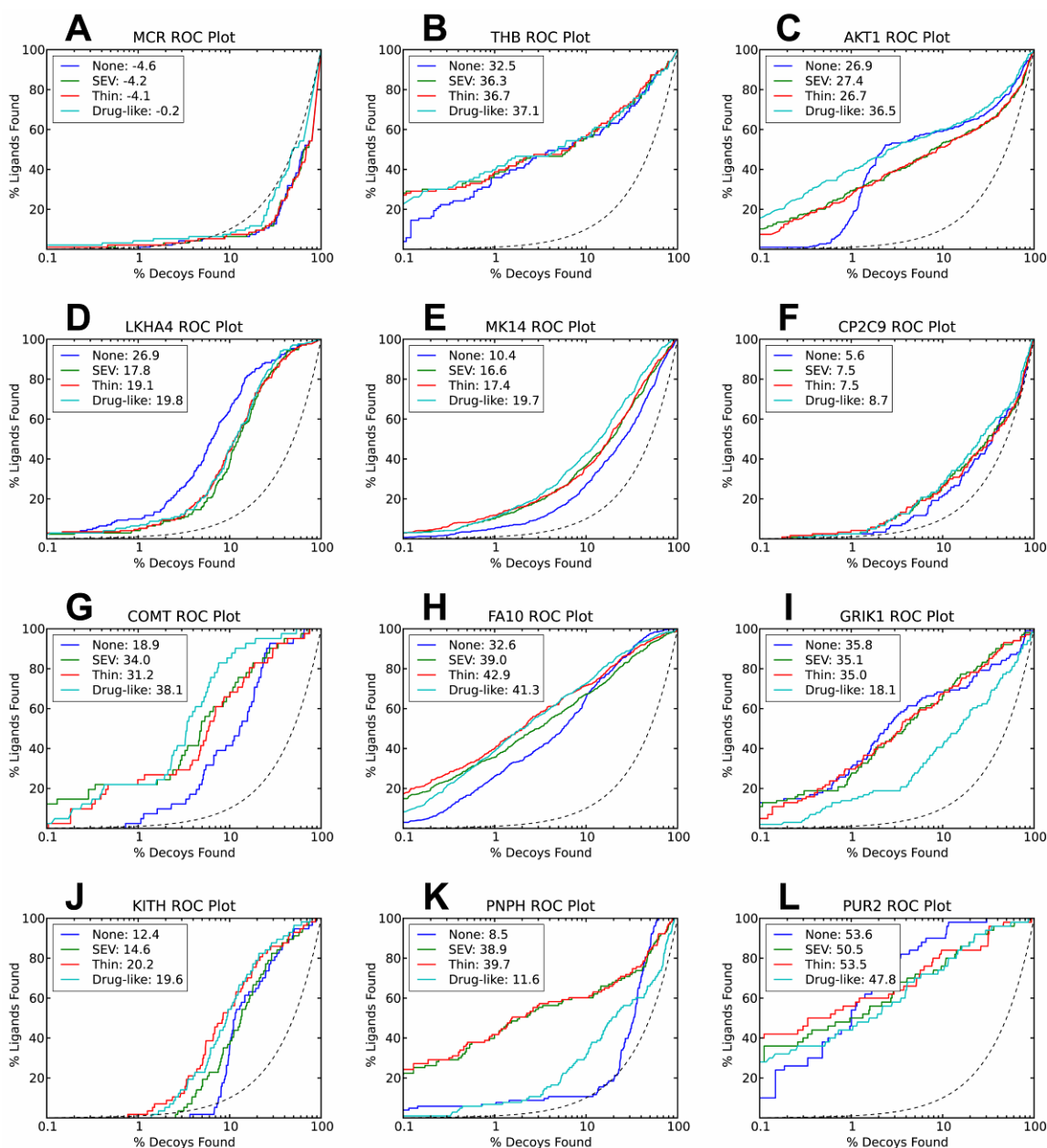
average difference from 0.5% to 1.0% LogAUC. Admittedly, these are moderate differences, but they exemplify how DUD-E may be used to test new docking methods, and hint that as we progress docking models, enrichment will improve.

Here we present three representative targets in greater detail to display a magnified view of DUD-E.

I. Mineralocorticoid Receptor (MCR)

MCR has the lowest enrichment in DUD-E. Across all 11 automatically docked structures, enrichment of DUD-E ligands to its decoys was negligible. Thus we selected the same PDB structure as the original DUD, 2AA2 at 1.95 Å resolution. While enrichment using the new DUD-E sets was worse than random at -4% LogAUC and 36% AUC (**Table 3.2**), using the original DUD ligands and decoys gave 45% LogAUC and 76% AUC. Despite poor enrichment in DUD-E, building and docking the crystal ligand from scratch, ignoring crystallographic information, results in good pose agreement (**Figure 3.6.A**). Taken together we can rationalize the enrichment differences, as 13 of 15 original ligands share a polycyclic scaffold with the well-docked crystal ligand, while the 94 new ligands have much more scaffold diversity. Thus the reduced enrichment in DUD-E reflects increased chemotype diversity, as a result of including more ligands and clustering them by Bemis-Murcko frameworks. Of the four lowest enriching targets in DUD-E, three are nuclear hormone receptors, with glucocorticoid receptor (GCR) and androgen receptor (ANDR) joining MCR. These receptors all have hydrophobic pockets with flexible binding site residues such as methionine and leucine, so that a single rigid receptor may be incapable of docking all of their ligands. Thus these targets may be good tests of flexible receptor docking methods.

II. Thyroid Hormone Receptor β1 (THB)

THB produced good enrichment when a structure with an open sub-pocket was selected. Enrichment for the 16 automatically docked structures varies significantly, ranging from 13% (1NQ0) to 37% LogAUC (1Q4X). The lower enriching structures have larger cavities near Arg320 (right side of **Figure 3.6.B**), opening to solvent in 1NQ0; the higher enriching structures have larger cavities at the other end of the binding site near Met420 (left side), opening to solvent in 1Q4X. We selected the automated preparation of 1Q4X despite its modest 2.80 Å resolution, because Thr273 is pushed away by the crystal ligand, making the left sub-pocket larger. Using SEV desolvation then yields enrichment statistics of 36% LogAUC, 79% AUC, and a ROC-based enrichment factor at 1% (EF_1) of 38 (**Table 3.2**). The re-docked crystal ligand has excellent pose agreement (**Figure 3.6.B**).

Figure 3.6 Representative Docking Poses

Crystallographic ligand rebuilt and docked. A-F) Crystal pose (magenta) is compared to docked pose (green). In C) more ligand conformations generated and re-docked pose shown (tan). Key hydrogen bonds shown (black dotted lines). Partially transparent protein surface colored by atom type.

| **A. MCR** | **B. THB** | **C. AKT1** |

| **D. LKHA4** | **E. MK14** | **F. CP2C9** |
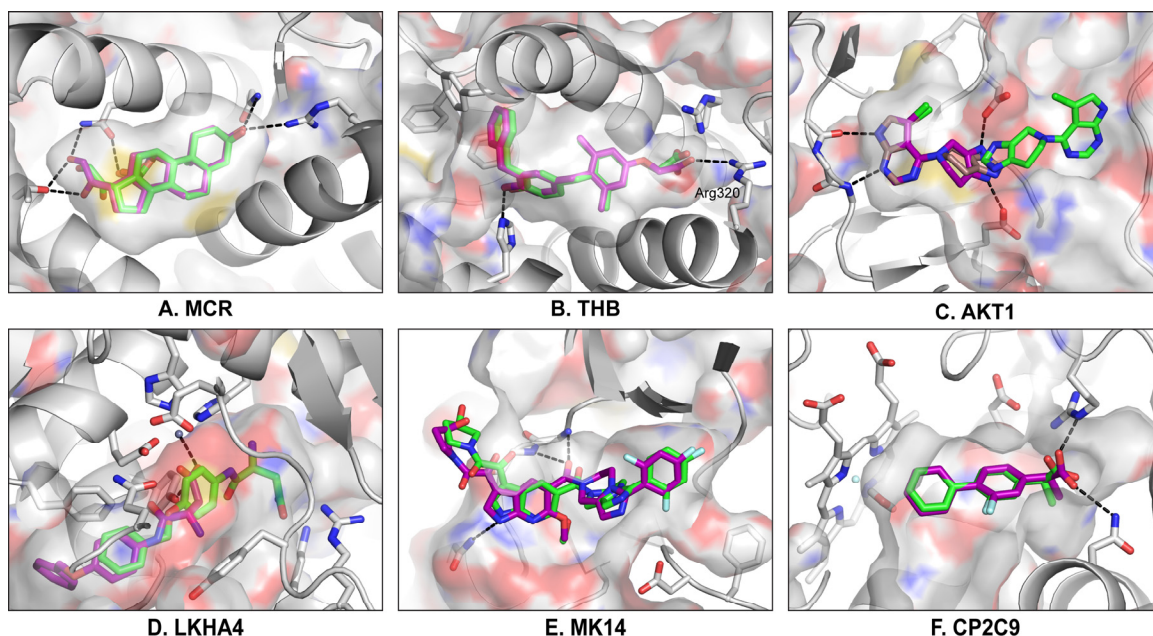
III. Serine/Threonine-Protein Kinase AKT (AKT1)

AKT1 is a newly added kinase that demonstrates several considerations during PDB structure selection. Whereas ten PDB structures were automatically docked, four got worse than random enrichment. All four correspond to structures of the Pleckstrin homology (PH) domain instead of the kinase domain. The structure with the best normal AUC, 3O96, corresponds to an allosteric site at the interface of the PH and kinase domains, not the traditional ATP binding pocket. While the best enriching structure by LogAUC, 3CQW at 2.00 Å, corresponds to the canonical site, its non-standard phospho-threonine amino acid evades the automated protocol. Preparing that residue manually results in 27% LogAUC, 72% AUC, and 29 EF_1 (**Table 3.2**). Nevertheless, the re-docked ligand (green) fails to generate the crystal ligand pose (magenta) (**Figure 3.6.C**). The ligand, however, is quite small with one central rotatable bond, and requires a specific rotation about that bond to fit in the binding site. Lowering the RMSD threshold for ligand conformation generation allows that rotation to be sampled, restoring the correct ligand binding pose (tan) (**Figure 3.6.C**).

3.4 Discussion

Docking remains a field that is judged by hit-rates prospectively, and enrichment retrospectively, because it cannot now hope to calculate affinities or even monotonic rank order. Like protein structure prediction, docking remains highly empirical, and yet we would argue also highly pragmatic. Its reliance on "enrichment" has driven the development of benchmarking sets, first explored by Rognan[11] and Jain[12], recently investigated by Boeckler[32] and Cavasotto[34]; the most widely used and cited of these remains the Directory of Useful Decoys (DUD) set[15]. Despite its widespread adoption DUD retains serious liabilities, including a lack of ligand diversity, lack of property-matching to net charge, and too many

false decoys. This enhanced DUD (DUD-E) was developed to address these shortcomings. By expanding the target list, DUD-E is substantially more representative of pharmacologically relevant space. By carefully linking ligands to decoys, and by releasing the methods online, we enable flexible use and even dynamic extension of DUD-E. We illustrate use of these benchmarking sets by examining our new ligand desolvation method, and finally we test a tweak to electrostatics to better balance this ligand desolvation.

Addressing major DUD shortcomings, we increase ligand diversity, improve decoy property-matching, and reduce false decoys that are likely to bind. When more than 100 ligands are available, DUD-E ligands are clustered by Murcko scaffold to reduce "analogue bias". With at least 40 ligands for every target, DUD-E provides robust enrichment metrics such as AUC and LogAUC. By property-matching decoys to each ligand individually while dynamically adapting to local chemical space, DUD-E decoy embedding is improved. Net charge, a key factor in ligand desolvation and electrostatics, is now property-matched. By selecting only the most dissimilar decoys for each ligand, the prevalence of false decoys is reduced.

We decomposed the changes between DUD-E and DUD, finding that DUD-E style decoys increase overall enrichment, new clustered ligands reduce enrichment, and new target preparations increase enrichment. Generating new decoys for the original DUD ligands increases average enrichment by 5% LogAUC, where we expected that the revised property-matching procedure would lower enrichment, as presented in the original DUD paper. However, to remove false decoys we now select only the 25% most dissimilar decoys, which results in this increased enrichment. This effect is much larger than we anticipated, indicating the original DUD decoys likely contain many ligand-like molecules. As expected, the larger, more diverse, clustered ligand sets reduce average enrichment by 3% LogAUC. The new

target preparations, carefully selected from a docking campaign to over 3500 structures, increase average enrichment by 6% LogAUC.

DUD-E is built to be a better platform for refinement and extension. Targets are independent of one another, both in ligand and decoy sets, allowing target addition, deletion, or replacement. The same protocol used to generate decoys for DUD-E, is available online to generate decoys for any target given only a list of ligand structures, which enables anyone to extend DUD-E to new targets of interest. The decoy server pulls directly from a purchasable subset of the ZINC database, inheriting its improvements and purchasing updates[46]. The final decoy selection from the applicable pool of decoys is random where possible, allowing the generation of multiple decoys sets to test over-fitting to the canonical DUD-E decoys. Each decoy belongs to one and only one ligand, so if one wants to filter a ligand, then the corresponding decoys can be easily removed. For example, we provide raw ligand and decoy sets before clustering by Murcko frameworks. If a different clustering method was desired, which selected a different subset of the raw ligands, then the corresponding decoys could be retained (furthermore we provide the python script used to generate clustered subsets from raw sets). We also include extra data that allows some design decisions to be altered, for instance we include the marginal ligands which are active above our 1 μM cutoff.

DUD-E should provide a more robust benchmarking set for exploring new docking methods, so we were keen to use it to explore the balance of ligand desolvation and electrostatics. Indeed, our new solvent-excluded volume (SEV) ligand desolvation algorithm shows a larger improvement over neglecting ligand desolvation using DUD-E, as compared to the original DUD. The larger increase when using the more physically realistic algorithm gives us further confidence in both the desolvation treatment and the new DUD-E dataset.

We finally use DUD-E to confirm a new hypothesis that a thin, more physically realistic dielectric layer, will help to better balance electrostatics with desolvation.

Certain caveats merit airing. Most importantly, DUD-E is a large dataset synthesized from several source databases, each of which is continuously evolving and improving. Thus individual errors are expected, though usually traceable to the source database at the time DUD-E was constructed. Also, some arbitrary choices and simplifying assumptions were made in the effort to provide one canonical dataset useful to compare docking algorithms. For instance, we assume a single PDB code can represent the target, but some targets are highly flexible, or they contain both orthosteric and allosteric binding pockets. Fundamentally, DUD and DUD-E are designed to measure value-added screening performance of 3-D methods over simple 1-D molecular properties. Decoys that might bind are removed using 2-D ligand similarity, so DUD-E is inappropriate to test 2-D methods. Through its construction, ligands light up against DUD-E decoys using these 2-D similarity methods, which create an artificially favorable enrichment bias for them. A final caution is that to filter more false decoys in DUD-E, we keep only a quarter of the most highly dissimilar decoys. However, while we show that this increased dissimilarity removes false decoys, it could also contribute to artificial increases in docking enrichment.

Notwithstanding these caveats, DUD-E is substantially improved over the original DUD. It is a larger, more diverse dataset with better matched decoys that resemble ligands less, correcting many flaws in its predecessor. Though we anticipate that it will be most widely used in the instantiation we describe here, it was developed with the idea that it could be flexibly extended and evolved; the tools to do so are even provided online (http://dude.docking.org). We hope that it and its descendants will provide a useful tool for

docking evaluation in the community, until such time as a more fundamental measurement of docking performance is possible.

3.5 Methods

I. ChEMBL and RCSB PDB data extraction

This enhanced DUD database has been constructed by combining ligand data from ChEMBL[38] and structural data from RCSB PDB[40]. Ligands assigned to protein targets (ChEMBL confidence score $\geq$ 4) with affinities (IC_{50}, EC_{50}, K_i, K_d and log variants thereof) of 1 µM or better were extracted from the ChEMBL09 database[38]. Similarly, we assigned experimental decoys as molecules with no measurable affinity at 30 µM or higher (">" relation only). The remaining ligands with affinities above 1 µM, and decoys with no measurable affinity below 30 µM, are included for completeness and designated "marginal". Via ChEMBL, ligands are associated with a particular target sequence by UniProt[41] accession code, and then mapped[47] from UniProt accession codes to protein data bank (PDB) structures (X-ray only) using http://www.uniprot.org/docs/pdbtosp.txt, obtained on February 23, 2011.

II. Target Selection Docking

Preliminary docking calculations were performed on each PDB structure that mapped to ChEMBL ligands and contained a single, unambiguous co-crystal ligand as prepared by DOCK Blaster[42]. Property-matched computational decoys were generated by the automated decoy generation procedure below, using Daylight fingerprints with a Tanimoto coefficient (T_c) threshold below 0.5. These decoys were docked and compared to their cognate ligands using DOCK 3.6 with solvent-excluded volume (SEV) ligand

desolvation[31]. Balancing the parallel goals of diversity, drug relevance, many ligands and structures, and at least modest automated docking enrichment, we selected 119 tentative targets for the new DUD. This list was reduced to the final 102 targets by factors such as ligand and PDB duplication between targets (e.g. FNTB duplicates FNTA), low resolution structures (RAF1), sterically constrained binding sites (NR1H2, THA), or over-representation (MK08, MTOR).

III. Target Preparation

For each target, we assembled all UniProt accession codes (species) with any raw ChEMBL compounds (ligands, decoys, marginal ligands, or marginal decoys). For only those accession codes, structures were extracted using the ChEMBL to PDB mapping, except P07700 was manually added to ADRB1 to include six more rare structures for that GPCR. This procedure neglects those PDB structures that belong to an accession code having no ChEMBL compounds. For example, 1KIM is the PDB structure of thymidine kinase (KITH) in the original DUD. This KITH structure is from herpes virus (UniProt P03176), an accession code with no raw compounds extracted from ChEMBL, and is thus not included in the ChEMBL/PDB intersection used to construct the new DUD. Still, 5025 PDB codes were sent to an updated DOCK Blaster pipeline for automated docking preparation. In some cases, an unambiguous ligand could not be found to indicate the binding site, but we were able to assign 565 additional ligands by manually inspecting over 1300 structures. Ultimately 3692 structures completed input grid preparation, and all but 2 finished docking and enrichment analysis. Clustered ligands sets were docked to property-matched decoys (both described below), using ECFP4 fingerprints and removing the most similar 75% of queried decoys. DOCK 3.6 was run using SEV ligand desolvation (as below).

For each target, enrichment, resolution, and organism were collected and sorted by enrichment in pdb_analyze.txt, available online at http://dude.docking.org. Crude notes on the selection process are recorded in pdb_selection.txt and the picked structure is listed in pdb_blessed.txt. AA2AR and DRD3 docking preparations were provided by Jens Carlson[44, 45], CXCR4 partially by Dahlia Weiss[3], ADRB1 by Peter Kolb (personal communication), and AMPC by Sarah Barelier, Oliv Eidam, and Inbar Fish (unpublished results).

IV. Ligand Preparation

To prepare ligand sets for each target, ChEMBL affinities and log variants were first normalized to nM units. Salts were removed, charges were normalized, and properties were calculated using Molinspiration's mib package (www.molinspiration.com). Ligands with 600 Da or higher molecular weight or with 20 or more rotatable bonds were removed. Smiles were put in canonical form using OpenEye's OEChem software[48]. Ligand sets from each species were combined, sorted by ascending normalized affinity, and then made unique based on canonical smiles. The same procedure was used to collate the experimental decoys, marginal ligands, and marginal decoys. For AmpC β-lactamase (AMPC), an original DUD target, the ChEMBL09 ligands are covalent in nature. To identify non-covalent ligands, we manually compiled ligands[6, 43, 49, 50] with affinities below 5 mM and experimental decoys[43, 51] from the literature.

V. Ligand Clustering

To reduce the sometimes large number of ChEMBL ligands down to a manageable size, while also increasing scaffold diversity as suggested by Good and Oprea[27], we clustered the ligands by their Bemis-Murcko frameworks[39], as generated by Molinspiration's mib. If there were 100 or more frameworks, we chose only the highest affinity ligand from each. If

there were fewer than 100 Murcko frameworks, we increased the number of highest affinity

ligands taken from each until we achieved at least 100 ligands (or until all ligands were

included). Conversely, if there were more than 600 Murcko frameworks, then we decreased

the ligand affinity threshold in the sequence [1 μM, 300 nM, 100 nM, 30 nM, 10 nM, 3 nM]

until fewer than 600 frameworks were present, where we then took the highest affinity ligand

from each framework. While clustered ligand sets are the default, the full unclustered ligand

sets and corresponding decoys are available. The script (subset_decoys.py) used to select the

clustered subset given the ligand ids is provided with the full ligand set to enable other

clustering algorithms or filtering methods to be substituted.

VI. Automated Decoy Generation

As in the original DUD, we property-matched decoys to ligands using molecular

weight, estimated water-octanol partition coefficient (miLogP), rotatable bonds, hydrogen

bond acceptors, and hydrogen bond donors, plus we added net charge. We generated all

ligand protonation states in pH range 6-8 using Schrödinger's Epik with arguments "-ph 7.0

-pht 1.0 -tp 0.20". Molecular properties were computed using Molinspiration's mib. For each

unique set of 6 properties, we aimed to generate 50 matched decoys. For example, a single

input ligand predicted to have 2 alternate charges would get 50 decoys property-matched to

each charge. Next a pool of decoys was selected from ZINC[46] using a dynamic protocol that

adapted to local chemical space by narrowing or widening windows in seven steps around

the 6 properties. The goal was to return 3000 to 9000 potential decoys that matched the

ligands's reference protonation state (predicted most prevalent form at pH 7.05). In the final

decoy procedure, ECFP4 fingerprints were generated by Scitegic's Pipeline Pilot for ligands

and potential decoys. The decoys were sorted by their maximum T_c to any ligand and the

most dissimilar 25% were retained through this dissimilarity filter. We then remove duplicate decoys between ligands by sorting decoys from least to most duplicated, and assigned each decoy to the ligand which has the least number of decoys already assigned. This ensures unique decoys were spread across the ligands as evenly as possible. Finally, if available, 50 decoys were picked randomly from this de-duplicated list.

VII. Original DUD Comparison

For the original DUD comparison, we downloaded ligands and decoys from dud.docking.org, and prepared docking flexibases with our modern ZINC toolchain[46]. The original DUD target preparations were copies of the original, modified to perform SEV desolvation calculations as described previously[31]. We also generated DUD-E style automated decoys and flexibases for the original DUD ligands. The analysis was performed on the 37 directly comparable targets, excluding the original targets PDGFrb, $ER_{agonist}$, and $ER_{antagonist}$.

