Pocket optimization and its application to identify small-molecule inhibitors of protein-protein interactions

By

David Keith Johnson

Submitted to the graduate degree program in Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson Dr. John Karanicolas

_____

Dr. Eric Deeds

_____

Dr. Wonpil Im

_____

Dr. Ilya Vakser

_____

Dr. Mario Rivera

Date Defended: December 8, 2015

The Dissertation Committee for David Keith Johnson

certifies that this is the approved version of the following dissertation:

Pocket optimization and its application to identify small-molecule inhibitors of

protein-protein interactions

_____

Chairperson Dr. John Karanicolas

Date approved: December 8, 2015

ABSTRACT

Because of their ubiquitous nature in many cellular processes, modulating protein-protein interactions offers tremendous therapeutic potential. However, protein-protein interactions remain a difficult class of drug targets, as most protein interaction sites have not evolved to bind small molecules. Indeed, some protein interaction sites are thought to be simply not amenable to binding any small molecule at all. Other sites feature small molecule binding pockets that simply are not present in the unbound or protein-bound conformations, making structure-based drug discovery difficult. Sometimes, inhibitors bind to multiple family members with high affinity, causing toxicity. In this dissertation I seek to address many of these challenges, by developing methodologies to assess the druggability of a target, assess the selectivity of known inhibitors, identify conformations that are sampled uniquely by a single protein, and identify inhibitors of protein-protein interactions.

To assess druggability, I developed the "pocket optimization" protocol which uses a biasing potential to create an ensemble of conformations that contain pockets at a specified location on the protein surface. I showed that low-resolution, low energy inhibitor shapes are encoded at druggable sites and sampled through low-energy fluctuations, whereas they are not present at random sites on protein surfaces.

To assess selectivity and screen for inhibitors, I developed "exemplars", representations of a pocket based on the perfect "non-physical" complementary ligand, allowing the comparison of pocket shapes independent of protein sequence. I predicted the selectivity of an array of inhibitors to a related family of proteins by comparing the exemplars from the known small-molecule bound conformation to the ensemble of exemplars from a "pocket optimized" ensemble. I identified distinct conformations that could be targeted for identifying selective inhibitors *de novo* by comparing ensembles of exemplars from related family members to one another. Finally, I developed a screening protocol that uses the speed of exemplar versus small molecule comparisons to screen very large compound libraries against ensembles of distinct, "pocket optimized" pocket conformations.

iii

This dissertation is outlined as follows. In the first chapter, I will introduce the pocket optimization protocol and show that low-resolution pocket shapes are encoded into the protein surface at druggable interfaces. In the second chapter, I will introduce exemplars and apply them to predict selectivity. Finally, in the third chapter I will combine pocket optimization and exemplars to present and benchmark exemplar screening for identifying novel inhibitors of protein-protein interactions.

The text of Chapter 1 is a reprint of the material from:

Johnson DK, Karanicolas J (2013) Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. PLoS Comput Biol 9(3): e1002951. doi:10.1371/journal.pcbi.1002951

Copyright © 2013 Johnson, Karanicolas.

The supporting information for this chapter is included as Appendix A.1.

The text of Chapter 2 is a reprint of the material from:

Johnson DK, Karanicolas J (2015) Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. PLoS Comput Biol 11(2): e1004081. doi:10.1371/journal.pcbi.1004081

Copyright © 2015 Johnson, Karanicolas.

The supporting information for this chapter is included as Appendix A.2.

The text of Chapter 3 is a reprint with permission from:

Johnson DK, Karanicolas J. (2016). Ultra-high-throughput structure-based virtual screening for small-molecule inhibitors of protein-protein interactions. J Chem Inf Model Just Accepted Manuscript. doi: 10.1021/acs.jcim.5b00572

The supporting information for this chapter is included as Appendix A.3.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. John Karanicolas for his enthusiasm, guidance, patience, and support throughout the course of my research.

I would like to thank the members of my dissertation committee: Dr. Eric Deeds, Dr. Wonpil Im, Dr. Ilya Vakser, and Dr. Mario Rivera for their guidance.

I would like to thank the faculty of the Center for Computational Biology: Dr. Ilya Vakser, Dr. Wonpil Im, Dr. John Karanicolas, Dr. Eric Deeds, Dr. Christian Ray, and Dr. Joanna Slusky for their support of my professional and academic career.

I am thankful to my colleagues Dr. Andrea Bazzoli, Dr. Ragul Gowthaman, Jimmy Budiardjo, Shipra Malhotra, Jittasak Khowsathit, Yusuf Adeshina, Nan Bai, and Jin Liu for their collaborations and discussions.

Finally, I would like to thank my wife Jennifer, my son Owen, my mother Johnnie, my father Keith, and the rest of my family for their love and support.