VIII. Docking Calculations

Except as noted, docking calculations were performed with DOCK 3.6 and solvent-excluded volume (SEV) ligand desolvation as described previously[31]. Ligand conformations were generated by OpenEye's Omega[52]. For sampling, the minimum number of graph matching nodes was changed to 3, and ligand overlap was changed to 0.1. Ligands were limited to between 5 and 100 heavy atoms. The timeout for an individual ligand hierarchy was 180 s. We performed 200 steps of simplex minimization, with initial translations of 0.2 Å and initial rotations of 5°. The thin dielectric layer Delphi spheres were created by walking out each DMS (http://www.cgl.ucsf.edu/Overview/ftp/dms.zip) surface normal by 1.8 Å

and placing a sphere. This thin sphere layer is then used as input to makespheres1.pl in place of the usual SPHGEN spheres. The random background calculations were performed using SEV desolvation by seeding the DUD-E ligands into the entire ChEMBL12_10 subset of ZINC, which includes 273,375 ligands with annotated affinities below 10 μM.

IX. Docking Metrics

The area under the curve (AUC) of the receiver operating characteristic (ROC) is one common metric to measure docking performance. However, ROC plots often use a semi-log transformation of the x-axis to zoom in on early changes. As described previously[31], LogAUC is completely analogous to AUC in this transformed space, measuring the percentage of the unit area under the curve. Formally, we use the adjusted $LogAUC_{0.001}$ here, which spans three decades of log space and subtracts the LogAUC of the random curve (14.462%) so that random enrichment is 0%. We typically refer to the adjusted $LogAUC_{0.001}$ as either adjusted LogAUC or simply LogAUC. The ROC-based enrichment factor at 1% (EF_1) is the % of ligands found when 1% of the decoys have been found, and is preferred over traditional enrichment factors[53].

3.6 Acknowledgments

Pilot; and ChemAxon (Budapest, Hungary) for cxcalc. We thank Oliv Eidam, Matthew

Merski, and Nir London for reading this manuscript.

3.7 References

1. Kitchen, DB; Decornez, H; Furr, JR; Bajorath, J. "Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications" *Nat Rev Drug Discovery* 2004 3: 935-949.

2. Kolb, P; Rosenbaum, DM; Irwin, JJ; Fung, JJ; Kobilka, BK; Shoichet, BK. "Structure-Based Discovery of Beta(2)-Adrenergic Receptor Ligands" *Proc Natl Acad Sci U S A* 2009 106: 6843-6848.

3. Mysinger, MM; Weiss, DR; Ziarek, JJ; Gravel, S; Doak, AK; Karpiak, J; Heveker, N; Shoichet, BK; Volkman, BF. "Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor Cxcr4" *Proc Natl Acad Sci U S A* 2012 109: 5517-5522.

4. Gruneberg, S; Stubbs, MT; Klebe, G. "Successful Virtual Screening for Novel Inhibitors of Human Carbonic Anhydrase: Strategy and Experimental Confirmation" *J Med Chem* 2002 45: 3588-3602.

5. Jain, AN; Nicholls, A. "Recommendations for Evaluation of Computational Methods" *J Comput-Aided Mol Des* 2008 22: 133-139.

6. Babaoglu, K; Simeonov, A; Irwin, JJ; Nelson, ME; Feng, B; Thomas, CJ; Cancian, L; Costi, MP; Maltby, DA; Jadhav, A; Inglese, J; Austin, CP; Shoichet, BK. "Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against Beta-Lactamase" *J Med Chem* 2008 51: 2502-2511.

7. Ferreira, RS; Simeonov, A; Jadhav, A; Eidam, O; Mott, BT; Keiser, MJ; McKerrow, JH; Maloney, DJ; Irwin, JJ; Shoichet, BK. "Complementarity between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors" *J Med Chem* 2010 53: 4891-4905.

8. Gohlke, H; Klebe, G. "Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors" *Angew Chem, Int Ed Engl* 2002 41: 2644-2676.

9. Enyedy, IJ; Egan, WJ. "Can We Use Docking and Scoring for Hit-to-Lead Optimization?" *J Comput-Aided Mol Des* 2008 22: 161-168.

10. Stahl, M; Rarey, M. "Detailed Analysis of Scoring Functions for Virtual Screening" *J Med Chem* 2001 44: 1035-1042.

11. Bissantz, C; Folkers, G; Rognan, D. "Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations" *J Med Chem* 2000 43: 4759-4767.

12. Pham, TA; Jain, AN. "Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data" *J Med Chem* 2006 49: 5856-5868.

13. Kellenberger, E; Rodrigo, J; Muller, P; Rognan, D. "Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy" *Proteins* 2004 57: 225-242.

14. Ferrara, P; Gohlke, H; Price, DJ; Klebe, G; Brooks, CL, 3rd. "Assessing Scoring Functions for Protein-Ligand Interactions" *J Med Chem* 2004 47: 3032-3047.

15. Huang, N; Shoichet, BK; Irwin, JJ. "Benchmarking Sets for Molecular Docking" *J Med Chem* 2006 49: 6789-6801.

16. Christofferson, AJ; Huang, N. 13. How to Benchmark Methods for Structure-Based Virtual Screening of Large Compound Libraries. In *Computational Drug Discovery and Design (Methods in Molecular Biology)*, 2011/12/21 ed.; Baron, R., Ed. Springer Protocols: 2012; Vol. 819, pp 187-195.

17. Verdonk, ML; Berdini, V; Hartshorn, MJ; Mooij, WT; Murray, CW; Taylor, RD; Watson, P. "Virtual Screening Using Protein-Ligand Docking: Avoiding Artificial Enrichment" *J Chem Inf Comput Sci* 2004 44: 793-806.

18. Kuntz, ID; Chen, K; Sharp, KA; Kollman, PA. "The Maximal Affinity of Ligands" *Proc Natl Acad Sci U S A* 1999 96: 9997-10002.

19. Fan, H; Irwin, JJ; Webb, BM; Klebe, G; Shoichet, BK; Sali, A. "Molecular Docking Screens Using Comparative Models of Proteins" *J Chem Inf Model* 2009.

20. Repasky, MP; Murphy, RB; Banks, JL; Greenwood, JR; Tubert-Brohman, I; Bhat, S; Friesner, RA. "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* 2012.

21. Brozell, SR; Mukherjee, S; Balius, TE; Roe, DR; Case, DA; Rizzo, RC. "Evaluation of Dock 6 as a Pose Generation and Database Enrichment Tool" *J Comput-Aided Mol Des* 2012.

22. Neves, MA; Totrov, M; Abagyan, R. "Docking and Scoring with Icm: The Benchmarking Results and Strategies for Improvement" *J Comput-Aided Mol Des* 2012.

23. Spitzer, R; Jain, AN. "Surflex-Dock: Docking Benchmarks and Real-World Application" *J Comput-Aided Mol Des* 2012.

24. Schneider, N; Hindle, S; Lange, G; Klein, R; Albrecht, J; Briem, H; Beyer, K; Claussen, H; Gastreich, M; Lemmen, C; Rarey, M. "Substantial Improvements in Large-Scale Redocking and Screening Using the Novel Hyde Scoring Function" *J Comput-Aided Mol Des* 2012.

25. Liebeschuetz, JW; Cole, JC; Korb, O. "Pose Prediction and Virtual Screening Performance of Gold Scoring Functions in a Standardized Test" *J Comput-Aided Mol Des* 2012.

26. Novikov, FN; Stroylov, VS; Zeifman, AA; Stroganov, OV; Kulkov, V; Chilov, GG. "Lead Finder Docking and Virtual Screening Evaluation with Astex and Dud Test Sets" *J Comput-Aided Mol Des* 2012.

27. Good, AC; Oprea, TI. "Optimization of Camd Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection?" *J Comput-Aided Mol Des* 2008 22: 169-178.

28. Mackey, MD; Melville, JL. "Better Than Random? The Chemotype Enrichment Problem" *J Chem Inf Model* 2009 49: 1154-1162.

29. Hawkins, PC; Warren, GL; Skillman, AG; Nicholls, A. "How to Do an Evaluation: Pitfalls and Traps" *J Comput-Aided Mol Des* 2008 22: 179-190.

30. Irwin, JJ. "Community Benchmarks for Virtual Screening" *J Comput-Aided Mol Des* 2008 22: 193-199.

31. Mysinger, MM; Shoichet, BK. "Rapid Context-Dependent Ligand Desolvation in Molecular Docking" *J Chem Inf Model* 2010 50: 1561-1573.

32. Vogel, SM; Bauer, MR; Boeckler, FM. "Dekois: Demanding Evaluation Kits for Objective in Silico Screening--a Versatile Tool for Benchmarking Docking Programs and Scoring Functions" *J Chem Inf Model* 2011 51: 2650-2665.

33. Wallach, I; Lilien, R. "Virtual Decoy Sets for Molecular Docking Benchmarks" *J Chem Inf Model* 2011 51: 196-202.

34. Gatica, EA; Cavasotto, CN. "Ligand and Decoy Sets for Docking to G Protein-Coupled Receptors" *J Chem Inf Model* 2012 52: 1-6.

35. Cereto-Massague, A; Guasch, L; Valls, C; Mulero, M; Pujadas, G; Garcia-Vallve, S. "Decoyfinder: An Easy-to-Use Python Gui Application for Building Target-Specific Decoy Sets" *Bioinformatics* 2012.

36. Rohrer, SG; Baumann, K. "Maximum Unbiased Validation (Muv) Data Sets for Virtual Screening Based on Pubchem Bioactivity Data" *J Chem Inf Model* 2009 49: 169-184.

37. Ripphausen, P; Wassermann, AM; Bajorath, J. "Reprovis-Db: A Benchmark System for Ligand-Based Virtual Screening Derived from Reproducible Prospective Applications" *J Chem Inf Model* 2011 51: 2467-2473.

38. Gaulton, A; Bellis, LJ; Bento, AP; Chambers, J; Davies, M; Hersey, A; Light, Y; McGlinchey, S; Michalovich, D; Al-Lazikani, B; Overington, JP. "Chembl: A Large-Scale Bioactivity Database for Drug Discovery" *Nucleic Acids Res* 2012 40: D1100-1107.

39. Bemis, GW; Murcko, MA. "The Properties of Known Drugs. 1. Molecular Frameworks" *J Med Chem* 1996 39: 2887-2893.

40. Berman, HM; Westbrook, J; Feng, Z; Gilliland, G; Bhat, TN; Weissig, H; Shindyalov, IN; Bourne, PE. "The Protein Data Bank" *Nucleic Acids Res* 2000 28: 235-242.

41. Apweiler, R; Bairoch, A; Wu, CH; Barker, WC; Boeckmann, B; Ferro, S; Gasteiger, E; Huang, H; Lopez, R; Magrane, M; Martin, MJ; Natale, DA; O'Donovan, C; Redaschi, N; Yeh, LS. "Uniprot: The Universal Protein Knowledgebase" *Nucleic Acids Res* 2004 32: D115-119.

42. Irwin, JJ; Shoichet, BK; Mysinger, MM; Huang, N; Colizzi, F; Wassam, P; Cao, Y. "Automated Docking Screens: A Feasibility Study" *J Med Chem* 2009 52: 5712-5720.

43. Powers, RA; Morandi, F; Shoichet, BK. "Structure-Based Discovery of a Novel, Noncovalent Inhibitor of Ampc Beta-Lactamase" *Structure* 2002 10: 1013-1023.

44. Carlsson, J; Yoo, L; Gao, ZG; Irwin, JJ; Shoichet, BK; Jacobson, KA. "Structure-Based Discovery of A2a Adenosine Receptor Ligands" *J Med Chem* 2010 53: 3748-3755.

45. Carlsson, J; Coleman, RG; Setola, V; Irwin, JJ; Fan, H; Schlessinger, A; Sali, A; Roth, BL; Shoichet, BK. "Ligand Discovery from a Dopamine D3 Receptor Homology Model and Crystal Structure" *Nat Chem Biol* 2011 7: 769-778.

46. Irwin, JJ; Shoichet, BK. "Zinc--a Free Database of Commercially Available Compounds for Virtual Screening" *J Chem Inf Model* 2005 45: 177-182.

47. Velankar, S; McNeil, P; Mittard-Runte, V; Suarez, A; Barrell, D; Apweiler, R; Henrick, K. "E-Msd: An Integrated Data Resource for Bioinformatics" *Nucleic Acids Res* 2005 33: D262-265.

48. Hawkins, PC; Skillman, AG; Nicholls, A. "Comparison of Shape-Matching and Docking as Virtual Screening Tools" *J Med Chem* 2007 50: 74-82.

49. Teotico, DG; Babaoglu, K; Rocklin, GJ; Ferreira, RS; Giannetti, AM; Shoichet, BK. "Docking for Fragment Inhibitors of Ampc Beta-Lactamase" *Proc Natl Acad Sci U S A* 2009 106: 7455-7460.

50. Tondi, D; Morandi, F; Bonnet, R; Costi, MP; Shoichet, BK. "Structure-Based Optimization of a Non-Beta-Lactam Lead Results in Inhibitors That Do Not up-Regulate Beta-Lactamase Expression in Cell Culture" *J Am Chem Soc* 2005 127: 4632-4639.

51. Graves, AP; Brenk, R; Shoichet, BK. "Decoys for Docking" *J Med Chem* 2005 48: 3714-3728.

52. Hawkins, PC; Skillman, AG; Warren, GL; Ellingson, BA; Stahl, MT. "Conformer Generation with Omega: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database" *J Chem Inf Model* 2010 50: 572-584.

53. Jain, AN. "Bias, Reporting, and Sharing: Computational Evaluations of Docking Methods" *J Comput-Aided Mol Des* 2008 22: 201-212.

Chapter 4:

Future Perspectives

Science is always wrong. It never solves a problem without creating ten more.

-- George Bernard Shaw

By its very nature, the process of answering scientific questions raises new thoughts and leads to more hypotheses. Here I attempt to capture some of my musings pertaining to DOCK 3.6 itself, the balance of desolvation and electrostatics, and some possible ways to approach receptor desolvation.

4.1 A DOCK State of Mind

With solvent-excluded volume (SEV) based ligand desolvation, I believe the scoring function of DOCK is fast and yet surprisingly useful. The two pathologies we now experience most are ligand internal clashes and stranded ligand polar groups. Adding an internal energy scoring term would remove the ligand clashes we observe; yet incorporating this term permanently into DOCK will require changes to the conformational hierarchy. As they are large anti-correlated forces, treatment of ligand internal energy should simultaneously include desolvation to have the largest impact, for example in the global energy minimum search. Ironically, I believe that incorporating and balancing receptor desolvation will likely help with the stranded polar groups of the ligands. Any residual tendency to stand polar groups after incorporating receptor desolvation may be due to double counting ligand dispersion forces, by also including them in the non-polar ligand

desolvation term. Other scoring terms, especially entropy terms, are on the critical path towards enabling molecular docking to someday predict ligand binding affinities. However, I think the areas where DOCK fails completely today do not correspond to scoring deficiencies, but instead to inadequate sampling of the local environment.

In creating DUD-E, I observed that the targets that were most recalcitrant to DOCK were typically either highly flexible proteins, or those that involved many water mediated interactions. While the lab has studied both of these phenomena academically[1, 2], those methods have not been automated and efficiently incorporated into mainline DOCK. I believe the ability to address them routinely in DOCK is both worthwhile and sorely needed. To treat these effects in the past, we used multiple scoring grids, assumed to be independent. For receptor flexibility I think this is appropriate, but a method to automatically generate the relevant conformations is needed, taking as input an ensemble of receptor structures. However, I think water mediated interactions are a more natural fit to the ligand conformation ensemble. The conformational hierarchy could be extended past the physical ligand itself, into optional water shells beyond. With internal energy implemented, you could even incorporate estimates of the energy differences due to adding or removing the optional water molecules.

4.2 Balance of Ligand Desolvation and Electrostatics

The balance of ligand desolvation and electrostatics has been critical in getting ligand desolvation to perform well. To date, however, my efforts have focused more on the ligand desolvation side. While SEV based ligand desolvation is near the right magnitude for balance, there are qualitative indications that it slightly over weights desolvation relative to electrostatics. This is surprising since we neglect even more solvent-excluded volume that

arises from neighboring ligand atoms. Thus I think the solution lies not on the ligand desolvation side, but on the electrostatics side.

My hypothesis is that incorporating the SEV approximation into DelPhi, or another Poisson-Boltzman solver, will help docking in two ways. First, it will allow us to treat the dielectric boundary identically in the electrostatics and desolvation calculations, which will help the balance of electrostatics and desolvation. Second, it will reduce or eliminate our dependence on the kludge of dielectric spheres, which effectively require us to know where the ligands will bind before docking begins. Though this would be technically challenging, the optimized sevsolv code shows that it is possible to do the repeated surface generation and numerical integration fast enough.

There are also a few things I believe would further improve our ligand desolvation term. First, we could help the balance with electrostatics by using more of the same parameters. While I modified the receptor radii of ligand desolvation to better match electrostatics, this was just a poor compromise because DelPhi has more atom types. Ideally sevsolv would just read the DelPhi parameter file and use its atom types directly. Second, in chapter one I explored the use of separate hydrogen and heavy atom grids for SEV desolvation. While this helped the overall average LogAUC by 0.8% in DUD, it has no effect (< 0.1%) on the overall enrichment in DUD-E. However, looking at the target-by-target breakdown, I see that even in DUD-E the targets with positively charged ligands get substantially better (5% LogAUC range), while the targets with negatively charge ligands get substantially worse. I hypothesize that the reduced performance from the negative targets arises from increasing the heavy atom radius from 1.4 to 1.8 Å. We might be able keep the improvements for positive targets by leaving the hydrogen atom radius at 1.0 Å. , while hurting the negative targets less by reducing the heavy atom radius closer to 1.4 Å..

Additionally, I think this sensitivity to increases in heavy atom radius is related to the fact that SEV desolvation already tends to be over-weighted relative to electrostatics.

4.3 Adding Receptor Desolvation

In the limit where ligands of a similar size bind to the exact same protein binding site, receptor desolvation differences between ligands are small. However, receptors typically accommodate a distribution of ligand sizes, and in many actual drug targets they contact multiple independent side pockets. Therefore receptor desolvation may be worth pursuing, so here I consider two alternate strategies to account for receptor desolvation: 1) flip our current ligand desolvation term around, or 2) use the novel, but unproven method, of Solvation Energy Density Occlusion (SEDO).

In the first approach, we keep our ligand desolvation term and complement it with a similar scheme for the receptor. Due to the size of the protein, we would need to replace the semi-empirical AMSOL calculations with a pairwise-decomposable MM-GB/SA computation in order to get the initial polar and non-polar atomic contributions to the receptor's desolvation energy. Then we could use the same volume-based approximation to scale the contributions as in chapter one, but now the ligand volume elements desolvate the receptor atoms. The receptor desolvation would then be given by:

$$\Delta G_{desolv}^{R} = \sum_{k}^{in\_lig,not\_rec} \left[\frac{dV}{4\pi} \sum_{j}^{rec} \frac{a_j \Delta G_j^R}{r_{jk}^4} \right] \quad \textbf{(Equation 1)}$$

where dV is the volume of a grid element, ΔG_j^R is atomic contribution of receptor atom j, a_j is the receptor atomic radii, and r_{jk} is distance between receptor atom j and ligand volume element k. The part in brackets would be pre-calculated and stored on a grid. During

docking we would determine the volume elements of the ligand (and the excluded volume) and sum their desolvation contributions off the receptor desolvation grid. This approach is symmetrical to our ligand desolvation term, and could potentially be speed up by replacing the volume element summation with analytical pairwise approximations similar to those used in GB methods[3]. Another advantage of this approach is that it simply scales the atomic contributions ΔG_j^R, so they could be enhanced with more sophisticated estimates (like explicit solvent) in the future.

In the second approach, we could add a receptor desolvation term based on the Solvation Energy Density Occlusion (SEDO) method[4]. Here we write the electrostatic component of the desolvation free energy without approximation as (**Equation 2**), where S is the solvation energy density, given by (**Equation 3**).

$$\Delta G_{pol} = \int_{ex} S(\vec{x})dV \qquad \textbf{(Equation 2)}$$

$$S = \frac{1}{8\pi}\left(\mathbf{E}_{wat} \cdot \mathbf{D}_{hex} - \mathbf{E}_{hex} \cdot \mathbf{D}_{wat}\right) \qquad \textbf{(Equation 3)}$$

If we consider the approach of a second molecule, its effect on the first causes S to go to zero in the second molecule's interior, as well as perturbing S around the second molecule. In SEDO we keep the first solvent occlusion effect while neglecting the second perturbation effect, giving (**Equation 4**) below:

$$\Delta G_{desolv}^R = -\int_{lig} S_{rec}(\vec{x})dV \qquad \textbf{(Equation 4)}$$

We can thus use the Poisson-Boltzman solver DelPhi to pre-compute S once for the receptor, and perform the volume integration during docking to estimate receptor

desolvation. The advantage of this approach is that it is conceptually simple, if indeed it can

still be accurate enough after neglecting the perturbation of the second molecule.

4.4 References

1. Wei, BQ; Weaver, LH; Ferrari, AM; Matthews, BW; Shoichet, BK. "Testing a Flexible-Receptor Docking Algorithm in a Model Binding Site" J Mol Biol 2004 337: 1161-1182.

2. Huang, N; Shoichet, BK. "Exploiting Ordered Waters in Molecular Docking" J Med Chem 2008 51: 4862-4865.

3. Hawkins, GD; Cramer, CJ; Truhlar, DG. "Pairwise Solute Descreening of Solute Charges from a Dielectric Medium" Chem Phys Lett 1995 246: 122-129.

4. Arora, N; Bashford, D. "Solvation Energy Density Occlusion Approximation for Evaluation of Desolvation Penalties in Biomolecular Interactions" Proteins 2001 43: 12-27.

Appendix A:

Supporting Information

A.1 Chapter 1 Supporting Information

Table A.1.1 DUD Matched Decoys – Adjusted LogAUC versus AUC

Adjusted LogAUC versus AUC compared against matched decoys over the 40 DUD targets. We also provide the ROC-derived enrichment factors (EF) at 1 and 10 %. "None" contains no ligand desolvation term, "SEV" uses a single context-dependent ligand desolvation map, "Full" uses the full ligand transfer free energy, and "Hsolv" uses two separate maps for hydrogen and heavy atoms.

Method	None				SEV			
Metric	LogAUC	AUC	EF1	EF10	LogAUC	AUC	EF1	EF10
Average	*14.3*	*68.8*	*8.6*	*3.3*	*15.0*	*68.7*	*11.6*	*3.2*
ace	7.9	59.7	4.1	3.1	11.1	59.5	12.2	2.9
ache	6.1	67.4	0.0	1.2	8.5	68.4	0.0	2.6
ada	10.4	66.0	4.4	2.6	2.3	59.2	0.0	0.4
alr2	10.7	63.3	4.4	3.5	13.4	65.0	4.4	3.5
ampc	0.5	58.7	0.0	0.5	6.7	66.9	0.0	1.0
ar	7.1	58.9	5.4	2.3	6.3	58.9	4.1	1.9
cdk2	26.7	78.7	20.0	4.8	25.7	78.0	22.0	5.2
comt	14.9	83.3	0.0	2.7	36.8	92.0	27.3	7.3
cox1	9.2	59.7	8.3	2.5	10.1	58.0	12.5	2.5
cox2	28.6	81.6	28.4	5.8	32.1	81.2	29.7	5.3
dhfr	38.3	88.4	24.4	8.3	32.1	83.7	26.9	6.1
egfr	13.8	72.9	6.1	3.5	28.3	85.6	21.7	5.4
er_agonist	10.8	66.0	3.0	3.0	7.7	63.9	0.0	2.4
er_antagonist	18.0	69.2	18.0	3.1	22.1	70.7	23.1	3.6
fgfr1	-1.5	41.7	0.9	1.2	-3.4	47.1	0.0	0.3
fxa	19.6	80.3	12.1	3.9	16.5	78.4	7.1	3.4
gart	38.8	89.4	19.1	7.1	48.6	90.5	66.7	7.6
gpb	5.4	65.3	0.0	1.2	4.3	63.2	0.0	1.0
gr	4.6	53.6	7.7	1.2	3.5	54.3	3.9	1.8

Method	None				SEV			
Metric	LogAUC	AUC	EF1	EF10	LogAUC	AUC	EF1	EF10
hivpr	5.4	52.0	3.8	2.6	11.6	56.2	13.2	2.8
hivrt	6.0	59.9	2.6	1.8	8.1	60.5	7.7	2.3
hmga	3.5	50.7	0.0	2.3	3.5	49.2	0.0	2.0
hsp90	18.1	82.2	4.4	3.5	14.5	71.4	17.4	2.2
inha	15.7	73.9	7.1	3.4	7.1	67.3	0.0	2.2
mr	42.0	72.0	53.3	6.0	40.0	72.0	46.7	6.0
neua	35.5	90.8	12.2	7.1	27.3	75.3	18.4	6.5
p38	12.6	69.8	4.3	3.3	14.2	73.4	4.7	4.0
parp	30.5	88.9	12.1	7.9	28.2	86.0	12.1	6.7
pde5	12.4	65.5	16.0	3.0	17.6	72.1	16.0	2.8
pdgfrb	1.4	55.0	1.3	1.2	3.0	56.7	2.6	1.3
pnp	7.1	62.1	4.0	2.0	10.3	62.7	12.0	2.0
ppar	0.2	55.4	0.0	0.6	1.2	57.4	0.0	0.6
pr	-0.9	50.0	0.0	0.7	-2.0	49.7	0.0	0.4
rxr	33.2	80.1	30.0	6.5	34.3	77.2	40.0	4.5
sahh	14.2	80.1	0.0	3.8	16.1	81.8	0.0	4.4
src	2.9	54.1	3.2	1.1	2.9	59.0	0.7	1.4
thrombin	15.7	75.6	6.2	4.0	13.2	68.7	6.2	2.8
tk	15.4	77.4	4.6	3.2	10.2	75.3	0.0	2.7
trypsin	20.9	87.0	6.8	2.7	16.2	80.3	2.3	4.1
vegfr2	8.6	65.6	5.5	2.3	9.8	71.8	2.7	2.2

Method	Full				Hsolv			
Metric	LogAUC	AUC	EF1	EF10	LogAUC	AUC	EF1	EF10
Average	*10.9*	*66.6*	*6.2*	*2.6*	*15.8*	*68.4*	*13.2*	*3.3*
ace	6.4	56.4	6.1	2.0	11.5	59.8	12.2	2.9
ache	4.3	64.4	0.0	1.1	10.9	70.7	1.0	3.2
ada	2.0	58.4	0.0	0.4	3.1	59.9	0.0	1.3
alr2	14.4	71.5	4.4	3.9	13.6	67.1	4.4	3.5
ampc	1.3	58.3	0.0	0.0	7.5	70.1	0.0	1.4
ar	7.9	60.2	6.8	3.0	7.1	59.5	4.1	2.4
cdk2	18.5	70.7	12.0	4.0	26.2	77.1	24.0	4.6
comt	22.8	79.6	9.1	4.6	39.8	92.9	36.4	8.2
cox1	5.0	59.0	0.0	2.9	10.3	57.4	16.7	2.5
cox2	30.6	81.2	29.0	5.5	32.3	81.1	29.4	5.4

Method	Full				Hsolv			
Metric	LogAUC	AUC	EF1	EF10	LogAUC	AUC	EF1	EF10
dhfr	21.9	80.6	7.0	5.7	47.5	86.6	50.8	7.2
egfr	23.7	79.2	14.8	5.1	28.4	83.1	24.2	5.2
er_agonist	9.4	67.0	1.5	2.5	7.7	64.0	0.0	2.2
er_antagonist	11.5	72.8	5.1	1.8	22.6	74.0	23.1	3.6
fgfr1	1.8	61.3	0.0	0.7	-7.1	34.4	0.0	0.1
fxa	9.9	70.3	1.4	2.5	19.5	78.1	13.5	3.8
gart	10.1	77.0	0.0	1.0	44.8	89.6	47.6	7.6
gpb	3.3	62.7	0.0	1.0	3.7	62.1	0.0	0.8
gr	2.2	54.5	1.3	1.8	3.4	54.4	5.1	1.5
hivpr	10.3	59.6	13.2	2.3	10.7	52.9	15.1	2.5
hivrt	7.1	60.1	5.1	1.8	7.1	59.7	7.7	1.8
hmga	7.5	61.6	2.9	2.6	2.8	49.1	0.0	1.7
hsp90	11.4	70.1	4.4	2.6	13.1	69.4	13.0	2.2
inha	3.9	59.1	0.0	1.8	17.5	74.4	11.8	3.8
mr	34.1	70.9	33.3	6.0	41.8	72.5	53.3	6.0
neua	-4.0	38.6	0.0	1.0	35.0	75.3	30.6	6.9
p38	16.7	74.5	9.8	3.7	14.7	71.2	6.6	3.9
parp	16.4	75.9	6.1	3.9	25.7	86.1	9.1	7.0
pde5	10.2	67.0	6.0	3.0	18.0	70.8	12.0	2.8
pdgfrb	3.6	58.3	2.6	1.2	2.8	56.4	1.9	1.2
pnp	10.5	63.5	8.0	2.0	10.0	62.6	16.0	2.0
ppar	0.8	56.3	0.0	0.7	0.9	57.0	0.0	0.6
pr	0.9	56.0	0.0	0.4	-1.2	51.6	0.0	0.4
rxr	35.6	78.2	45.0	6.5	34.0	77.4	40.0	5.5
sahh	17.0	82.2	0.0	5.0	15.6	81.3	0.0	4.7
src	3.6	63.0	0.7	1.0	0.7	51.7	1.3	1.3
thrombin	12.1	67.0	6.2	2.5	14.9	71.0	12.3	3.4
tk	8.7	74.2	0.0	1.4	10.1	75.4	0.0	2.3
trypsin	14.2	76.0	4.6	3.0	15.8	80.9	2.3	3.4
vegfr2	8.9	67.9	2.7	2.1	9.1	65.9	4.1	2.1

Figure A.1.2 Using Separate Hydrogen and Heavy Atom Desolvation Grids

Graph of adjusted LogAUC against matched decoys, showing the effect of using SEV desolvation with separate desolvation maps for hydrogen and heavy atoms (hydrogen-sev).
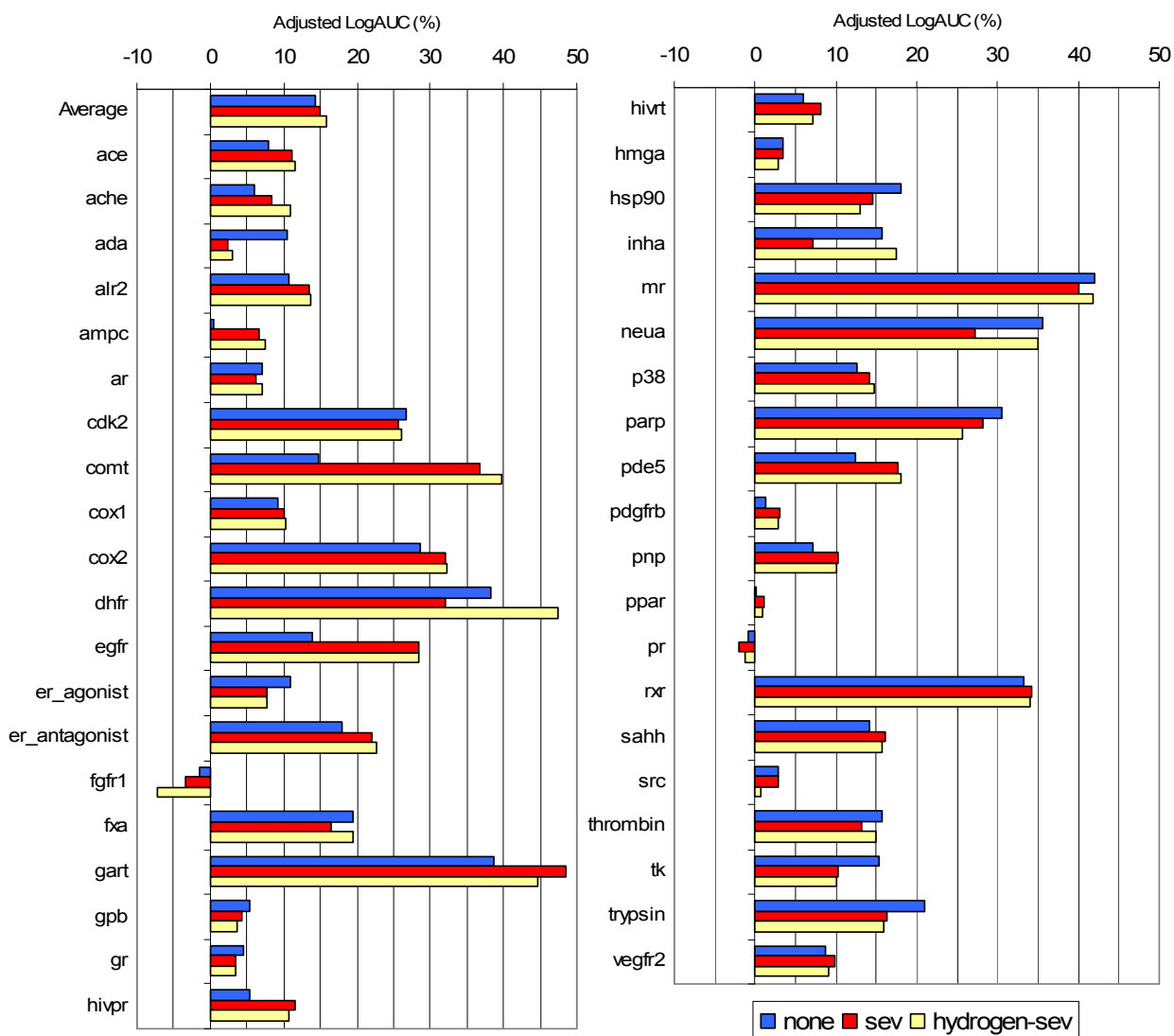
Figure A.1.3 All DUD Enrichment Plots versus Matched Decoys

ROC plots for each of the 40 DUD targets against matched decoys. Each legend contains the adjusted LogAUCs.

MR_AUTO ROC Plot - Own Decoys
none: 41.98
full: 34.09
sev: 40.04

NEUA_AUTO ROC Plot - Own Decoys
none: 35.50
full: -3.99
sev: 27.25

P38_SEMI ROC Plot - Own Decoys
none: 12.57
full: 16.74
sev: 14.18

PARP_SEMI ROC Plot - Own Decoys
none: 30.54
full: 16.43
sev: 28.21

PDE5_SEMI ROC Plot - Own Decoys
none: 12.37
full: 10.15
sev: 17.58

PDGFRB_AUTO ROC Plot - Own Decoys
none: 1.38
full: 3.57
sev: 2.98

PNP_AUTO ROC Plot - Own Decoys
none: 7.12
full: 10.49
sev: 10.32

PPAR_AUTO ROC Plot - Own Decoys
none: 0.18
full: 0.84
sev: 1.19

PR_AUTO ROC Plot - Own Decoys
none: -0.94
full: 0.90
sev: -1.95

RXR_AUTO ROC Plot - Own Decoys
none: 33.24
full: 35.62
sev: 34.29

SAHH_AUTO ROC Plot - Own Decoys
none: 14.23
full: 17.01
sev: 16.05

SRC_SEMI ROC Plot - Own Decoys
none: 2.94
full: 3.63
sev: 2.86

THROMBIN_AUTO ROC Plot - Own Decoys
none: 15.69
full: 12.05
sev: 13.17

TK_AUTO ROC Plot - Own Decoys
none: 15.42
full: 8.68
sev: 10.23

TRYPSIN_AUTO ROC Plot - Own Decoys
none: 20.91
full: 14.20
sev: 16.24

VEGFR2_SEMI ROC Plot - Own Decoys
none: 8.61
full: 8.85
sev: 9.84

Figure A.1.4 Effect of Changing Grid Parameters

Fractional desolvation for a fully desolvated atom (centered in a 48 Å cube of carbon atoms spaced 2 Å apart). A) Smaller grid resolutions indicate greater sampling, resulting in better exclusion of grid points inside the probe ligand atom. B) Increased integration cutoffs lead to fractional desolvation values closer to 1 (as expected).
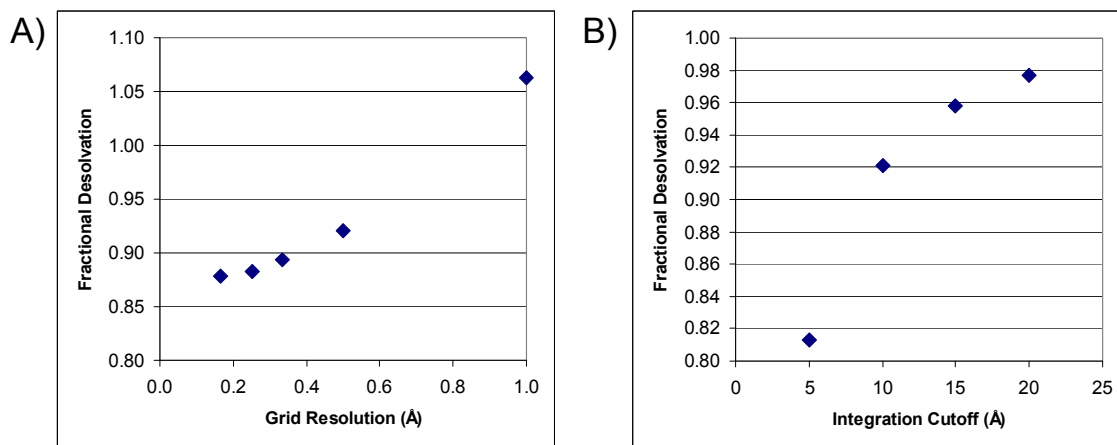
Figure A.1.5 Rank Comparison

Comparison of ligand (blue) and own matched decoy (red) ranks when using "none" or "sev" desolvation. A) gart, B) dhfr, C) cox1, and D) fxa.
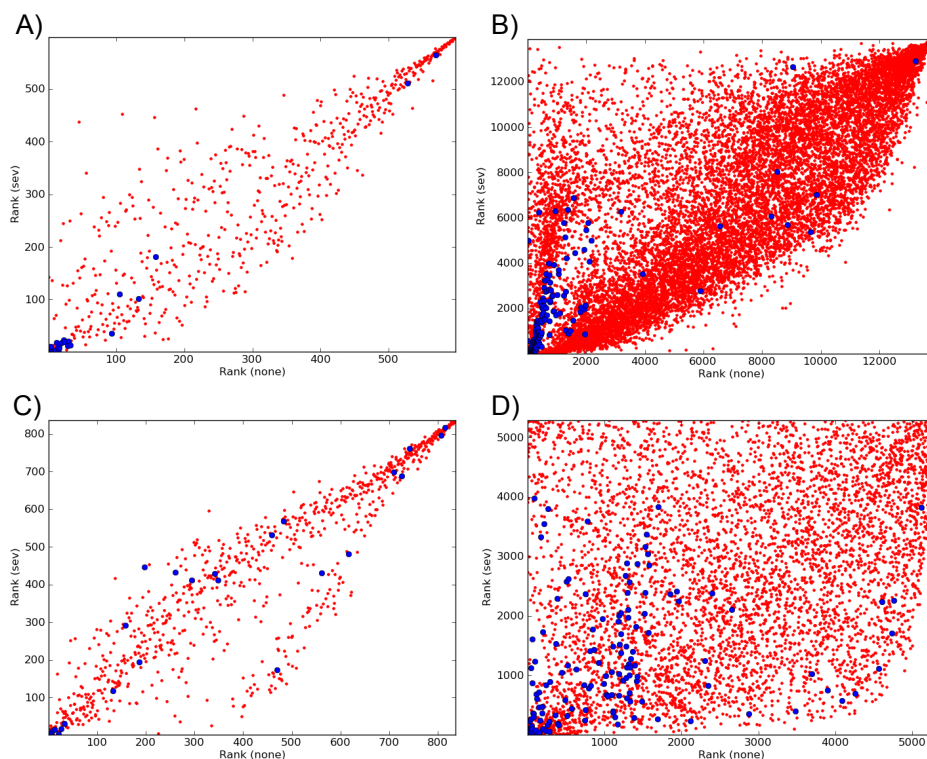
I. Analysis of Error in Self-Energy Term

Here we estimate the error in the self-energy term of the electrostatic component of ligand desolvation that arises from the fact that we neglect the ligand component of fractional desolvation. In generalized-born (GB) theory, the "true" self-energy term of ligand desolvation would be given by equation A.1.1, where ε_p and ε_W are the protein and water dielectrics, and α_i^L and α_i^{LR} are the effective born radii of the ligand alone and the ligand-receptor complex. Our context-dependent approximation is to multiply the fractional desolvation, including the receptor and solvent-excluded volume components but not the ligand component, by the transfer free energy computed by AMSOL. Since AMSOL is based on generalized-born, our self-energy term is approximately that in equation A.1.2, which simplifies to equation A.1.3. The error is thus given by the difference of equations A.1.3 and A.1.1, as shown in equation A.1.4.

(Equation A.1.1)

$$\Delta G_{L,desolv}^{GB,self} = \frac{1}{8\pi}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\sum_i\left[\frac{q_i^2}{\alpha_i^L} - \frac{q_i^2}{\alpha_i^{LR}}\right] = A\sum_i\left[\frac{q_i^2}{\alpha_i^L} - \frac{q_i^2}{\alpha_i^{LR}}\right]$$

$$\Delta G_{L,desolv}^{context,self} \cong \sum_i D_i \Delta G_{i,trans}^{GB,self} = \sum_i\left(1 - \frac{a_i}{\alpha_i^R}\right)\left(A\frac{q_i^2}{\alpha_i^L}\right) \text{ (Equation A.1.2)}$$

$$\Delta G_{L,desolv}^{context,self} \cong A\sum_i\left[\frac{q_i^2}{\alpha_i^L} - \frac{a_i q_i^2}{\alpha_i^R \alpha_i^L}\right] \qquad \text{(Equation A.1.3)}$$

$$\Delta G_{L,desolv}^{error,self} = \Delta G_{L,desolv}^{context,self} - \Delta G_{L,desolv}^{GB,self} \cong A \sum_i \left[\frac{q_i^2}{\alpha_i^{LR}} - \frac{a_i q_i^2}{\alpha_i^R \alpha_i^L} \right]$$

We can then use our sevsolv program with the dielectric boundary set to the molecular surface to numerically integrate the effective born radii of all the ligand atoms, in all three situations above: on the ligand alone, on the receptor alone, and on the final ligand-receptor complex. Additionally, we can estimate the effective born radii using "sev" method, which includes the receptor and atomic solvent-excluded volume. In this table, we estimate this error for the best pose of the rebuilt and re-docked crystallographic ligand of Factor Xa in DUD (pdb code 1F0R). We compute the error in the energies by scaling with the total atomic transfer free energy as computed by AMSOL instead of Aq_i^2 as found in equation A.1.4.

Table A.1.6: Analysis of Error in Self-Energy Term

Atom	Total Atomic Desolvation (kcal/mol)	1/alpha (LR)	1/alpha (L)	1/alpha (R)	a/(alpha(L) * alpha(R))	Error (1/alpha(LR))	Error (kcal/mol)
N1	-9.60	0.29	0.35	0.65	0.32	0.03	-0.32
C2	0.62	0.19	0.30	0.59	0.25	0.06	0.04
C3	-2.23	0.17	0.34	0.51	0.24	0.07	-0.16
C4	1.09	0.14	0.34	0.47	0.22	0.08	0.09
N5	-6.84	0.16	0.31	0.52	0.22	0.06	-0.43
C6	4.84	0.20	0.33	0.57	0.26	0.07	0.32
O7	-4.38	0.29	0.44	0.56	0.34	0.05	-0.24
C8	2.00	0.14	0.35	0.45	0.22	0.08	0.15
H9	0.85	0.19	0.38	0.56	0.29	0.10	0.09
H10	1.87	0.16	0.42	0.44	0.26	0.10	0.18
H11	1.67	0.25	0.43	0.52	0.32	0.06	0.11
H12	2.48	0.11	0.42	0.41	0.24	0.13	0.33
H13	1.47	0.18	0.43	0.41	0.25	0.07	0.10
S14	12.58	0.27	0.32	0.65	0.29	0.03	0.33
H15	6.56	0.40	0.46	0.66	0.42	0.02	0.13

Atom	Total Atomic Desolvation (kcal/mol)	1/alpha (LR)	1/alpha (L)	1/alpha (R)	a/(alpha(L) * alpha(R))	Error (1/alpha(LR))	Error (kcal/mol)
C16	-3.10	0.25	0.31	0.67	0.29	0.04	-0.11
O17	-0.56	0.29	0.41	0.60	0.35	0.05	-0.03
O18	-0.09	0.42	0.45	0.69	0.43	0.01	0.00
C19	-1.60	0.27	0.39	0.55	0.30	0.03	-0.04
C20	-1.01	0.20	0.38	0.49	0.26	0.06	-0.06
C21	-1.21	0.21	0.33	0.57	0.26	0.05	-0.06
C22	-0.67	0.28	0.33	0.66	0.30	0.03	-0.02
C23	-0.35	0.31	0.35	0.69	0.34	0.02	-0.01
N24	-0.80	0.36	0.40	0.68	0.38	0.02	-0.02
C25	-0.01	0.34	0.40	0.65	0.36	0.02	0.00
H26	2.96	0.38	0.52	0.52	0.38	0.00	0.01
H27	1.63	0.17	0.50	0.41	0.29	0.12	0.20
H28	0.70	0.44	0.46	0.70	0.45	0.01	0.01
H29	2.64	0.49	0.53	0.68	0.50	0.01	0.02
S30	-0.34	0.21	0.35	0.55	0.27	0.06	-0.02
C31	-2.50	0.10	0.31	0.39	0.17	0.07	-0.17
H32	2.19	0.22	0.46	0.44	0.29	0.07	0.16
H33	0.61	0.15	0.44	0.42	0.26	0.11	0.07
C34	-1.41	0.10	0.35	0.35	0.17	0.07	-0.09
C35	-3.02	0.08	0.36	0.28	0.14	0.06	-0.19
C36	0.12	0.07	0.33	0.28	0.13	0.06	0.01
C37	-4.36	0.07	0.38	0.20	0.11	0.04	-0.17
C38	2.14	0.07	0.40	0.18	0.10	0.03	0.07
C39	9.29	0.08	0.36	0.24	0.12	0.04	0.37
C40	-3.51	0.08	0.31	0.27	0.12	0.04	-0.15
C41	-1.51	0.09	0.33	0.36	0.16	0.08	-0.11
N42	-17.02	0.11	0.46	0.24	0.15	0.05	-0.78
N43	-12.28	0.08	0.41	0.22	0.12	0.05	-0.58
H44	3.24	0.16	0.46	0.35	0.23	0.07	0.22
H45	3.50	0.08	0.49	0.24	0.17	0.09	0.32
H46	4.61	0.07	0.50	0.19	0.13	0.07	0.30
H47	6.59	0.07	0.53	0.16	0.12	0.06	0.36
H48	1.85	0.09	0.43	0.33	0.20	0.11	0.20
H49	9.42	0.10	0.50	0.26	0.18	0.08	0.72
H50	11.75	0.14	0.55	0.22	0.17	0.03	0.40
H51	15.15	0.08	0.51	0.20	0.14	0.06	0.93
Sum	36.02						2.45

A.2 Chapter 2 Supporting Information

I. Supporting Results

Competition Models. Based on ligand IT1t RMSD and ligand-protein contacts, our best competition[1] model (#5) placed 5th out of 106. Even our worst model (#2) placed 0.4 standard deviations above the mean of other predictions. We were helped by correct modeling of D2.63, implicated for ligand binding by the limited mutational data[2], which required introduction of a single residue gap in CXCR4 at the top of helix II. Still, even our best model had a 6.1 Å ligand RMSD from the crystal structure, suggesting that there were substantial differences between them. Conversely, the binding site similarity was more reasonable, with a TM binding site heavy-atom RMSD of 2.2 Å (**Table A.2.2**).

Counter-Screen for Aggregation. To further investigate the role of promiscuous aggregation, all compounds were tested for inhibition of soluble enzymes AmpC β-lactamase and the thiol-protease cruzain (**Table A.2.4**, **Figure A.2.9**); these are widely used as counter-screens for aggregation-based inhibition[3, 4]. Consistent with the spin-down result, compound **6** inhibited cruzain with an IC50 value of 2 μM in the absence of the non-ionic detergent Triton X-100. Detergent addition increased (made worse) the apparent IC50 by 8-fold. We also investigated particle formation by dynamic light scattering (DLS), where compound **6** formed particles of radius 757 nm that scattered light at over 3 million counts/s at 100 μM. All observations are consistent with its status as a colloidal aggregator at relevant concentrations, so it was not considered a hit. Conversely, compounds **1-5** were better behaved. Although DLS particle formation was observed for compounds **2-4** at their CXCR4 calcium-flux IC50 concentration, with sizes between 60 and 92 nm, no compound inhibited AmpC at concentrations up to 100 μM. We observed some inhibition of cruzain,

but it was not reversible by even 0.1% Triton X-100. In fact, compounds **2-4** only inhibited cruzain at values substantially higher than their IC50 values in the calcium flux or radioligand binding assays. For instance, compound **3** showed 40% inhibition at 300 μM against cruzain, but its IC50 value in the radioligand assay was 306 nM. Taking these results together, we conclude that whereas compound **6** likely acts via promiscuous, aggregation-based inhibition, compounds **1-5** are well-behaved and act as specific inhibitors against CXCR4. The one caution would be for compound **1**, whose steep dose-response curve (Hill slope of 5) in the calcium flux assay and inhibition properties against cruzain may suggest some aggregation at the upper end of the concentrations measured here.

II. Supporting Methods

A flowchart (**Figure A.2.1**) visually illustrates our overall homology modeling procedure.

Alignment. To prepare for alignment, we truncated extraneous parts, like the T4 lysozyme insertion, from the four template crystal structures: rhodopsin, PDB ID code 1U19[5]; β1 adrenergic receptor, 2VT4[6]; β2 adrenergic receptor, 2RH1[7]; and A_{2A} adenosine receptor, 3EML[8]. We then used PROMALS-3D[9] to align those four structures, the human sequences of the CXCR and CCR receptors, as well as 12 related sequences, diverse in both sequence similarity and organism, as identified by the Basic Local Alignment Search Tool (BLAST)[10], such that 30 sequences were initially aligned. While this generally produced good alignment in the trans-membrane regions, the more distantly related extra-cellular loops required extensive manual adjustment, using more closely related sequences to build a plausible manual alignment. For example, adjusting the alignment to create an isoleucine insertion (BW 4.61, Ballesteros-Weinstein numbering[11]) at the top of β1 and β2 relative to all the other structures increases hydrophobic and hydrophilic residue overlap in the early part of extra-

126

cellular loop 2 (ECL2). Even in the trans-membrane region, $D^{2.63}$ in CXCR4, implicated to be important for binding by mutational data[2] initially lay outside the central GPCR binding site. To compensate, we introduced a single residue gap in CXCR4 at the top of helix II, experimenting with several different gap positions, to find the best compromise such that $D^{2.63}$ went into binding site while minimally disturbing the protein backbone. This gap was found to be critical in correctly modeling that region in the GPCR Dock 2010 assessment study[1].

Normal-Mode Based Virtual Template Generation. In addition to the four crystallographic templates mentioned earlier, we built additional homology models using computationally generated templates based on normal mode perturbations of elastic network models of the 4 crystallographic templates. Using Yang and Sharp's 3K-ENM program[12] we generated 198, 361, 350, and 455 backbone perturbations for 1U19, 2RH1, 2VT4, and 3EML, respectively. These were clustered by trans-membrane helix RMSD using K-means clustering[13] to obtain 128 representative structures for each of the four crystallographic templates. These additional 512 structures were then used as templates for further homology modeling and docking.

Homology Modeling. MODELLER v9.8 was used for homology model generation[14]. Four conserved aromatic residue sidechains were constrained to preserve the dihedral angles found in the template structures. These were $W^{4.50}$, $W^{6.48}$, $Y^{6.51}$ and $Y^{6.52}$ (Ballesteros-Weinstein numbering[11]). For each of the four crystallographic templates we built 64 homology models. For 1U19, 2VT4, and 2RH1, we generated 4 homology models for each of the 128 normal-mode based templates, while for 3EML we generated only 2 per normal-

127

mode template, as this template initially gave less promising docked poses for known

ligands. With 4 modeling failures, we generated and docked a total of 2044 homology

models.

Homology Model Docking. To inform homology model selection, we measured how well

the various homology models were able to retrospectively enrich the known CXCR4 ligands

in ChEMBL04[15]. We docked 60 known ligands, excluding any large macrocyclic or peptidic

ligands because of conformational sampling constraints. To better determine the enrichment

due to molecular docking itself, and not simple 1-dimensional chemical properties, we

generated 35 DUD[16] style property-matched decoys for each ligand using an automated

procedure (M.M.M. and B.K.S., see **Chapter 3**), docking 2516 molecules for each homology

model. To quantify overall enrichment while giving extra weight to early enrichment, we

used the adjusted LogAUC metric[17]. LogAUC is completely analogous to ROC AUC, being

the area under curve divided the maximum possible area, except it is computed after taking

the log of the x-axis to zoom in on early enrichment. As a metric, it shares attractive features

with ROC AUC, as it is easily interpreted, robust, and independent of similar extensive

variables[18].

Initial docking failed because extra-cellular loop 2 invaded the central GPCR binding

pocket. Thus, prior to docking we removed ECL2 (residues 173-195) and the linear

unmodeled N and C-terminal residues (1-35 and 320-352). Even with ECL2 removed, the

binding pocket was often still ill-formed for rigid protein docking, due to large number of

aromatic side-chains that had a tendency to crowd the binding pocket. To create initial space

we used PyMol to align the crystallographic ligands from 2VT4 and 3EML (cyanopindolol

and ZM241385, respectively) into the homology model, and then used Jacobson's Protein

Local Optimization Program (PLOP) to optimize the local protein side chains around the aligned ligands[19]. After docking, the best looking pose of ligand IT1t was selected and that pose was used as the basis of PLOP side-chain optimization in the next run. After several rounds, this bootstrapping procedure resulted in a stable, plausible pose of ligand IT1t. All future docking used this IT1t pose as the basis for PLOP side-chain optimization and as the starting point for DOCK matching sphere generation.

Molecular docking was performed with DOCK 3.6 including rapid context-dependent ligand desolvation using the solvent-excluded volume method with a single solvation grid for all atoms[17]. Parameters for docking were the same as in[17] except as noted. Matching spheres were still generated as described in[16] except the target number of spheres was 60 instead of 35. The bin size for both receptor and ligand were still set to 0.4 Å but the overlap size was reduced to 0.1 Å. Final energies were computed after 100 steps of rigid-body minimization. Docking flexibase were generated as described previously[17] where Schrödinger's Epik[20] replaced Ligprep for protomer and tautomer generation.

Model Selection. We examined several criteria to select the model to carry forward to the loop building stage. We measured the enrichment using adjusted LogAUC[17], both for the overall set of 60 small molecule CXCR4 ligands, as well as two subsets: 1) ligands that resembled ligand IT1t, and 2) other non-peptidic, non-macrocyclic ligands. We noted the rank of ligand IT1t rank in the 2516 docked ligand and decoys. The ROC-based enrichment factor at 1% of the decoys was tabulated for each model. We also calculated the number of ligands whose single best pose had a nitrogen atom within 3.4 Å of $E^{7.39}$, the most important residue for ligand binding in the limited mutational data that was available[2, 21]. Finally, in perhaps the most important and yet most qualitative measure, we looked at the docking

poses. For the top 100 or so models by quantitative measures, we manually looked through the poses to ensure that the binding site looked reasonable and that the IT1t-like ligands poses had good protein contacts and were similar to one another.

Loop Modeling. We then choose a single model to carry forward for loop modeling with Rosetta[22]. 500 ECL2 models were generated using the fragment-based loop modeling application in Rosetta v3.1. The CXCR4 ECL2 was defined as residues 171-198 to add flexibility to the tops of the connecting TM domains. Briefly, the centroid remodel and subsequent full-atom refine options were implemented, using the quick cyclic coordinate descent (CCD) method for refinement and loop closure after Monte Carlo fragment sampling. Fragment libraries were generated automatically by the Robetta server[23], in order to predict possible secondary structure elements in the CXCR4 ECL2. Two length constraints were used in the loop modeling. The first was to enforce the conserved disulfide bond between C186 and C109 by using a harmonic potential centered at 2.05 Å, with a standard deviation of 0.1 Å. The second, in order to anchor the resultant loop close to the ligand by burying a tryptophan side-chain, was a harmonic potential between W195 and the ligand, centered at 4.5 Å with a standard deviation of 0.5 Å. The +2 charged ligand in the final, docked pose was taken into account during the whole loop modeling procedure, with Rosetta parameters for the ligand being generated from OpenBabel 2.2.3 (http://openbabel.sourceforge.net) and the Rosetta-provided molfile_to_params script. 500 more ECL2 models were produced from a modified input structure, in which the rotamers of 3 residue side-chains (F199, Q200, and I259) were altered away from the ligand in order to open up the binding site, where we used same loop modeling procedure as above. The 1,000 ECL2 loop-models generated by Rosetta were stripped of polar hydrogens, and then

sent directly to docking grid generation. Docking was otherwise the same as described above.

Homology Model Virtual Screen. Using the same model selection criteria as above, we selected a single homology model with ECL2 to carry forward for a blind prospective experimental screen. We docked the ZINC[24] lead-like set, which at that time contained 3.312 million molecules, using the same docking procedure as above. Docking resulted in molecules that bound much higher (more extra-cellular) in the putative active site than previously known GPCR crystallographic ligands. Perhaps naively in retrospect, we manually added matching spheres to sample more orientations at the bottom of the active site. After filtering for chemotypes that were mischarged by Epik[20] or simply broken during our database preparation, we eventually looked through the top 900 scoring molecules from both runs, one with automated and one with manual matching spheres. Since these two runs used the same scoring grids, we can take the best result from either, and use the combined result to generate the final virtual screening statistics. We report the combined rank after filtering for mischarged and broken molecules. After looking through these poses, we narrowed the list manually, selecting 24 molecules for purchasing and experimental testing. Tanimoto coefficients were calculated to the best ChEMBL09[15] CXCR4 ligand using Pipeline Pilot from Accelrys using ECFP4 fingerprints folded to 1024-bits.

Crystal Structure Preparation. For docking, we had to choose from numerous CXCR4 structures simultaneously published[25] by Wu *et al.*: several CXCR4 structures in complex with ligand IT1t and one in complex the cyclic peptide CVX15. Additionally, the best resolution (2.5 Å) ligand IT1t structure, 3ODU, has two CXCR4 copies in the asymmetric unit.

Whereas 3ODU has 2 lipids that invade each copy of the TM helical bundle, the peptide structure, 3OE0, is free of any lipid perturbations but has a lower resolution (2.9 Å). Thus, which binding site to use for prospective screening was not entirely straightforward. So we manually assigned histidine protonation states in all the binding sites, and docked the known ChEMBL04[15] ligands and the corresponding DUD-style[16] property-matched decoys. Including lipid during docking did not help enrichment for both 3ODU binding sites, and was not considered further. Further, the 3ODU B chain binding site had smaller experimental B-factors, and the invading lipid head group, although charged, was much further away from ligand IT1t. The 3ODU B chain also proved adequate at enriching known ligands against DUD-style decoys, and was thus selected for prospective testing.

Still, the binding site in CXCR4 seemed like a historically challenging site for docking, being enormous, highly charged, and highly solvent exposed. In particular the enormity of the site caused a particular dilemma for how to fill the active site with low-dielectric spheres before calculating the electrostatic grid with Delphi[26]. Our usual automatic procedure produced low-dielectric near the crystallographic ligand, but left 2/3 of the binding pocket empty. Thus charge-charge interactions with the numerous carboxylic acids in the binding cavity that were not close to where ligand IT1t bound would be too highly screened by solvent, potentially preventing us from finding some novel ligands. Manually filling the entire pocket with low-dielectric spheres on the other hand, results in overweighting charge-charge interactions for atoms placed at the center of the cavity, even though they are relatively far away from the protein. Our new solution was to create a single layer of low-dielectric spheres all over the binding pocket, but not extending into the center of the pocket. Thus ligand-protein interaction remained strong everywhere, while changing

bulk electrostatics as little as possible. For prospective screening, we combined this electrostatic treatment with solvent-excluded volume based ligand desolvation[17].

The large binding cavity also presented a challenge for exhaustive ligand sampling. To compensate, we created one set of scoring grids that covered the entire binding site, but divided the binding pocket into three partially overlapping sub-sites, one centered next to $E^{7.39}$, one centered next to $D^{2.63}$, and one centered where the napthalen-2-yl-3-alanine residue of CVX15 binds in 3OE0. We used 60 matching spheres in the $E^{7.39}$ and $D^{2.63}$ sites and 45 in the deeper, more remote naphthalene site. We then docked the $ZINC^{24}$ lead-like set, which now contained 4.17 million compounds, to all three sub-sites and took the best scoring pose generated from any of the three. The net effect was to divide the enormous sampling problem into three sub-problems and then combine the results since the same scoring grids where used. After filtering out problematic molecules, we manually looked through the top 500 molecules, selecting 23 molecules for purchasing and experimental testing. We used TravelDepth to highlight depth in some figures of the CXCR4 cavity[27].

Compound Sources. Compounds were obtained from the following sources: compound **1** Developmental Therapeutics Program at the National Cancer Institute, compounds **2-4** from Chembridge and compound **5** from Princeton Biomolecular Research. All compounds were sourced at 95% or greater purity as described by the vendors. All active compounds were further tested for purity by LC/MS, at UCSF, and were found to be pure as judged by peak height and identity.

Compound Solubility and Auto-fluorescence. Compound stocks were prepared in dimethylsulfoxide (DMSO) to either 10 mM or 100 mM. To test for solubility, compounds

were resuspended in assay buffer: Hanks buffered saline solution (HBSS), 20 mM HEPES (pH 7.4), 0.5% (w/v) bovine serum albumin (BSA), and either 1% (v/v) DMSO (compounds **1** and **5**) or 0.5% (v/v) DMSO (compounds **2**-**4** and **6**). Absorbance was measured from 620-820 nm and percent transmittance was calculated at 620 nm using a FlexStation 3 Microplate Reader (Molecular Devices). Subsequent analysis of cellular calcium flux utilizes fluorescence excitation and emission at 485 nm and 515 nm, respectively. The fluorescence emission spectrum for each compound was scanned from 500 nm to 620 nm at an excitation of 485 nm. No compound exhibited a fluorescence spectrum distinct from the vehicle control indicating compatibility with the assay.

Calcium Flux-based Assays. THP-1 monocytes obtained from American Type Culture Collection were washed twice and resuspended in 96-well format at 2×10^5 cells/well in assay buffer containing FLIPR Calcium4 dye (Molecular Devices) as per manufacturer's instructions. Compounds were added for a final concentration of 100µM (single point) or the indicated concentrations (dose-response) and the plates were then incubated for 1 h at 37°C in 5% CO_2. Fluorescence was measured at 37 °C using a FlexStation3 Microplate Reader with excitation and emission wavelengths at 485 nm and 515 nm, respectively. After a 20 s baseline measurement, CXCL12 was added for a final concentration of 30 nM and the resulting calcium response was measured for an additional 50 s. CCL2, a chemokine that targets CCR2 – a distinct receptor of THP-1 cells, was then added as a control for compound specificity. Approximately 60 s after CXCL12 addition, 6 nM CCL2 was added to each well and the calcium mobilization was measured for an additional 40 s. Percent calcium flux for each agonist was calculated from the maximum fluorescence minus the minimum fluorescence as a percent of baseline. A two-tailed student's t-test between either the 30 nM

CXCL12 control or the 6 nM CCL2 control, and the compound of interest was used to identify statistically significant inhibitory compounds (**Figure A.2.8**). For significant CXCR4 inhibitors, the assay was repeated in a dose-response format, where IC_{50} values were determined by non-linear fitting to a four parameter logistic function.

Chemotaxis and Viability. Chemotaxis experiments were performed using Transwell (5 μm pore size; Costar) in 24-well format. CXCL12 ligand (30 nM) and respective compounds (100 μM) were re-suspended in RPMI 1640 containing 0.2% (w/v) BSA and 600 μL was added to the lower chamber; samples were prepared in triplicate and two runs performed. THP-1 cells were washed twice and resuspended at 5×10^6 cells/mL and 100 μM compound. 100 μL cell solution was added to the upper chamber and added to a TC-treated 96-well plate (BD Falcon) for viability studies. Both plates were incubated for 3 h at 37°C in 5% CO_2. Final DMSO concentration was 0.1% for compounds **1**-**4** and 1% for compound **5**. Migrated cells were counted using a TC-10 Automated Cell Counter (BioRad) and hemocytometer. Percent maximal migration was number of migrated cells with compound divided by number that migrated to CXCL12 alone. Cell viability was number of cells excluding Trypan Blue stain divided by total cell number; experiments were performed in duplicate.

Radioligand Binding. Binding studies were performed on pre-B leukemia REH cells. Cells were resuspended in binding buffer (50 mM HEPES, pH 7.4, 5 mM $MgCl_2$ 1 mM $CaCl_2$ and 0.2% BSA) and seeded in 96-wells in increasing concentrations of unlabeled competitor. The competition binding assays were carried out at 4°C for 1 h using 50 pM $[I^{125}]$-CXCL12 (PerkinElmer Life and Analytical Sciences, Waltham, MA) as a tracer. A combi cell

harvester was used to remove the excess tracer (Molecular Devices, Menlo Park, CA). The samples were captured on glass-fiber filter paper (Skatron Instruments INC. USA). The counts were measured using a Wallace Wizard 1470 gamma counter (PerkinElmer Life and Analytical Sciences, Waltham, MA). Experiments were performed in duplicate.

III. Supporting References

1. Kufareva, I; Rueda, M; Katritch, V; Stevens, RC; Abagyan, R; Participants, GD. "Status of Gpcr Modeling and Docking as Reflected by Community-Wide Gpcr Dock 2010 Assessment" Structure 2011 19: 1108-1126.

2. Wong, RS; Bodart, V; Metz, M; Labrecque, J; Bridger, G; Fricker, SP. "Comparison of the Potential Multiple Binding Modes of Bicyclam, Monocylam, and Noncyclam Small-Molecule Cxc Chemokine Receptor 4 Inhibitors" Mol Pharmacol 2008 74: 1485-1495.

3. Coan, KE; Shoichet, BK. "Stoichiometry and Physical Chemistry of Promiscuous Aggregate-Based Inhibitors" J Am Chem Soc 2008 130: 9606-9612.

4. Doak, AK; Wille, H; Prusiner, SB; Shoichet, BK. "Colloid Formation by Drugs in Simulated Intestinal Fluid" J Med Chem 2010 53: 4259-4265.

5. Palczewski, K; Kumasaka, T; Hori, T; Behnke, CA; Motoshima, H; Fox, BA; Le Trong, I; Teller, DC; Okada, T; Stenkamp, RE; Yamamoto, M; Miyano, M. "Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor" Science 2000 289: 739-745.

6. Warne, T; Serrano-Vega, MJ; Baker, JG; Moukhametzianov, R; Edwards, PC; Henderson, R; Leslie, AG; Tate, CG; Schertler, GF. "Structure of a Beta1-Adrenergic G-Protein-Coupled Receptor" Nature 2008 454: 486-491.

7. Cherezov, V; Rosenbaum, DM; Hanson, MA; Rasmussen, SG; Thian, FS; Kobilka, TS; Choi, HJ; Kuhn, P; Weis, WI; Kobilka, BK; Stevens, RC. "High-Resolution Crystal Structure of an Engineered Human Beta2-Adrenergic G Protein-Coupled Receptor" Science 2007 318: 1258-1265.

8. Jaakola, VP; Griffith, MT; Hanson, MA; Cherezov, V; Chien, EY; Lane, JR; Ijzerman, AP; Stevens, RC. "The 2.6 Angstrom Crystal Structure of a Human A2a Adenosine Receptor Bound to an Antagonist" Science 2008 322: 1211-1217.

9. Pei, J; Kim, BH; Grishin, NV. "Promals3d: A Tool for Multiple Protein Sequence and Structure Alignments" Nucleic Acids Res 2008 36: 2295-2300.

10. Altschul, SF; Gish, W; Miller, W; Myers, EW; Lipman, DJ. "Basic Local Alignment Search Tool" J Mol Biol 1990 215: 403-410.

11. Ballesteros, JA; Weinstein, H. "Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors" Methods in Neurosciences 1995 25: 366-428.

12. Yang, Q; Sharp, KA. "Building Alternate Protein Structures Using the Elastic Network Model" Proteins 2009 74: 682-700.

13. Arthur, D; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. In Symposium on Discrete Algorithms (SODA), 2007.

14. Sali, A; Blundell, TL. "Comparative Protein Modelling by Satisfaction of Spatial Restraints" J Mol Biol 1993 234: 779-815.

15. Overington, J. "Chembl. An Interview with John Overington, Team Leader, Chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (Embl-Ebi). Interview by Wendy A. Warr" J Comput-Aided Mol Des 2009 23: 195-198.

16. Huang, N; Shoichet, BK; Irwin, JJ. "Benchmarking Sets for Molecular Docking" J Med Chem 2006 49: 6789-6801.

17. Mysinger, MM; Shoichet, BK. "Rapid Context-Dependent Ligand Desolvation in Molecular Docking" J Chem Inf Model 2010 50: 1561-1573.

18. Nicholls, A. "What Do We Know and When Do We Know It?" J Comput-Aided Mol Des 2008 22: 239-255.

19. Jacobson, MP; Friesner, RA; Xiang, Z; Honig, B. "On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations" J Mol Biol 2002 320: 597-608.

20. Shelley, JC; Cholleti, A; Frye, LL; Greenwood, JR; Timlin, MR; Uchimaya, M. "Epik: A Software Program for Pk(a) Prediction and Protonation State Generation for Drug-Like Molecules" J Comput-Aided Mol Des 2007 21: 681-691.

21. Rosenkilde, MM; Gerlach, LO; Hatse, S; Skerlj, RT; Schols, D; Bridger, GJ; Schwartz, TW. "Molecular Mechanism of Action of Monocyclam Versus Bicyclam Non-Peptide Antagonists in the Cxcr4 Chemokine Receptor" J Biol Chem 2007 282: 27354-27365.

22. Wang, C; Bradley, P; Baker, D. "Protein-Protein Docking with Backbone Flexibility" J Mol Biol 2007 373: 503-519.

23. Kim, DE; Chivian, D; Baker, D. "Protein Structure Prediction and Analysis Using the Robetta Server" Nucleic Acids Res 2004 32: W526-531.

24. Irwin, JJ; Shoichet, BK. "Zinc--a Free Database of Commercially Available Compounds for Virtual Screening" J Chem Inf Model 2005 45: 177-182.

25. Wu, B; Chien, EY; Mol, CD; Fenalti, G; Liu, W; Katritch, V; Abagyan, R; Brooun, A; Wells, P; Bi, FC; Hamel, DJ; Kuhn, P; Handel, TM; Cherezov, V; Stevens, RC. "Structures of the Cxcr4 Chemokine Gpcr with Small-Molecule and Cyclic Peptide Antagonists" Science 2011 330: 1066-1071.

26. Nicholls, A; Honig, B. "A Rapid Finite-Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson-Boltzmann Equation" J Comput Chem 1991 12: 435-445.

27. Coleman, RG; Sharp, KA. "Travel Depth, a New Shape Descriptor for Macromolecules: Application to Ligand Binding" J Mol Biol 2006 362: 441-458.

IV. Supporting Tables

Table A.2.1 Chemotaxis and Cell Viability

Inhibition of CXCL12 mediated chemotaxis in THP-1 cells by compounds **1-5** at 100 μM and cell viability at 3 hours.

Compound	Maximal Migration (%)	Cell Viability (%)
CXCL12	100 ± 21	97.2 ± 1.0
1	3.2 ± 4.8	95.3 ± 2.4
2	1.4 ± 1.6	95.0 ± 0.8
3	8.8 ± 5.1	96.5 ± 1.0
4	40 ± 8.9	97.3 ± 1.0
5	6.6 ± 3.9	96.0 ± 0.8
plerixafor (AMD3100)	2.9 ± 2.6	98.0 ± 0.0
DMSO control	4.3 ± 4.9	97.1 ± 1.0

Table A.2.2 Model Performance

Model performance for the prospectively docked model, the two highest scoring competition models: the de Graaf *et al.* model (VU-5) and the Vaidehi *et al.* model (COH-1), a homology model we retrospectively found to be closest to the crystal structure, and the competition homology models we submitted to 2010 GPCR Docking Assessment.
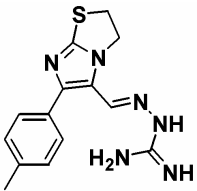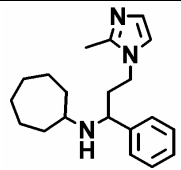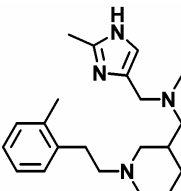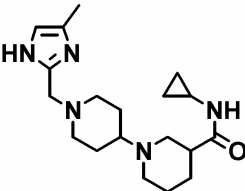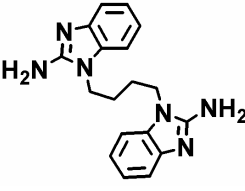
	Trans-membrane backbone RMSD (Å)	TM bindingsite heavy-atom RMSD (Å)	Ligand IT1t RMSD (Å)	LogAUC (%)	ROC EF1 (fold)
Prospective model docked	3.5	2.2 (6.7*)	9.5	30 (21*)	38 (15*)
VU-5	3.1	4.0 (5.1*)	4.9	4 (5*)	2 (3*)
COH-1	5.2	3.6 (5.6*)	6.0	5 (6*)	2 (5*)
Our retrospective best homology model	4.2	2.6	2.9	22	22
Competition model #1	3.6	2.1	7.7	14	3
Competition model #2	3.5	2.9	7.2	9	10
Competition model #3	3.3	2.2	7.3	14	10
Competition model #4	3.2	2.5	5.1	15	3
Competition model #5[†]	3.3	2.0	6.1	8	0

* with ecl2 modeled

[†] placed 5th in competition

Table A.2.3 Inhibitor Ranks
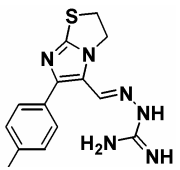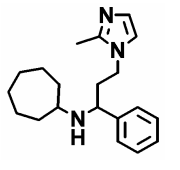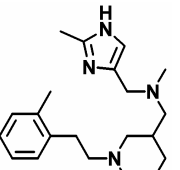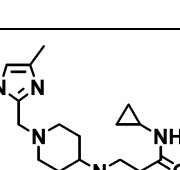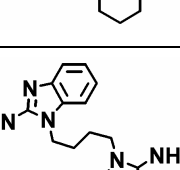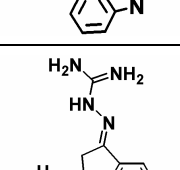
Ranking of the five newly discovered ligands against the background of the ZINC lead-like set for the crystal structure, the prospective homology model, the two highest scoring competition models: the de Graaf *et al.* model (VU-5) and the Vaidehi *et al.* model (COH-1), and the model we retrospectively found to be closest to the crystal structure.

#	Structure	Crystal	Prospective Model	VU-5	COH-1	Retrospective best model
1		55	2803	312290	19977	326177
2		521	50121	2037738	440711	15344
3		604	30898	1186418	117316	20330
4		178	5380	412852	63803	22123
5		620	5800	3282	107651	15823
Range (min-max)		55-620	2803-50121	3282-2037738	19977-440711	15344-326177
ChEMBL04 LogAUC (%)		28	21	4.8	6.2	22

*Ranks reported are not filtered for broken molecules.

Table A.2.4 Aggregation Counter Screens

Aggregation counter screening assays of compounds **1-6** discovered by docking screens and screened for inhibition of CXCL12 induced calcium flux.

#	Structure	IC_{50} (µM)[a]	Hill Slope[a]	DLS Intensity (count/s)	DLS R (nm)	Cruzain Dose-Response IC_{50} (µM) No / 0.1% Triton X	Before/After Spindown Assay (% calcium flux relative to CXCL12)	AmpC % Remaining Activity (at concentration)
1		107	-5.0	11845	32.5	73 / 92	66% / 62%	101% (100 µM)
2		76	-2.4	164956	60.3	233 / 207	43% / 32%	96% (80 µM)
3		57	-2.1	543409	77.6	- / -	48% / 31%	95% (60 µM)
4		77	-1.5	1198000	92.3	258 / 307	58% / 47%	98% (80 µM)
5		55	-2.9	68565	121.2	17 / 10	26% / 16%	99% (60 µM)
6		103	-10	3009770	757.4	2 / 15	36% / 108%	101% (100 µM)

[a] From CXCL12 induced calcium flux DR curves

V. Supporting Figures

Figure A.2.1 Flowchart of Homology Modeling Process
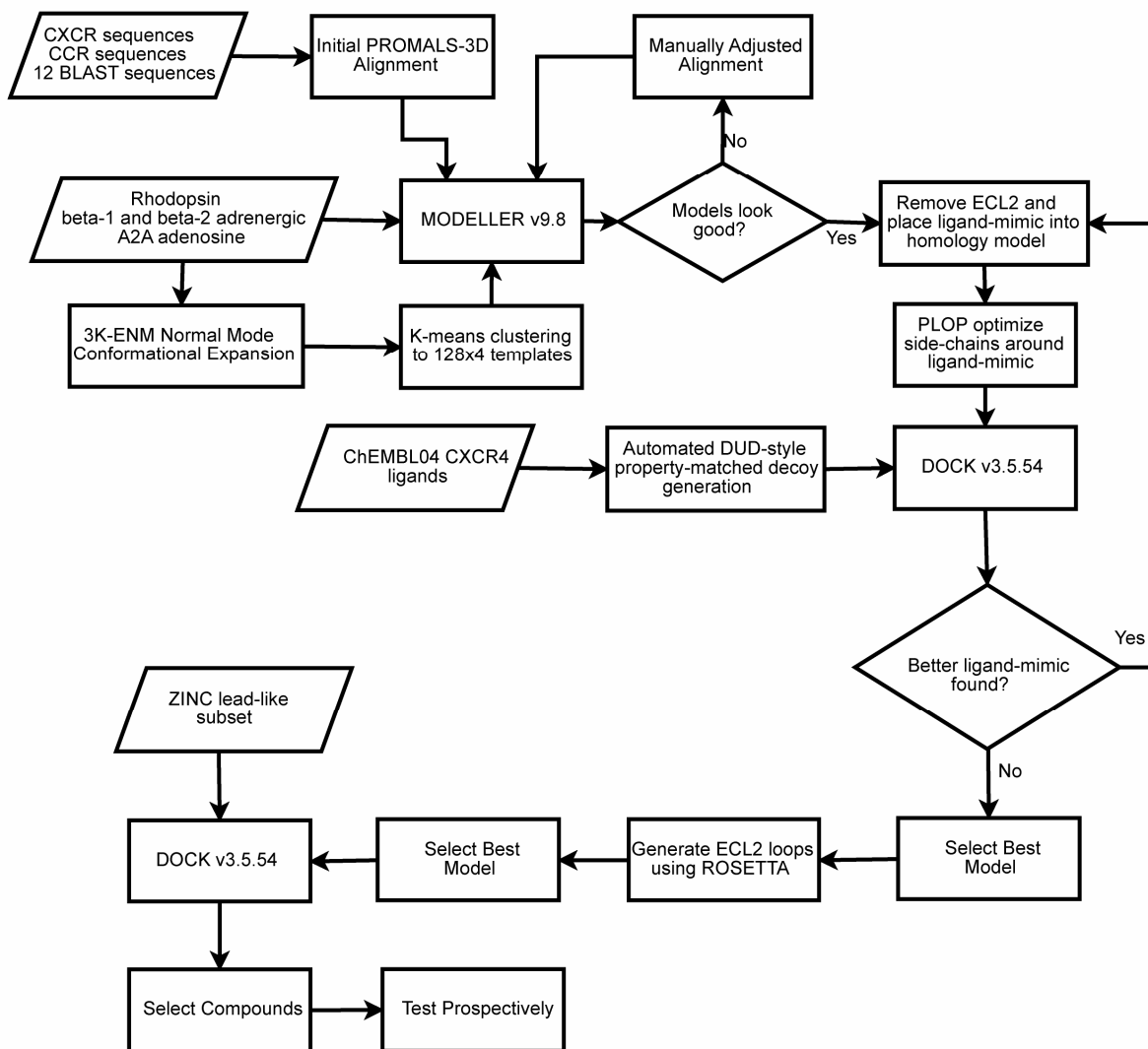
Figure A.2.2 CXCR4 Alignment with Available GPCR Templates

BW residues are marked as A, binding site residues as B, and transmembrane helix regions as H. Sequence alignment image was generated in PFAAT (http://pfaat.sourceforge.net) with default coloring.
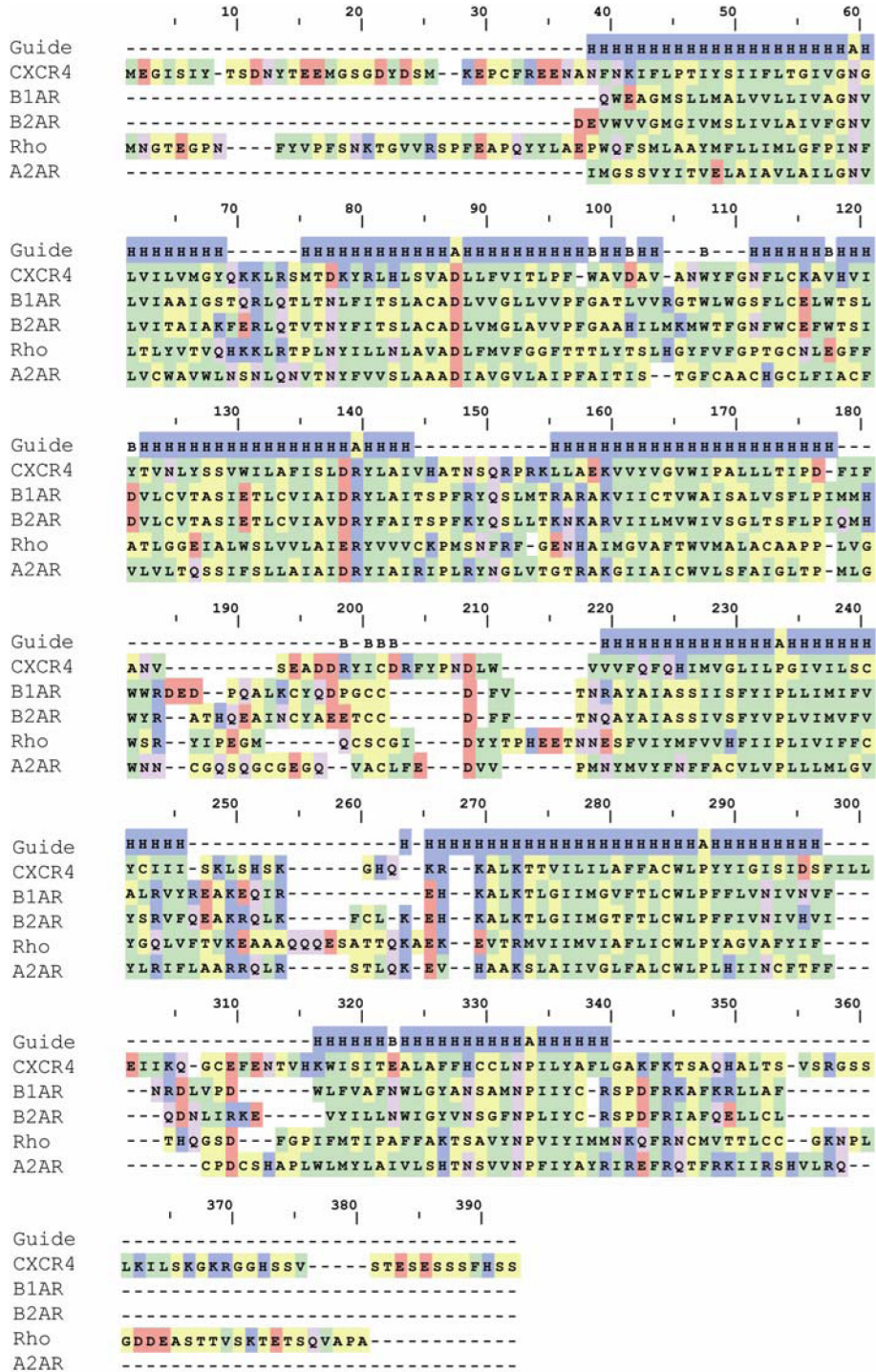
Figure A.2.3 Radioligand Binding Assays

All figures show pooled data from 2 to 5 assays in REH cells [(n=2, compound **1**), (n=3, CXCL12, compound **3**, **5**-**6**), (n=4, compound **2**) or (n=5, compound **4** and DMSO alone)] independent experiments performed in duplicate, ± S.E.M. DMSO alone had a minor effect, increasing [125I]-CXCL12 binding, however not to the degree as compound 4. IC50s, LogIC50, ± S.E.M. : CXCL12 (1.5 nM, -8.8 ± 0.1), Compound **1** (225 μM, -3.6 ± 0.2), Compound **2** (-), Compound **3** (306 nM, -6.5 ± 0.2), Compound **4** (2.16 μM, -5.7 ± 0.3), Compound **5** (14 μM, -4.9 ± 0.2).
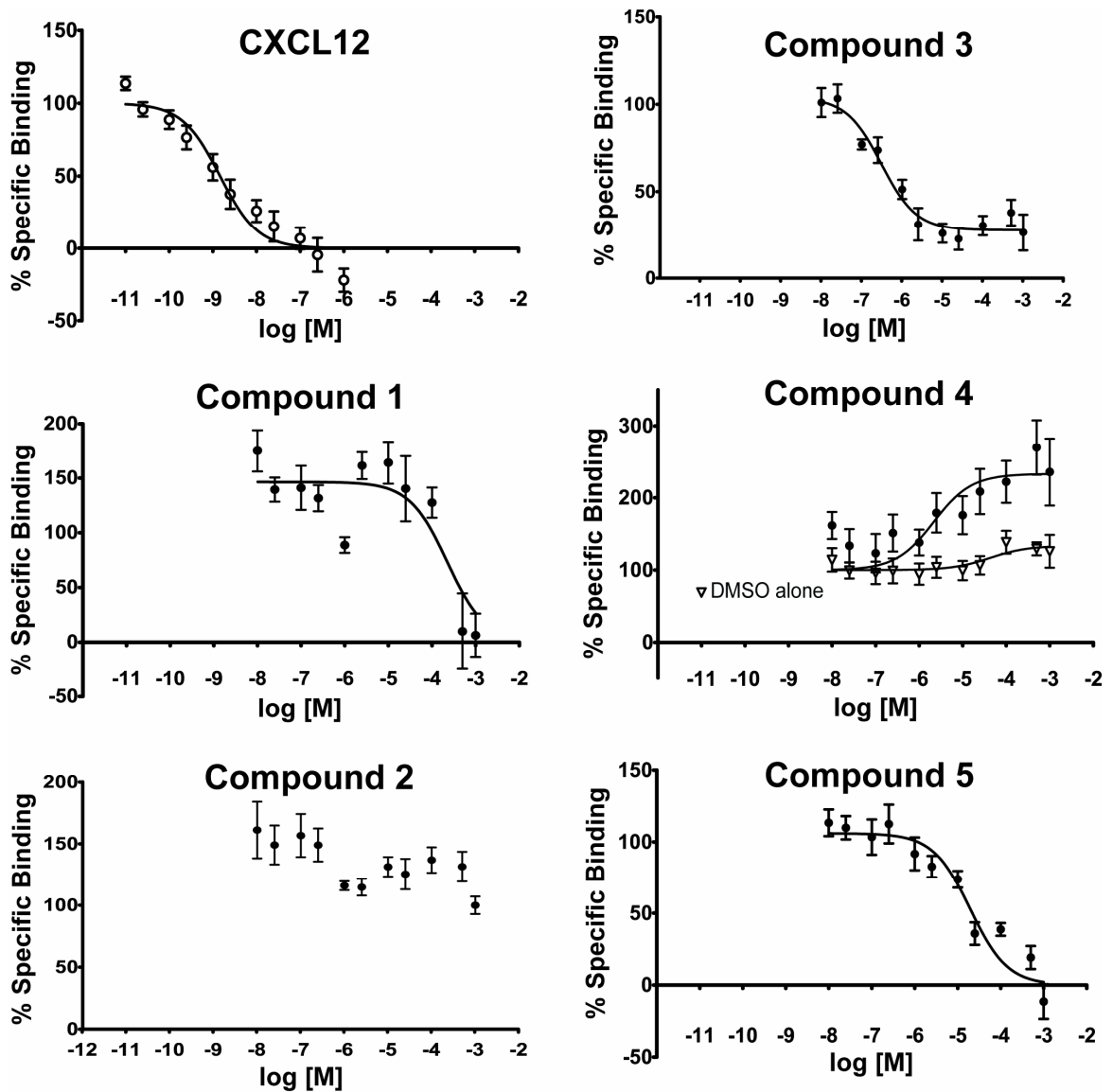
Figure A.2.4 Binding Site Comparison

(A) Rhodopsin (1U19) structure with retinal (yellow). (B) The CXCR4 homology model used for virtual screening, with the 1T1t small molecule (yellow). (C) The CXCR4/1T1t crystal structure (3ODU). Depth of the binding site is indicated, as calculated by Travel Depth[27].
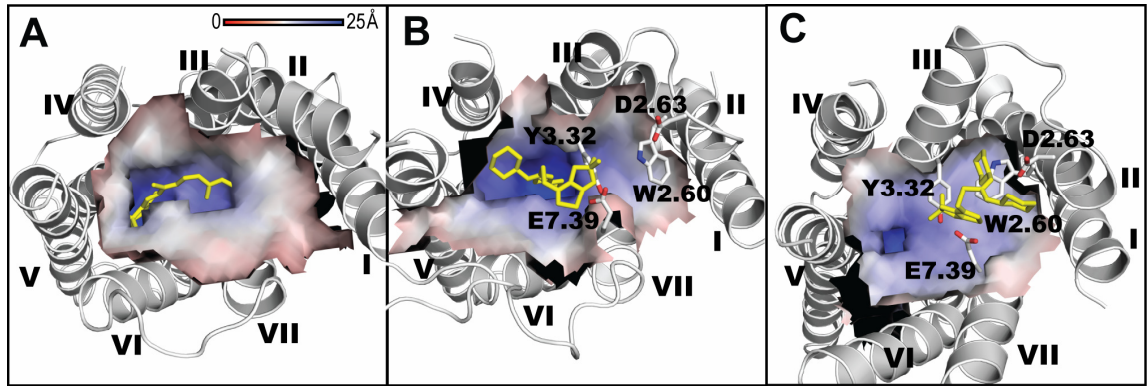
Figure A.2.5 Retrospective Enrichment of CXCR4 Models

ROC enrichment curves of the crystal structure (blue), the prospective homology model (green), the retrospective best homology model (red), the VU-5 model (cyan) and the COH-1 model (magenta). (A) Retrospective enrichment of the known ChEMBL04 CXCR4 ligands against property matched decoys. (B) Retrospective enrichment of the five newly discovered ligands against the ZINC lead-like database background.
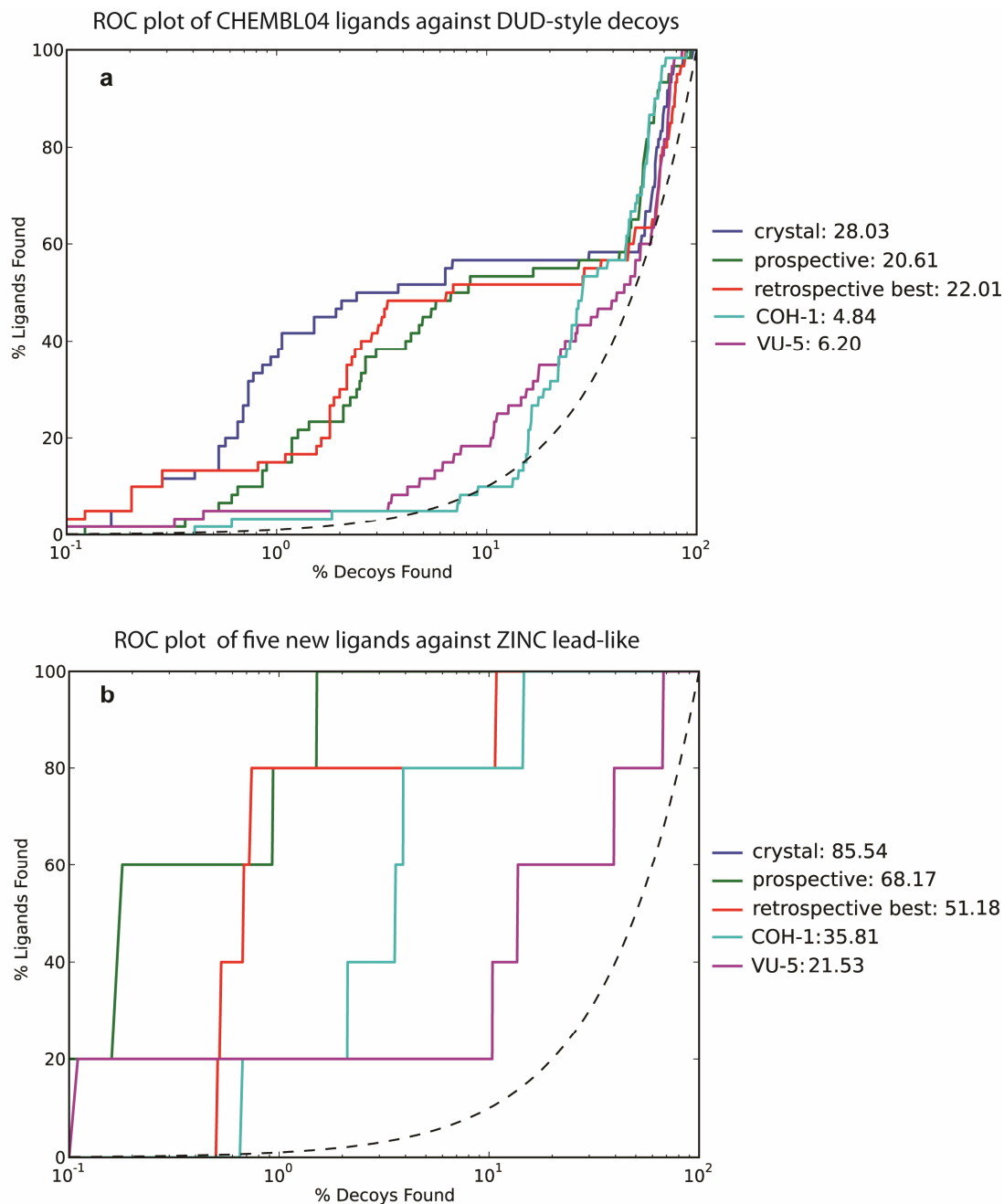
Figure A.2.6 Loopless Homology Model Statistics

Distributions of the 2044 loopless homology models generated, from which the prospective model was picked. The prospective model is marked with a pink star. (A) Retrospective enrichment of the known ChEMBL04 CXCR4 ligands. (B) Binding-site RMSD. (C) Ligand RMSD. (D) Enrichment was not found to correlate with binding-site RMSD, one measure of model accuracy. (E) Binding-site RMSD did correlate to some extent with normalized DOPE score, a measure of model quality. (F) Normalized DOPE score does not correlate with retrospective enrichment.
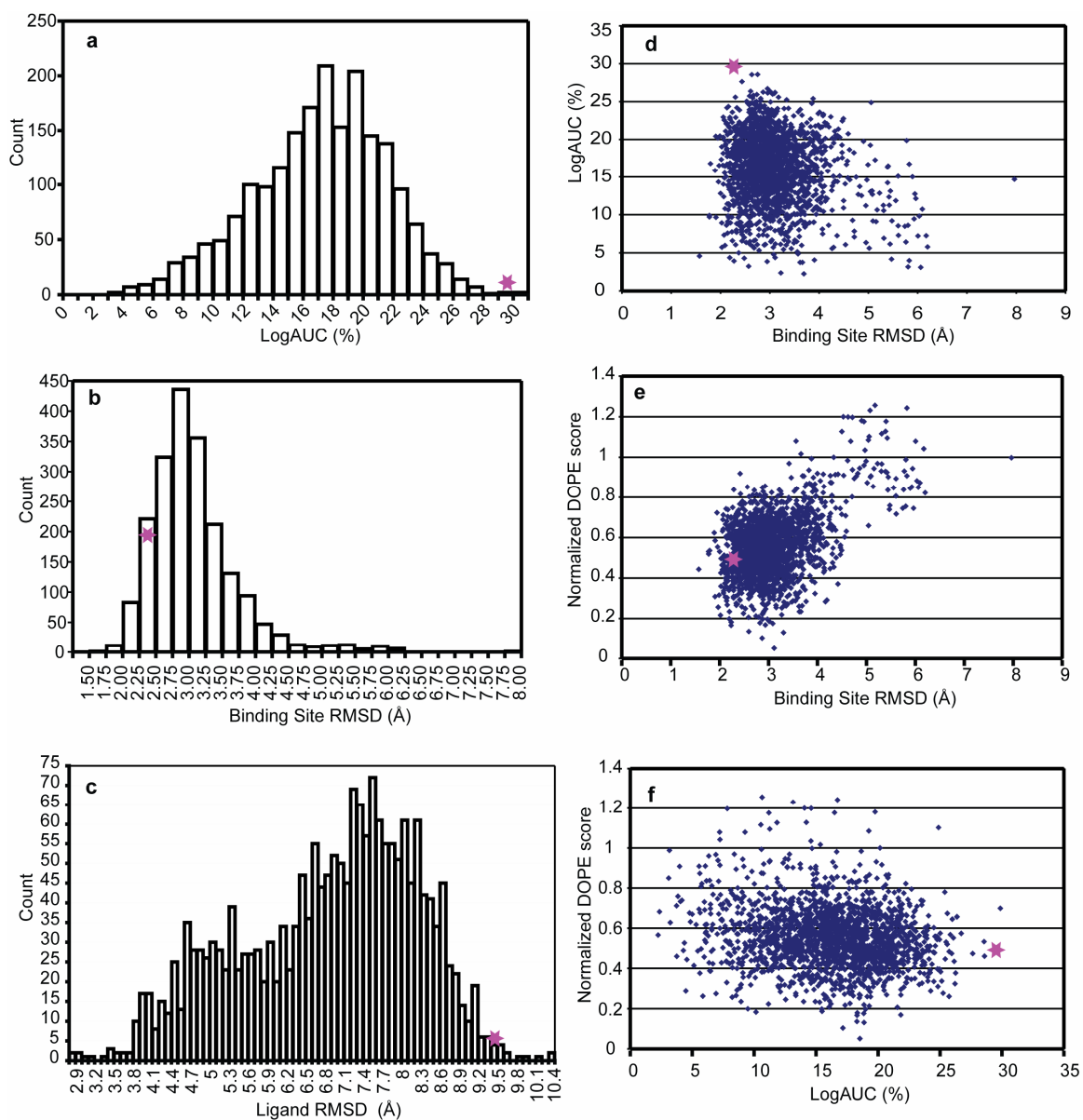
Figure A.2.7 Promiscuous Aggregator Inhibiting Membrane Protein

Compound **6**, not reported as a hit, showed substantial inhibition in the calcium flux assay (top) but had a steep dose response curve in both calcium flux and direct binding assays (bottom), often a red flag for aggregation. Indeed, in further screening, this compound was found to be an aggregator which inhibits cruzain in a detergent sensitive manner.
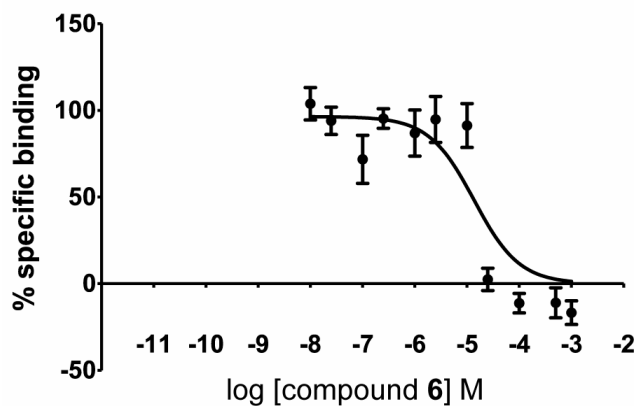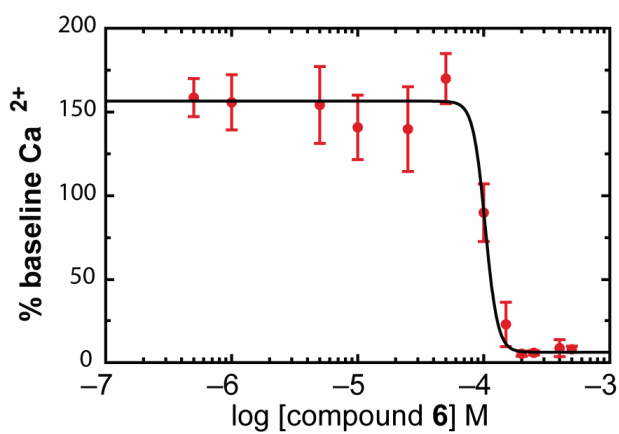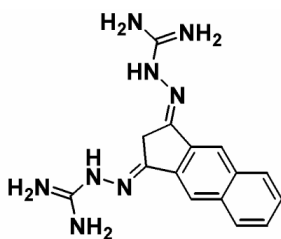
Figure A.2.8 Testing for CXCR4-Specific Inhibitors

THP-1 monocytes were incubated with either (A) DMSO or (B) 100 μM compound for 1 h. Cells were stimulated with 30 nM CXCL12 at ~20 s and then exposed to 6 nM CCL2 at ~80 s. The resulting calcium response to each agonist was measured as the maximum peak height minus the baseline signal. A two-tailed Student's t-test between either the 30 nM CXCL12 control or the 6 nM CCL2 control, and the compound of interest was used to identify statistically significant inhibitory compounds. (C) Dose-response curve of plerixafor in calcium flux assay. (D) Dose-response curve of CXCL12 in calcium flux assay.
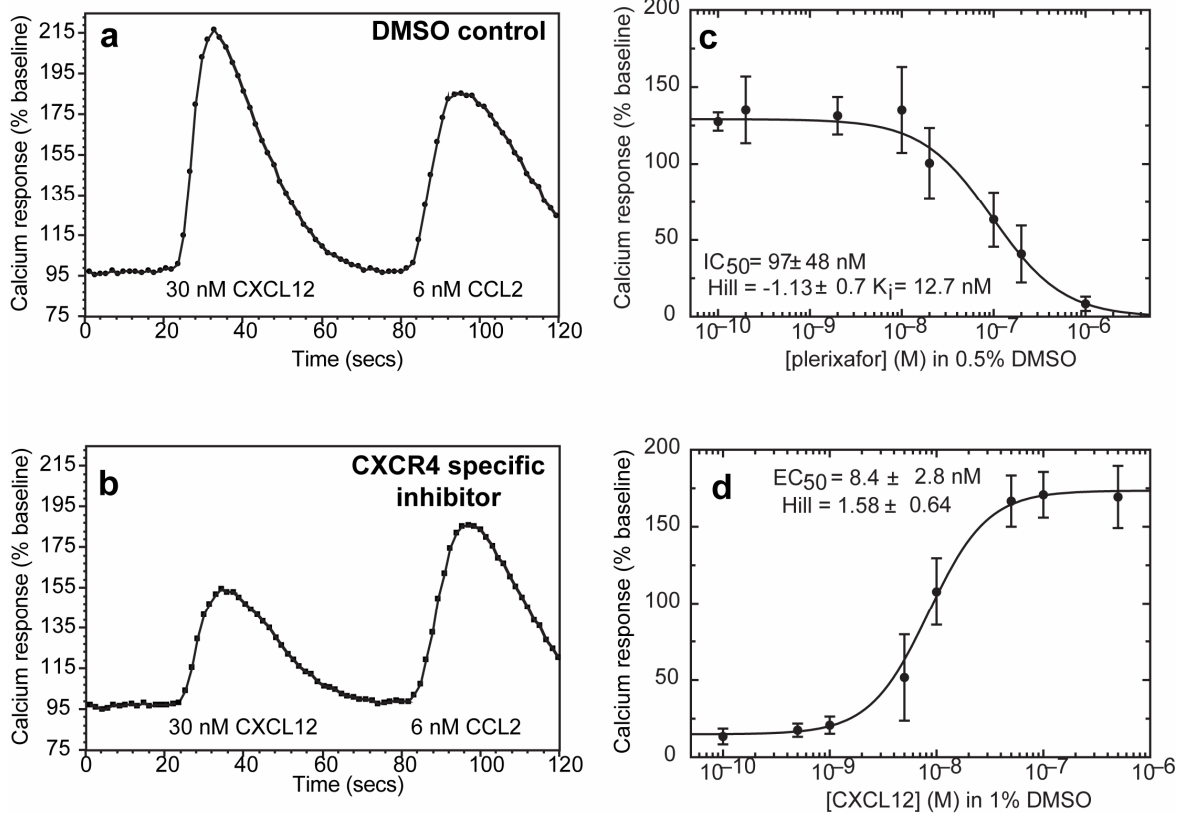
Figure A.2.9 Aggregation Dose-Response Conterscreen Against Cruzain

Dose-response curves of cruzain inhibition measured with and without the detergent Triton X for compounds **1-6**. Compound **3** in the presence of 0.1% Triton dose-response curve does not converge, and therefore no trend line is shown.
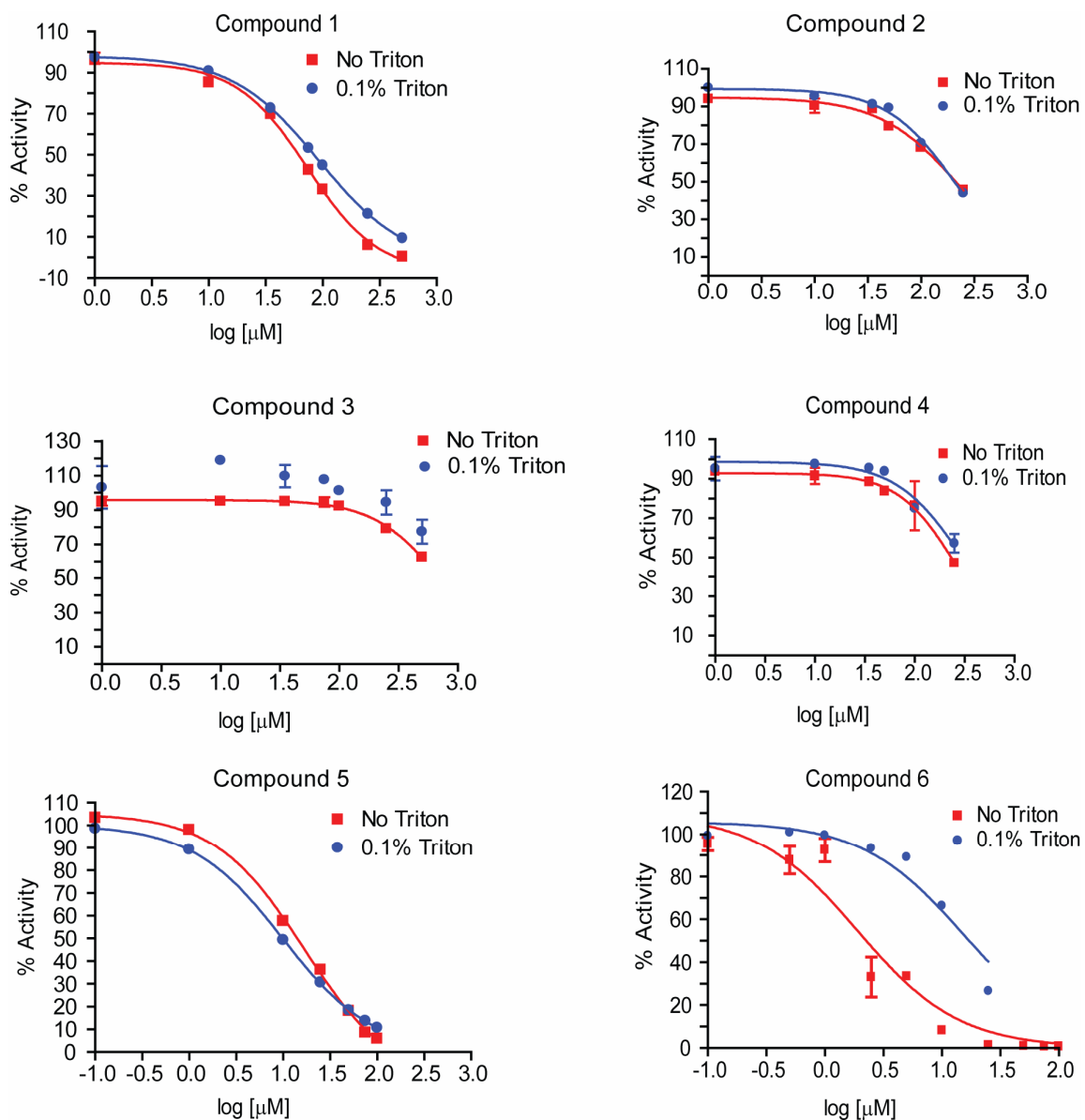
A.3 Chapter 3 Supporting Information

Table A.3.1: Full Target List

Id	Class	Description	Total Ligands	Clustered Ligands	Experi-mental Decoys	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
AA2AR	GPCR	Adenosine A2a receptor	3057	482	192	31,550	3EML	28	22	83
ABL1	Kinase	Tyrosine-protein kinase ABL	409	182	84	10,750	2HZI	26	19	81
ACE	Protease	Angiotensin-converting enzyme	749	282	55	16,900	3BKL	22	20	72
ACES	Other Enzymes	Acetylcholinesterase	1581	453	487	26,250	1E66	25	20	81
ADA	Other Enzymes	Adenosine deaminase	98	93	42	5,450	2E1W	24	22	76
ADA17	Protease	ADAM17	1341	532	31	35,900	2OI0	26	22	77
ADRB1	GPCR	Beta-1 adrenergic receptor	648	247	69	15,850	2VT4	19	11	76
ADRB2	GPCR	Beta-2 adrenergic receptor	602	231	118	15,000	3NY8	12	4	69
AKT1	Kinase	Serine/threonine-protein kinase AKT	585	293	53	16,450	3CQW	27	29	72
AKT2	Kinase	Serine/threonine-protein kinase AKT2	234	117	23	6,900	3D0E	16	8	77
ALDR	Other Enzymes	Aldose reductase	604	159	78	9,000	2HV5	31	33	76
AMPC	Other Enzymes	Beta-lactamase	48	48	84	2,850	1L2S	16	8	79
ANDR	Nuclear Receptor	Androgen Receptor	1046	269	19	14,350	2AM9	5	6	51
AOFB	Other Enzymes	Monoamine oxidase B	438	122	129	6,900	1S3B	17	8	75
BACE1	Protease	Beta-secretase 1	595	283	41	18,100	3L5D	11	8	66
BRAF	Kinase	Serine/threonine-protein kinase B-raf	317	152	28	9,950	3D4Q	22	11	83
CAH2	Other Enzymes	Carbonic anhydrase II	1924	492	27	31,172	1BCD	12	3	73
CASP3	Protease	Caspase-3	470	199	37	10,700	2CNK	16	12	72
CDK2	Kinase	Cyclin-dependent kinase 2	1310	474	136	27,850	1H00	21	14	79
COMT	Other Enzymes	Catechol O-methyltransferase	41	41	1	3,850	3BWM	34	22	89
CP2C9	Cytochrome P450	Cytochrome P450 2C9	145	120	176	7,450	1R9O	7	3	60
CP3A4	Cytochrome P450	Cytochrome P450 3A4	302	170	267	11,800	3NXU	7	2	63

Id	Class	Description	Total Ligands	Clustered Ligands	Experi-mental Decoys	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
CSF1R	Kinase	Macrophage colony stimulating factor receptor	385	166	5	12,150	3KRJ	32	25	86
CXCR4	GPCR	C-X-C chemokine receptor type 4	40	40	14	3,406	3ODU	36	18	90
DEF	Other Enzymes	Peptide deformylase	171	102	8	5,700	1LRU	23	22	69
DHI1	Other Enzymes	11-beta-hydroxysteroid dehydrogenase 1	836	387	29	19,350	3FRJ	11	6	68
DPP4	Protease	Dipeptidyl peptidase IV	1939	533	167	40,950	2I78	41	41	87
DRD3	GPCR	Dopamine D3 receptor	2113	480	14	34,050	3PBL	13	4	69
DYR	Other Enzymes	Dihydrofolate reductase	1082	231	117	17,200	3NXO	43	44	87
EGFR	Kinase	Epidermal growth factor receptor erbB1	1612	542	407	35,050	2RGP	29	22	84
ESR1	Nuclear Receptor	Estrogen receptor alpha	1297	383	136	20,685	1SJ0	18	15	67
ESR2	Nuclear Receptor	Estrogen receptor beta	1343	367	126	20,199	2FSZ	19	18	65
FA10	Protease	Coagulation factor X	3090	537	176	28,325	3KL6	39	36	87
FA7	Protease	Coagulation factor VII	303	114	39	6,250	1W7X	56	66	88
FABP4	Miscel-laneous	Fatty acid binding protein adipocyte	47	47	16	2,750	2NNQ	46	43	90
FAK1	Kinase	Focal adhesion kinase 1	101	100	11	5,350	3BZ3	24	20	75
FGFR1	Kinase	Fibroblast growth factor receptor 1	327	139	146	8,700	3C4F	18	13	73
FKB1A	Other Enzymes	FK506-binding protein 1A	159	111	68	5,800	1J4H	16	9	75
FNTA	Other Enzymes	Protein farnesyltransferase type I alpha subunit	1430	592	132	51,500	3E37	16	7	76
FPPS	Other Enzymes	Farnesyl diphosphate synthase	85	85	27	8,850	1ZW5	51	67	84
GCR	Nuclear Receptor	Glucocorticoid receptor	972	258	6	15,000	3BQD	4	9	44
GLCM	Other Enzymes	Beta-glucocerebrosidase	54	54	27	3,800	2V3F	30	24	81
GRIA2	Ion Channel	Glutamate receptor ionotropic, AMPA 2	476	158	201	11,845	3KGC	23	23	71
GRIK1	Ion Channel	Glutamate receptor ionotropic kainate 1	136	101	235	6,550	1VSO	35	27	86

Id	Class	Description	Total Ligands	Clustered Ligands	Experi-mental Decoys	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
HDAC2	Other Enzymes	Histone deacetylase 2	407	185	62	10,300	3MAX	24	20	77
HDAC8	Other Enzymes	Histone deacetylase 8	309	170	73	10,450	3F07	29	24	80
HIVINT	Other Enzymes	Human immunodeficiency virus type 1 integrase	167	100	268	6,650	3NF7	8	2	64
HIVPR	Protease	Human immunodeficiency virus type 1 protease	1468	536	96	35,750	1XL2	7	5	60
HIVRT	Other Enzymes	Human immunodeficiency virus type 1 reverse transcriptase	1178	338	258	18,891	3LAN	11	7	64
HMDH	Other Enzymes	HMG-CoA reductase	527	170	8	8,750	3CCW	26	24	74
HS90A	Miscel-laneous	Heat shock protein HSP 90-alpha	88	88	25	4,850	1UYG	15	3	69
HXK4	Other Enzymes	Hexokinase type IV	136	92	14	4,700	3F9M	25	17	80
IGF1R	Kinase	Insulin-like growth factor I receptor	370	148	75	9,300	2OJ9	18	5	79
INHA	Other Enzymes	Enoyl-[acyl-carrier-protein] reductase	44	44	22	2,300	2H7L	19	2	85
ITAL	Miscel-laneous	Leukocyte adhesion glycoprotein LFA-1 alpha	324	138	29	8,500	2ICA	15	14	66
JAK2	Kinase	Tyrosine-protein kinase JAK2	246	130	6	6,500	3LPB	29	23	82
KIF11	Miscel-laneous	Kinesin-like protein 1	272	116	29	6,850	3CJO	34	35	77
KIT	Kinase	Stem cell growth factor receptor	378	166	8	10,450	3G0E	12	4	73
KITH	Other Enzymes	Thymidine kinase	57	57	68	2,850	2B8T	15	0	80
KPCB	Kinase	Protein kinase C beta	331	135	153	8,700	2I0E	15	17	61
LCK	Kinase	Tyrosine-protein kinase LCK	916	420	148	27,400	2OF2	25	16	82
LKHA4	Protease	Leukotriene A4 hydrolase	343	171	21	9,450	3CHP	18	4	82
MAPK2	Kinase	MAP kinase-activated protein kinase 2	184	101	81	6,150	3M2W	33	27	86
MCR	Nuclear Receptor	Mineralocorticoid receptor	201	94	2	5,150	2AA2	-4	2	36

Id	Class	Description	Total Ligands	Clustered Ligands	Experi-mental Decoys	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
MET	Kinase	Hepatocyte growth factor receptor	333	166	17	11,250	3LQ8	24	16	80
MK01	Kinase	MAP kinase ERK2	79	79	35	4,550	2OJG	19	9	79
MK10	Kinase	c-Jun N-terminal kinase 3	199	104	23	6,600	2ZDT	24	11	82
MK14	Kinase	MAP kinase p38 alpha	2205	578	73	35,850	2QD9	17	10	74
MMP13	Protease	Matrix metalloproteinase 13	1632	572	26	37,200	830C	12	5	71
MP2K1	Kinase	Dual specificity mitogen-activated protein kinase kinase 1	308	121	12	8,150	3EQH	16	4	78
NOS1	Other Enzymes	Nitric-oxide synthase, brain	311	100	73	8,050	1QW6	19	11	78
NRAM	Other Enzymes	Neuraminidase	200	98	21	6,200	1B9V	44	46	93
PA2GA	Other Enzymes	Phospholipase A2 group IIA	173	99	14	5,150	1KVO	26	29	72
PARP1	Other Enzymes	Poly [ADP-ribose] polymerase-1	1031	508	12	30,050	3L3M	25	21	79
PDE5A	Other Enzymes	Phosphodiesterase 5A	958	398	93	27,550	1UDT	17	11	72
PGH1	Other Enzymes	Cyclooxygenase-1	534	195	1070	10,800	2OYU	3	5	53
PGH2	Other Enzymes	Cyclooxygenase-2	1707	435	444	23,150	3LN1	13	13	62
PLK1	Kinase	Serine/threonine-protein kinase PLK1	227	107	46	6,800	2OWB	28	19	84
PNPH	Other Enzymes	Purine nucleoside phosphorylase	225	103	10	6,950	3BGS	39	42	81
PPARA	Nuclear Receptor	Peroxisome proliferator-activated receptor alpha	1040	373	82	19,399	2P54	19	5	81
PPARD	Nuclear Receptor	Peroxisome proliferator-activated receptor delta	699	240	79	12,250	2ZNP	32	20	89
PPARG	Nuclear Receptor	Peroxisome proliferator-activated receptor gamma	1245	484	43	25,300	2GTK	21	12	80
PRGR	Nuclear Receptor	Progesterone receptor	1114	293	1	15,650	3KBA	8	8	56
PTN1	Other Enzymes	Protein-tyrosine phosphatase 1B	264	130	153	7,250	2AZR	36	30	84
PUR2	Other Enzymes	GAR transformylase	50	50	12	2,700	1NJS	51	50	92
PYGM	Other Enzymes	Muscle glycogen phosphorylase	77	77	7	3,950	1C8K	17	9	72

Id	Class	Description	Total Ligands	Clustered Ligands	Experi-mental Decoys	Matched Decoys	PDB	Log AUC (%)	ROC EF1	AUC (%)
PYRD	Other Enzymes	Dihydroorotate dehydrogenase	226	111	88	6,450	1D3G	30	28	83
RENI	Protease	Renin	391	104	46	6,958	3G6Z	21	19	66
ROCK1	Kinase	Rho-associated protein kinase 1	216	100	15	6,300	2ETR	17	10	74
RXRA	Nuclear Receptor	Retinoid X receptor alpha	298	131	7	6,950	1MV9	25	24	71
SAHH	Other Enzymes	Adenosylhomocystei nase	63	63	5	3,450	1LI4	39	33	85
SRC	Kinase	Tyrosine-protein kinase SRC	1269	524	287	34,500	3EL8	18	9	77
TGFR1	Kinase	TGF-beta receptor type I	235	133	7	8,500	3HMM	36	28	88
THB	Nuclear Receptor	Thyroid hormone receptor beta-1	246	103	29	7,450	1Q4X	36	38	79
THRB	Protease	Thrombin	2109	461	255	27,004	1YPE	32	30	81
TRY1	Protease	Trypsin I	924	449	117	25,980	2AYW	57	62	93
TRYB1	Protease	Tryptase beta-1	216	148	16	7,650	2ZEC	37	31	87
TYSY	Other Enzymes	Thymidylate synthase	390	109	63	6,750	1SYN	28	21	82
UROK	Protease	Urokinase-type plasminogen activator	372	162	44	9,850	1SQT	65	72	95
VGFR2	Kinase	Vascular endothelial growth factor receptor 2	2320	409	142	24,950	2P2I	21	12	79
WEE1	Kinase	Serine/threonine-protein kinase WEE1	221	102	15	6,150	3BIZ	52	55	91
XIAP	Miscel-laneous	Inhibitor of apoptosis protein 3	100	100	7	5,150	3HL5	52	55	88

Table A.3.2: Effect of Clustering on Enrichment

Target	Raw Ligands	Clustered Ligands	Clustered Percentage	Raw LogAUC (%)	Clustered LogAUC (%)
TRYB1	216	148	69	43	37
CP3A4	302	170	56	5	7
HDAC8	309	170	55	25	29
...			...		
VGFR2	2320	409	18	16	21
FA10	3090	537	17	37	39
AA2AR	3057	482	16	24	28

Table A.3.3: Ligand and Decoy Property Distribution

Average (Standard Deviation)	Molecular Weight [Daltons]		logP		Rotable Bonds		Hydrogen Bond Acceptors		Hydrogen Bond Donors		Net Charge	
Target	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs
AA2AR	409.8 (74.9)	401.9 (71.8)	2.4 (1.3)	2.4 (1.4)	5.1 (2.4)	5.4 (2.3)	8.8 (2.4)	7.8 (1.9)	2.9 (1.6)	2.5 (1.3)	0.4 (0.7)	0.3 (0.6)
ABL1	465.9 (79.7)	446.8 (69.1)	4.3 (1.3)	4.0 (1.3)	5.8 (2.0)	5.9 (2.2)	7.2 (1.5)	6.6 (1.4)	2.6 (1.2)	2.3 (1.0)	0.5 (0.6)	0.5 (0.6)
ACE	416.0 (80.7)	390.6 (77.8)	0.8 (1.5)	1.8 (1.6)	8.2 (2.9)	6.5 (2.6)	6.8 (1.8)	7.4 (1.8)	1.5 (1.3)	0.9 (1.0)	-1.2 (0.5)	-1.1 (0.5)
ACES	443.8 (91.3)	415.8 (76.4)	4.5 (2.6)	3.9 (1.8)	8.5 (3.8)	7.0 (2.9)	5.4 (1.8)	5.4 (1.9)	2.4 (1.5)	2.0 (1.1)	1.2 (0.7)	0.9 (0.6)
ADA	324.1 (74.6)	323.9 (72.1)	2.0 (1.4)	2.0 (1.3)	7.1 (2.5)	6.2 (2.1)	6.5 (1.4)	5.9 (1.4)	3.3 (1.1)	3.2 (0.9)	0.3 (0.5)	0.3 (0.5)
ADA17	475.8 (62.8)	451.3 (52.5)	2.6 (1.2)	2.9 (1.4)	7.0 (1.9)	6.8 (2.0)	8.6 (1.6)	8.0 (1.5)	2.6 (1.2)	2.3 (1.1)	-0.0 (0.5)	-0.0 (0.6)
ADRB1	443.4 (86.5)	417.1 (73.7)	3.4 (1.5)	3.2 (1.5)	9.6 (3.3)	7.9 (2.6)	6.6 (2.0)	6.6 (2.1)	3.9 (1.4)	3.2 (1.1)	0.6 (0.7)	0.5 (0.6)
ADRB2	453.4 (98.4)	421.1 (79.6)	3.3 (1.5)	3.2 (1.6)	9.9 (3.5)	8.1 (2.8)	6.9 (2.3)	6.7 (2.2)	4.0 (1.6)	3.1 (1.2)	0.6 (0.6)	0.5 (0.6)
AKT1	449.5 (71.6)	422.2 (67.1)	3.7 (1.5)	3.2 (1.6)	6.4 (1.8)	6.3 (2.1)	6.9 (1.9)	6.4 (1.8)	4.0 (1.4)	3.3 (1.2)	0.9 (0.6)	0.9 (0.6)
AKT2	450.9 (88.8)	424.0 (81.9)	3.2 (1.9)	3.0 (1.9)	5.3 (1.9)	5.9 (2.1)	7.6 (1.9)	6.7 (1.8)	4.3 (1.7)	3.5 (1.4)	0.8 (0.5)	0.8 (0.5)
ALDR	351.7 (65.9)	340.7 (65.1)	2.4 (1.5)	2.4 (1.5)	3.6 (2.2)	4.0 (1.9)	6.0 (1.9)	5.7 (1.6)	0.8 (1.2)	0.7 (1.1)	-0.8 (0.5)	-0.8 (0.5)
AMPC	293.4 (63.7)	294.9 (65.3)	1.8 (1.4)	1.9 (1.4)	3.8 (1.4)	3.7 (1.6)	5.8 (1.4)	5.7 (1.4)	0.6 (0.7)	0.6 (0.6)	-1.5 (0.6)	-1.5 (0.5)
ANDR	368.1 (71.6)	358.1 (71.6)	3.8 (1.4)	3.8 (1.4)	3.0 (2.6)	3.4 (2.0)	4.2 (2.0)	4.1 (1.9)	0.9 (0.8)	0.9 (0.8)	0.0 (0.3)	0.0 (0.3)
AOFB	278.4 (67.3)	277.2 (64.7)	2.9 (1.3)	2.9 (1.2)	3.6 (2.3)	3.3 (1.9)	3.7 (1.6)	3.7 (1.5)	1.2 (1.1)	1.2 (1.0)	0.3 (0.5)	0.3 (0.5)
BACE1	508.9 (68.9)	466.6 (53.4)	4.0 (1.4)	3.9 (1.5)	8.6 (3.8)	7.4 (2.8)	7.1 (1.4)	6.9 (1.7)	3.8 (1.1)	2.9 (0.9)	0.7 (0.5)	0.6 (0.5)
BRAF	456.6 (83.4)	439.0 (69.6)	4.0 (1.3)	3.8 (1.3)	5.8 (2.2)	5.8 (2.3)	7.2 (1.6)	6.6 (1.5)	2.7 (1.2)	2.4 (1.0)	0.4 (0.6)	0.4 (0.6)
CAH2	406.6 (97.3)	382.0 (83.0)	0.9 (1.8)	1.5 (1.6)	6.0 (2.9)	5.6 (2.4)	8.1 (2.6)	7.4 (2.1)	3.5 (1.6)	2.8 (1.3)	-0.1 (0.8)	-0.0 (0.7)
CASP3	461.8 (80.6)	437.3 (66.1)	1.9 (1.6)	2.5 (1.5)	9.1 (4.2)	7.7 (3.1)	9.1 (2.5)	8.2 (1.9)	1.9 (1.3)	1.4 (1.3)	-0.4 (0.8)	-0.2 (0.7)
CDK2	394.9 (74.0)	386.0 (72.1)	2.9 (1.2)	2.8 (1.3)	4.8 (2.0)	5.1 (2.1)	7.2 (1.8)	6.6 (1.7)	3.0 (1.4)	2.8 (1.2)	0.2 (0.6)	0.2 (0.6)
COMT	291.8 (76.7)	300.0 (74.8)	2.3 (1.2)	2.1 (1.2)	4.1 (2.1)	3.8 (1.9)	7.0 (1.6)	6.2 (1.6)	1.8 (0.9)	1.6 (0.9)	-0.5 (0.6)	-0.5 (0.6)
CP2C9	422.5 (107.8)	406.9 (93.1)	4.1 (1.4)	4.0 (1.3)	6.0 (2.6)	5.6 (2.5)	5.7 (2.0)	5.6 (2.0)	1.5 (1.3)	1.4 (1.2)	-0.1 (0.7)	-0.1 (0.7)
CP3A4	449.7 (97.1)	428.6 (83.1)	4.0 (1.5)	3.9 (1.4)	6.6 (3.4)	6.2 (2.9)	6.5 (2.3)	6.1 (2.0)	2.0 (1.6)	1.7 (1.3)	0.4 (0.6)	0.4 (0.6)
CSF1R	436.9 (66.6)	424.4 (59.5)	3.5 (1.1)	3.3 (1.2)	5.3 (2.0)	5.4 (2.2)	7.3 (1.8)	6.7 (1.6)	2.7 (1.4)	2.4 (1.1)	0.4 (0.7)	0.4 (0.7)
CXCR4	394.0 (65.2)	368.3 (58.6)	2.4 (1.4)	2.8 (1.3)	6.1 (1.5)	6.0 (1.9)	5.1 (1.2)	5.1 (1.6)	3.8 (1.9)	3.4 (1.3)	1.8 (0.9)	1.6 (0.6)
DEF	383.7 (83.9)	375.2 (76.6)	2.5 (1.2)	2.6 (1.2)	7.1 (3.3)	6.4 (2.9)	6.7 (1.4)	6.5 (1.4)	2.5 (0.9)	2.4 (0.8)	0.1 (0.4)	0.1 (0.4)
DHI1	392.8 (77.3)	384.3 (72.9)	3.8 (1.1)	3.9 (1.1)	4.6 (1.8)	4.4 (1.8)	4.8 (1.5)	4.8 (1.5)	1.0 (1.0)	0.9 (1.0)	0.0 (0.5)	0.0 (0.5)
DPP4	377.4 (63.5)	362.3 (65.9)	1.6 (1.4)	1.9 (1.4)	4.5 (1.8)	4.9 (1.9)	6.2 (1.5)	6.0 (1.6)	2.8 (1.0)	2.7 (0.9)	0.7 (0.6)	0.7 (0.6)
DRD3	418.4 (73.9)	406.0 (71.5)	4.4 (1.2)	4.2 (1.2)	6.4 (2.4)	6.1 (2.4)	4.8 (1.5)	4.8 (1.6)	1.7 (0.9)	1.6 (0.9)	0.8 (0.5)	0.8 (0.5)
DYR	381.7 (82.7)	361.1 (81.8)	1.9 (2.1)	1.9 (1.7)	5.5 (3.1)	5.3 (2.3)	7.8 (2.8)	6.6 (2.5)	5.1 (1.2)	4.1 (1.1)	0.1 (1.1)	0.3 (0.8)

Average (Standard Deviation)	Molecular Weight [Daltons]		logP		Rotable Bonds		Hydrogen Bond Acceptors		Hydrogen Bond Donors		Net Charge	
Target	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs
EGFR	449.0 (82.6)	434.1 (71.2)	4.2 (1.2)	3.9 (1.2)	6.7 (2.6)	6.2 (2.5)	7.1 (1.9)	6.6 (1.7)	2.4 (1.2)	2.2 (1.1)	0.4 (0.6)	0.4 (0.6)
ESR1	421.0 (93.2)	404.6 (84.5)	5.1 (1.4)	4.7 (1.3)	5.2 (3.4)	5.3 (2.8)	4.6 (1.8)	4.4 (1.8)	2.3 (1.0)	2.1 (0.9)	0.5 (0.6)	0.5 (0.6)
ESR2	410.2 (95.3)	394.6 (86.8)	4.9 (1.5)	4.5 (1.4)	4.9 (3.3)	5.2 (2.8)	4.6 (1.9)	4.4 (1.9)	2.3 (1.0)	2.1 (0.8)	0.5 (0.6)	0.5 (0.6)
FA10	513.5 (48.8)	470.9 (48.4)	2.9 (1.6)	3.1 (1.8)	7.1 (2.1)	6.9 (2.2)	8.6 (1.9)	7.9 (1.9)	3.9 (2.3)	2.9 (1.6)	0.7 (0.7)	0.5 (0.7)
FA7	472.1 (66.8)	427.5 (67.7)	2.9 (1.8)	2.4 (2.0)	8.2 (2.5)	6.9 (2.5)	8.1 (1.9)	7.8 (2.3)	6.2 (1.6)	4.8 (1.3)	0.6 (0.8)	0.5 (0.8)
FABP4	390.7 (78.6)	393.8 (79.5)	5.5 (1.4)	5.1 (1.2)	6.3 (2.9)	5.7 (2.6)	4.1 (1.0)	4.4 (1.3)	0.5 (0.8)	0.5 (0.8)	-0.9 (0.4)	-0.8 (0.4)
FAK1	438.1 (58.0)	433.1 (53.7)	3.4 (0.8)	3.3 (0.8)	7.0 (2.1)	6.6 (2.0)	8.3 (1.6)	7.7 (1.4)	2.4 (1.2)	2.2 (1.1)	0.2 (0.6)	0.2 (0.6)
FGFR1	448.7 (78.7)	430.0 (70.4)	3.7 (1.6)	3.4 (1.5)	5.7 (3.1)	5.9 (2.6)	7.7 (1.5)	6.8 (1.5)	3.2 (1.2)	2.8 (1.1)	0.5 (0.6)	0.5 (0.6)
FKB1A	440.4 (65.2)	429.7 (58.4)	4.0 (1.4)	4.0 (1.4)	8.4 (2.8)	7.7 (2.3)	6.2 (1.9)	6.0 (1.7)	0.2 (0.5)	0.3 (0.5)	-0.1 (0.2)	-0.1 (0.2)
FNTA	480.8 (58.7)	454.5 (49.8)	3.5 (1.7)	3.8 (1.5)	7.0 (3.4)	6.6 (2.6)	6.6 (1.5)	6.6 (1.7)	1.5 (1.3)	1.3 (1.1)	0.3 (0.7)	0.3 (0.7)
FPPS	292.0 (40.6)	311.2 (69.8)	-1.0 (1.7)	0.3 (2.1)	6.0 (3.1)	5.6 (2.5)	7.5 (0.8)	7.3 (1.6)	2.7 (1.2)	1.3 (1.3)	-2.1 (0.9)	-1.7 (0.6)
GCR	440.0 (65.5)	425.5 (60.8)	5.2 (1.2)	5.0 (1.2)	4.3 (2.2)	4.6 (2.0)	4.4 (1.6)	4.4 (1.7)	1.3 (0.9)	1.2 (0.8)	0.1 (0.5)	0.1 (0.5)
GLCM	352.9 (93.4)	344.7 (93.6)	1.9 (2.2)	1.8 (1.9)	7.7 (4.8)	6.5 (3.3)	6.5 (1.9)	6.3 (2.2)	4.8 (2.1)	4.0 (1.5)	0.6 (0.6)	0.5 (0.6)
GRIA2	362.7 (89.6)	353.8 (81.3)	1.0 (2.1)	1.3 (2.0)	5.2 (4.4)	5.1 (2.9)	8.2 (2.4)	7.3 (2.0)	2.5 (2.2)	1.9 (1.6)	-0.3 (1.5)	-0.4 (1.2)
GRIK1	313.4 (94.5)	314.9 (81.3)	-0.6 (2.0)	-0.1 (1.8)	4.5 (2.4)	4.6 (2.0)	7.9 (2.3)	7.5 (1.7)	2.5 (1.1)	2.0 (1.1)	-0.8 (0.6)	-0.8 (0.6)
HDAC2	390.2 (84.6)	381.9 (75.9)	3.3 (1.3)	3.2 (1.2)	7.9 (2.7)	6.9 (2.4)	6.3 (1.8)	6.0 (1.8)	3.1 (1.2)	2.8 (1.1)	0.2 (0.5)	0.2 (0.5)
HDAC8	384.4 (80.1)	376.2 (71.6)	3.3 (1.3)	3.2 (1.3)	7.4 (2.4)	6.6 (2.2)	6.1 (1.8)	5.8 (1.7)	2.9 (1.1)	2.7 (1.0)	0.3 (0.6)	0.3 (0.6)
HIVINT	383.1 (78.0)	371.2 (69.8)	1.9 (1.4)	2.1 (1.3)	5.2 (2.5)	4.9 (2.0)	7.2 (2.3)	6.7 (2.0)	2.1 (1.9)	1.9 (1.6)	-0.6 (0.9)	-0.5 (0.7)
HIVPR	512.2 (75.6)	472.2 (57.8)	5.1 (1.5)	4.6 (1.5)	9.3 (3.2)	7.8 (2.7)	6.9 (2.2)	6.9 (2.1)	2.3 (1.8)	1.7 (1.4)	-0.3 (0.7)	-0.3 (0.7)
HIVRT	357.9 (68.6)	352.4 (66.1)	3.2 (1.8)	3.2 (1.6)	4.3 (2.5)	4.2 (2.1)	5.5 (2.5)	5.2 (2.0)	1.5 (1.2)	1.4 (1.2)	-0.0 (0.7)	0.0 (0.5)
HMDH	471.8 (69.5)	446.3 (55.1)	4.3 (1.5)	4.2 (1.6)	9.0 (2.8)	7.6 (2.5)	6.1 (2.0)	6.5 (2.0)	2.0 (1.0)	1.3 (0.8)	-0.6 (0.5)	-0.6 (0.5)
HS90A	431.5 (76.7)	418.4 (63.0)	3.6 (1.3)	3.6 (1.1)	5.2 (1.9)	5.3 (2.0)	7.5 (1.8)	6.7 (1.8)	3.1 (1.2)	2.7 (0.9)	0.1 (0.4)	0.1 (0.4)
HXK4	419.7 (53.8)	417.0 (53.1)	3.3 (1.0)	3.3 (1.0)	6.2 (1.8)	5.7 (1.7)	7.1 (1.2)	6.9 (1.2)	1.4 (0.7)	1.4 (0.7)	-0.1 (0.4)	-0.1 (0.4)
IGF1R	498.5 (62.9)	464.4 (52.0)	4.3 (1.4)	3.7 (1.5)	6.6 (1.9)	6.6 (2.2)	8.3 (1.8)	7.1 (1.7)	3.8 (1.4)	3.0 (1.1)	0.6 (0.6)	0.5 (0.6)
INHA	344.4 (55.8)	347.6 (55.7)	5.1 (0.8)	4.8 (0.8)	6.1 (2.5)	5.2 (2.0)	4.0 (1.1)	3.6 (1.1)	1.2 (0.8)	1.1 (0.7)	-0.0 (0.4)	-0.0 (0.4)
ITAL	516.3 (57.4)	486.0 (46.9)	4.8 (1.5)	4.9 (1.4)	6.6 (2.0)	6.7 (2.1)	6.1 (1.9)	6.3 (1.8)	0.8 (0.9)	0.8 (0.8)	-0.1 (0.6)	-0.1 (0.6)
JAK2	417.1 (81.5)	407.9 (75.2)	3.0 (1.1)	3.0 (1.2)	5.0 (2.2)	5.2 (2.0)	7.5 (1.8)	7.0 (1.6)	2.2 (1.2)	2.1 (1.1)	0.4 (0.7)	0.4 (0.7)
KIF11	406.4 (68.3)	394.6 (67.0)	3.8 (0.9)	3.8 (1.0)	5.1 (2.3)	5.0 (2.1)	4.8 (1.4)	4.7 (1.5)	1.9 (1.3)	1.9 (1.2)	0.5 (0.6)	0.5 (0.6)
KIT	459.9 (69.7)	440.5 (61.8)	3.9 (1.2)	3.6 (1.2)	6.1 (2.3)	6.1 (2.2)	7.7 (1.7)	6.8 (1.6)	3.3 (1.3)	2.8 (1.1)	0.5 (0.6)	0.5 (0.6)
KITH	418.9 (91.1)	403.7 (82.1)	2.0 (1.7)	2.1 (1.6)	5.3 (2.0)	5.7 (1.9)	7.7 (1.3)	7.4 (1.4)	2.8 (0.6)	2.7 (0.7)	0.1 (0.3)	0.1 (0.3)
KPCB	466.6 (72.8)	437.7 (59.9)	3.9 (1.2)	3.6 (1.4)	4.7 (3.6)	5.2 (2.6)	7.5 (2.0)	6.6 (1.8)	2.8 (1.9)	2.2 (1.5)	0.3 (0.9)	0.3 (0.7)

Average (Standard Deviation)	Molecular Weight [Daltons]		logP		Rotable Bonds		Hydrogen Bond Acceptors		Hydrogen Bond Donors		Net Charge	
Target	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs
LCK	465.8 (80.6)	443.3 (65.9)	4.5 (1.4)	4.0 (1.4)	5.9 (2.3)	6.0 (2.3)	7.5 (1.8)	6.6 (1.6)	3.0 (1.4)	2.5 (1.1)	0.5 (0.7)	0.5 (0.6)
LKHA4	372.8 (73.8)	370.9 (76.0)	4.0 (1.7)	3.9 (1.6)	7.5 (2.1)	6.5 (2.1)	4.9 (1.8)	4.8 (1.8)	1.7 (1.1)	1.6 (1.1)	0.4 (0.6)	0.4 (0.6)
MAPK2	374.9 (74.1)	362.7 (73.2)	2.5 (1.8)	2.4 (1.7)	3.2 (1.8)	4.0 (1.8)	6.3 (1.5)	5.8 (1.6)	3.2 (1.3)	2.9 (1.1)	0.2 (0.6)	0.2 (0.6)
MCR	411.4 (46.3)	405.3 (49.4)	4.8 (1.3)	4.6 (1.2)	4.6 (2.4)	4.6 (1.9)	4.7 (1.5)	4.6 (1.5)	1.1 (1.0)	1.1 (1.0)	-0.2 (0.5)	-0.2 (0.5)
MET	473.7 (77.9)	454.6 (67.6)	3.8 (1.3)	3.7 (1.3)	6.0 (2.2)	6.1 (2.3)	7.8 (1.7)	7.1 (1.6)	2.2 (1.3)	2.0 (1.2)	0.4 (0.6)	0.4 (0.6)
MK01	415.0 (73.5)	402.8 (74.1)	3.3 (1.8)	3.1 (1.7)	5.1 (2.6)	5.3 (2.3)	6.9 (1.6)	6.6 (1.6)	3.4 (1.1)	3.2 (1.1)	0.0 (0.4)	0.0 (0.4)
MK10	411.9 (71.2)	404.0 (66.5)	3.5 (1.4)	3.5 (1.3)	5.5 (2.0)	5.2 (1.9)	6.7 (1.6)	6.4 (1.5)	2.3 (1.2)	2.1 (1.1)	0.3 (0.6)	0.3 (0.6)
MK14	445.8 (71.5)	430.3 (62.9)	4.5 (1.4)	4.1 (1.3)	5.8 (2.0)	5.8 (2.2)	6.4 (1.9)	6.0 (1.7)	2.4 (1.2)	2.1 (1.0)	0.4 (0.6)	0.4 (0.6)
MMP13	473.5 (62.5)	449.8 (53.1)	2.6 (1.4)	2.8 (1.5)	7.5 (2.2)	7.0 (2.2)	8.6 (1.7)	8.0 (1.5)	2.0 (0.9)	1.8 (1.0)	-0.2 (0.6)	-0.2 (0.6)
MP2K1	454.5 (80.1)	436.6 (78.1)	4.1 (1.6)	3.9 (1.6)	6.8 (2.9)	6.6 (2.6)	6.8 (1.7)	6.5 (1.5)	2.8 (1.7)	2.5 (1.4)	0.4 (0.6)	0.4 (0.6)
NOS1	308.2 (96.1)	304.5 (87.4)	2.7 (1.7)	2.5 (1.6)	4.4 (2.7)	4.3 (2.2)	4.1 (1.7)	4.0 (1.9)	3.8 (1.7)	3.3 (1.2)	1.1 (0.7)	1.1 (0.7)
NRAM	343.8 (51.2)	333.6 (61.6)	-0.3 (1.2)	0.2 (1.3)	6.3 (2.1)	5.4 (2.1)	7.5 (1.4)	7.5 (1.6)	4.3 (1.5)	3.9 (1.3)	-0.1 (0.5)	-0.1 (0.5)
PA2GA	444.7 (76.0)	431.2 (65.0)	3.9 (1.8)	3.7 (1.5)	9.7 (4.0)	8.2 (3.3)	6.7 (1.3)	6.7 (1.5)	1.6 (0.9)	1.5 (0.8)	-0.7 (0.6)	-0.6 (0.6)
PARP1	361.3 (86.4)	350.9 (84.4)	2.3 (1.3)	2.3 (1.3)	3.6 (1.8)	4.1 (1.9)	5.9 (1.9)	5.6 (1.9)	2.4 (1.3)	2.4 (1.2)	0.4 (0.6)	0.4 (0.6)
PDE5A	454.5 (82.4)	439.8 (68.8)	3.5 (1.2)	3.4 (1.2)	6.5 (2.8)	6.2 (2.5)	8.3 (2.1)	7.5 (1.8)	1.4 (1.0)	1.4 (1.0)	0.2 (0.7)	0.2 (0.7)
PGH1	341.8 (64.8)	340.9 (63.8)	4.3 (1.5)	4.0 (1.4)	3.6 (1.7)	3.6 (1.6)	4.1 (1.7)	4.0 (1.7)	1.1 (1.1)	1.0 (1.0)	-0.3 (0.5)	-0.2 (0.5)
PGH2	375.3 (68.7)	369.4 (67.8)	4.0 (1.4)	3.9 (1.4)	4.2 (1.8)	4.0 (1.7)	4.5 (1.6)	4.5 (1.6)	0.8 (0.9)	0.7 (0.9)	-0.2 (0.4)	-0.2 (0.4)
PLK1	464.3 (73.6)	447.6 (61.2)	3.8 (1.4)	3.7 (1.4)	6.8 (1.9)	6.3 (1.9)	7.4 (1.6)	6.9 (1.4)	2.9 (1.2)	2.5 (0.9)	0.4 (0.7)	0.4 (0.7)
PNPH	277.7 (44.9)	273.8 (50.9)	0.1 (1.7)	0.3 (1.7)	3.3 (1.9)	4.1 (1.8)	7.0 (1.8)	6.2 (1.6)	4.4 (1.7)	3.8 (1.6)	-0.2 (0.9)	-0.2 (0.9)
PPARA	472.4 (58.6)	460.9 (52.8)	5.6 (1.3)	5.0 (1.1)	10.2 (2.0)	8.4 (2.1)	6.1 (1.4)	6.5 (1.5)	0.3 (0.6)	0.5 (0.6)	-0.8 (0.4)	-0.8 (0.4)
PPARD	484.5 (56.5)	463.2 (45.9)	5.8 (1.5)	5.1 (1.2)	9.9 (1.6)	8.1 (1.8)	5.8 (1.5)	6.4 (1.4)	0.2 (0.5)	0.4 (0.6)	-0.9 (0.3)	-0.9 (0.3)
PPARG	463.4 (59.9)	451.8 (53.0)	5.4 (1.5)	4.8 (1.3)	9.5 (2.3)	7.9 (2.3)	6.1 (1.5)	6.4 (1.5)	0.4 (0.6)	0.6 (0.6)	-0.8 (0.5)	-0.7 (0.5)
PRGR	368.7 (72.6)	360.8 (71.1)	4.8 (1.2)	4.6 (1.2)	2.9 (2.1)	3.5 (1.8)	3.6 (1.3)	3.6 (1.3)	1.0 (0.7)	0.9 (0.6)	0.0 (0.3)	0.0 (0.3)
PTN1	483.2 (86.9)	445.4 (67.2)	4.9 (2.6)	4.3 (2.1)	7.9 (3.3)	6.8 (2.8)	6.4 (2.3)	6.5 (2.1)	1.1 (1.0)	0.7 (0.9)	-1.2 (0.8)	-1.0 (0.7)
PUR2	463.5 (27.4)	418.1 (50.5)	-1.1 (0.9)	0.7 (1.7)	10.2 (1.4)	7.2 (1.9)	11.3 (1.0)	9.6 (1.9)	5.4 (0.8)	3.5 (1.1)	-1.9 (0.3)	-0.9 (0.6)
PYGM	412.2 (75.1)	397.9 (68.5)	3.3 (1.4)	3.1 (1.3)	5.4 (2.3)	5.2 (2.2)	7.1 (1.7)	6.9 (1.6)	1.6 (1.3)	1.5 (1.4)	-0.9 (1.1)	-0.9 (1.1)
PYRD	377.6 (60.1)	369.6 (60.9)	4.1 (1.2)	3.9 (1.1)	4.5 (1.6)	4.5 (1.6)	4.6 (1.3)	4.7 (1.3)	0.7 (0.7)	0.7 (0.7)	-0.7 (0.4)	-0.7 (0.4)
RENI	534.3 (58.7)	480.5 (47.2)	4.3 (1.7)	4.0 (1.9)	13.6 (3.6)	10.8 (2.6)	8.1 (2.1)	7.7 (2.2)	4.1 (1.8)	2.6 (1.4)	0.5 (0.7)	0.3 (0.6)
ROCK1	360.6 (67.2)	353.3 (67.9)	2.6 (1.4)	2.5 (1.4)	4.4 (1.5)	4.8 (1.9)	6.6 (2.2)	5.9 (1.9)	3.4 (1.2)	3.1 (1.0)	0.8 (0.7)	0.8 (0.7)
RXRA	407.8 (66.8)	412.4 (58.6)	6.4 (1.2)	5.5 (0.9)	5.4 (2.3)	5.0 (2.1)	3.7 (1.6)	4.7 (1.5)	0.2 (0.5)	0.4 (0.6)	-0.9 (0.4)	-0.8 (0.4)
SAHH	270.6 (41.7)	275.4 (48.9)	-0.3 (0.9)	-0.0 (1.0)	2.5 (1.4)	3.5 (1.4)	7.9 (1.3)	6.9 (1.4)	3.9 (0.7)	3.6 (0.8)	-0.1 (0.4)	-0.1 (0.4)

Average (Standard Deviation)	Molecular Weight [Daltons]		logP		Rotable Bonds		Hydrogen Bond Acceptors		Hydrogen Bond Donors		Net Charge	
Target	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs	Ligs	Decs
SRC	482.3 (73.6)	457.8 (61.8)	4.7 (1.6)	4.3 (1.5)	6.5 (2.4)	6.3 (2.4)	7.6 (1.8)	6.7 (1.6)	2.4 (1.3)	2.0 (1.1)	0.4 (0.8)	0.4 (0.7)
TGFR1	379.0 (59.8)	374.8 (59.5)	3.2 (1.0)	3.2 (1.0)	4.2 (1.7)	4.1 (1.7)	6.1 (1.4)	5.8 (1.3)	1.5 (1.1)	1.5 (1.0)	0.2 (0.5)	0.2 (0.5)
THB	461.0 (69.5)	442.6 (58.3)	4.6 (1.2)	4.3 (1.2)	6.0 (2.2)	6.0 (2.1)	6.8 (2.5)	6.5 (2.1)	1.5 (0.9)	1.1 (0.9)	-0.5 (0.7)	-0.5 (0.7)
THRB	489.9 (58.2)	436.4 (53.4)	1.2 (1.7)	1.9 (1.8)	9.2 (2.7)	7.3 (2.4)	9.2 (1.8)	8.4 (2.0)	5.3 (1.9)	3.9 (1.4)	0.7 (0.7)	0.4 (0.7)
TRY1	481.5 (67.6)	423.8 (65.6)	1.6 (1.8)	1.8 (1.9)	8.6 (3.0)	6.9 (2.7)	8.9 (2.3)	8.1 (2.3)	6.1 (1.9)	4.5 (1.5)	1.1 (0.7)	0.6 (0.7)
TRYB1	488.6 (62.7)	449.7 (56.7)	1.6 (2.1)	2.0 (1.8)	8.8 (3.3)	7.9 (2.6)	9.2 (2.0)	8.5 (2.0)	4.1 (2.2)	3.1 (1.8)	0.6 (0.8)	0.4 (0.7)
TYSY	452.2 (93.4)	407.0 (79.6)	0.7 (2.3)	1.9 (1.9)	7.1 (3.3)	5.8 (2.5)	9.4 (3.0)	7.8 (2.4)	3.0 (1.4)	2.0 (1.5)	-1.0 (1.3)	-0.5 (0.9)
UROK	403.6 (80.1)	375.0 (79.8)	2.8 (1.8)	2.2 (1.8)	5.8 (2.8)	5.5 (2.4)	6.5 (2.4)	6.4 (2.5)	5.5 (1.5)	4.4 (1.2)	0.8 (0.8)	0.6 (0.8)
VGFR2	449.9 (69.8)	432.0 (62.9)	4.0 (1.5)	3.6 (1.4)	5.9 (2.2)	6.0 (2.2)	7.3 (1.7)	6.6 (1.6)	2.7 (1.4)	2.3 (1.1)	0.5 (0.6)	0.5 (0.6)
WEE1	474.6 (73.4)	453.8 (66.8)	4.1 (1.1)	4.0 (1.2)	5.1 (2.1)	5.7 (2.0)	7.3 (1.2)	6.6 (1.3)	2.6 (0.8)	2.4 (0.8)	0.2 (0.6)	0.2 (0.6)
XIAP	484.1 (73.7)	440.9 (59.0)	1.6 (1.8)	2.2 (1.6)	8.5 (2.4)	7.4 (2.2)	8.0 (1.9)	7.9 (2.3)	5.1 (1.7)	4.0 (1.2)	0.8 (0.7)	0.6 (0.7)

Table A.3.4: Per Target Docking Enrichments

Target	None	Full	SEV	Thin	Drug-like*
Average	*20.6*	*14.3*	*24.4*	*24.9*	*26.8*
AA2AR	20	25	28	3	25
ABL1	27	13	26	23	27
ACE	18	16	22	18	38
ACES	18	15	25	23	27
ADA	13	27	24	16	29
ADA17	25	13	26	26	33
ADRB1	23	3	19	12	20
ADRB2	19	1	12	11	13
AKT1	27	6	27	27	36
AKT2	21	4	16	20	24
ALDR	30	18	31	31	34
AMPC	20	6	16	17	26
ANDR	3	4	5	5	20
AOFB	19	16	17	18	33
BACE1	10	6	11	9	15
BRAF	22	11	22	22	25
CAH2	12	7	12	9	15
CASP3	11	5	16	16	19
CDK2	14	15	21	22	22
COMT	19	30	34	31	38
CP2C9	6	6	7	8	9
CP3A4	6	8	7	7	4

Target	None	Full	SEV	Thin	Drug-like*
CSF1R	21	24	32	35	26
CXCR4	20	11	36	20	53
DEF	17	23	23	26	25
DHI1	8	10	11	12	22
DPP4	36	26	41	41	36
DRD3	14	8	13	10	17
DYR	33	31	43	43	39
EGFR	22	21	29	35	32
ESR1	14	8	18	17	21
ESR2	17	8	19	17	18
FA10	33	11	39	43	41
FA7	53	14	56	67	53
FABP4	44	33	46	46	56
FAK1	9	16	24	25	25
FGFR1	24	9	18	16	18
FKB1A	9	18	16	16	22
FNTA	11	14	16	25	15
FPPS	55	35	51	53	71
GCR	1	4	4	3	9
GLCM	16	32	30	31	33
GRIA2	15	10	23	23	14
GRIK1	36	6	35	35	18
HDAC2	20	16	24	25	19
HDAC8	25	20	29	29	18
HIVINT	13	1	8	4	7
HIVPR	4	5	7	8	7
HIVRT	9	9	11	11	19
HMDH	17	35	26	25	34
HS90A	11	12	15	15	18
HXK4	21	16	25	27	30
IGF1R	8	8	18	25	22
INHA	12	18	19	20	19
ITAL	6	3	15	19	27
JAK2	12	19	29	28	28
KIF11	25	17	34	40	34
KIT	7	7	12	11	17
KITH	12	10	15	20	20
KPCB	13	7	15	17	13
LCK	18	14	25	22	28
LKHA4	27	13	18	19	20
MAPK2	15	19	33	43	30
MCR	-5	-4	-4	-4	0
MET	30	8	24	24	27
MK01	21	20	19	24	32
MK10	14	21	24	27	25
MK14	10	12	17	17	20

Target	None	Full	SEV	Thin	Drug-like*
MMP13	22	2	12	12	21
MP2K1	12	15	16	16	21
NOS1	12	25	19	19	33
NRAM	46	10	44	45	41
PA2GA	16	21	26	28	37
PARP1	22	17	25	37	21
PDE5A	11	16	17	19	20
PGH1	7	2	3	3	11
PGH2	10	11	13	13	20
PLK1	18	21	28	30	33
PNPH	9	32	39	40	12
PPARA	16	7	19	20	13
PPARD	28	22	32	34	43
PPARG	18	11	21	21	28
PRGR	9	8	8	9	25
PTN1	30	25	36	35	44
PUR2	54	-13	51	53	48
PYGM	16	2	17	18	9
PYRD	29	14	30	30	33
RENI	24	11	21	21	17
ROCK1	11	10	17	11	19
RXRA	30	18	25	27	32
SAHH	24	39	39	38	58
SRC	17	11	18	20	21
TGFR1	30	27	36	37	33
THB	33	38	36	37	37
THRB	31	5	32	34	27
TRY1	45	18	57	57	42
TRYB1	43	5	37	37	29
TYSY	31	7	28	27	21
UROK	45	41	65	64	47
VGFR2	17	13	21	20	22
WEE1	39	31	52	64	55
XIAP	61	7	52	60	53

*Drug-like background made from all ChEMBL12 ligands with affinities below 10 μM
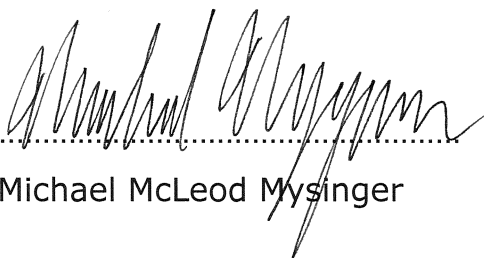
Table A.3.5: Comparison of DUD-E and DUD

Incremental Change		All Original	New Style Decoys	Switch Ligands	Switch Target Preparation
Decoys		DUD	DUD-E	DUD-E	DUD-E
Ligands		DUD	DUD	DUD-E	DUD-E
Receptor Preparation		DUD	DUD	DUD	DUD-E
LogAUC	*Average*	*14.8*	*19.7*	*16.4*	*22.8*
DUD-E Target	DUD target				
ACE	ace_auto	20	29	19	22
ACES	ache_auto	2	3	9	25
ADA	ada_semi	8	19	17	24
ALDR	alr2_auto	24	24	14	31
AMPC	ampc_auto	0	8	10	16
ANDR	ar_auto	19	3	3	5
CDK2	cdk2_semi	21	28	25	21
COMT	comt_auto	31	27	33	34
DYR	dhfr_semi	13	30	19	43
EGFR	egfr_semi	25	32	28	29
FA10	fxa_auto	13	28	25	39
FGFR1	fgfr1_auto	-10	-6	11	18
GCR	gr_auto	10	-3	-4	4
HIVPR	hivpr_semi	7	12	-2	7
HIVRT	hivrt_auto	6	8	8	11
HMDH	hmga_auto	20	14	16	26
HS90A	hsp90_semi	21	20	17	15
INHA	inha_auto	10	24	11	19
KITH	tk_auto	1	32	14	15
MCR	mr_auto	45	27	-4	-4
MK14	p38_semi	4	19	16	17
NRAM	neua_auto	14	14	32	44
PARP1	parp_semi	15	35	33	25
PDE5A	pde5_semi	18	18	17	17
PGH1	cox1_auto	8	15	-4	3
PGH2	cox2_auto	26	29	8	13
PNPH	pnp_auto	15	14	12	39
PPARG	ppar_auto	-4	2	15	21
PRGR	pr_auto	4	-5	4	8
PUR2	gart_semi	40	62	37	51
PYGM	gpb_semi	17	13	2	17
RXRA	rxr_auto	34	37	22	25
SAHH	sahh_auto	26	26	31	39
SRC	src_semi	1	10	19	18
THRB	thrombin_auto	27	39	28	32
TRY1	trypsin_auto	15	31	53	57
VGFR2	vegfr2_semi	3	11	15	21

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____   6/8/2012
Michael McLeod Mysinger   Date