TABLE OF CONTENTS

INTRODUCTION

Druggability of protein-protein interaction sites

Modulating protein-protein interactions through small-molecule agents has exciting therapeutic potential. Because of the ubiquitous nature of protein-protein interactions, manipulating these interactions provides the opportunity to treat cancer, bacterial and viral infections, and other autoimmune disorders (1-7). However, they are considered a challenging class of drug targets (8-12). Indeed high-throughput screening can have very low hit rates when applied to this target class, with rates as low as 0.01% having been reported against a large pharmaceutical library (13).

In the cases where a protein structure has been solved for the unbound protein, as well as in complex a small-molecule inhibitor, and with its biological (protein) partner, it was observed that binding the small molecule was not associated with large conformational changes (8). However, the concave pocket at the small molecule binding surface of protein was typically smaller or absent from the unbound protein structure. Thus, the unbound structure had to undergo a small conformational change in order to expose a binding site that was previously absently in the unbound structure (8).

Inspired by methods such "SAR by NMR" (14) and "multiple solvent crystal structures" (15) that assess druggability by the ability to interact with multiple probe molecules at the interaction site, I proposed two hypotheses. First, the ability to form a small molecule binding pocket at the protein-protein interaction site may be the limiting factor for druggability. Second, an inhibitor acting at a protein interaction site should have shape complementarity to specific surface pockets found in the low-energy ensemble. These two hypotheses imply that druggable sites differ from the rest of the protein surface in that, unlike random sites on the protein surface, druggable sites include a special subset of "pocket-containing" conformations observed in the surface fluctuations under physiological conditions.

To test these hypotheses, I have developed a protocol to sample "pocket-opened" conformations. This method probes the pockets at a specified location and preferentially samples conformations that increase the pocket volume at this site.

Selectivity of protein-protein inhibitors

Selectivity of a compound for its desired protein target(s) is an important property that is typically optimized during the small-molecule drug discovery process (12). While some diseases, such as chronic myeloid leukemia, can be treated by targeting a single, dysfunctional protein, others, including various cancers, are more complex and involve multiple targets. The ability for a single drug to hit multiple targets – "polypharmacology" – has garnered interest in the treatment of cancer and psychiatric conditions (16). Compound promiscuity, however, has a clear downside due to the potential for adverse effects (toxicity) from unintended interactions with unrelated proteins, or from other family members of the target protein (17). One such example is ABT-263 (navitoclax), a Bcl-2 inhibitor that exhibited toxicity (thrombocytopenia) from its interaction with Bcl-xL (18, 19).

A lead compound can be tuned for selectivity by exploiting differences in shape and electrostatics between target and off-target proteins. Structure-activity relationships (SAR) (20) or structural biology (21-23) can be used to identify features that would prevent an undesirable interactions. Unlike a number of "traditional" drug targets, such as kinases, that have had determinants of selectivity carefully mapped, for "non-traditional" drug targets, most notably small-molecule inhibitors of protein-protein interactions, rational design of selectivity is not trivial.

We sought to predict selectivity of an inhibitor across members of a protein family. We hypothesized that proteins that interact with a given small molecule would have, in their low-energy ensemble, a conformation that presented an interface with both shape and chemical complementarity, whereas proteins that would not interact would not present a complementary binding site. To compare

2

surface pockets, I introduce the concept of an "exemplar", a non-physical molecule that maps out the shape and chemical features of a pocket.

Discovery of novel small-molecule inhibitors

Pharmacophores, defined as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" (24) have been used in drug discovery efforts for decades to design analogs (25-27). The first pharmacophore-building algorithms drew information from the ligand alone by finding shared functional groups in a consensus structural alignment of multiple active compounds (28). This approach has evolved to define key interactions from one or more crystal structures of a receptor with assorted ligands bound, as seen with tools such as LigandScout (29). With many traditional drug targets, even in the absence of known active compounds, one or more of the natural ligand(s) are known, so a starting approach could be to recapitulate the key interactions of these natural ligand(s).

However some biological targets, such as protein-protein or protein-RNA interactions, have not evolved to bind *any* natural small-molecule ligand. Without a natural small-molecule binding partner to use as a template for building a pharmacophore model, it is not feasible to build a ligand-based pharmacophore until after an active compound has been discovered. Furthermore, it has been observed that some of the binding pockets on the protein surface are transient, seen neither in the unbound structure nor in that of the protein-bound complex (8).

Because of these challenges, the focus of several groups has recently been on developing computational approaches that select some of the side chains from the natural protein partner (the "hotspot" or "anchor" residues) as the template to build a pharmacophore (30-37). However, this approach is limited by the chemical space from which to sample, limited by potentially low binding affinity by only recapitulating a subset of the natural binding partner's interactions, and limited by not

3

being able to identify compounds that require conformational changes at the protein surface (because they do not resemble the natural binding partner).

By virtue of being the mapping of the shape and chemical complementarity of a binding pocket, exemplars are, in essence, pharmacophores. Using the exemplar as a pharmacophore, combined with "pocket optimization" to create ensembles of pocket-opened conformations, we introduce a virtual screening protocol that uses exemplars to screen ensembles of conformations against very large datasets.

In the chapters that follow I will describe the development of these tools, then conclude by assembling them into a unified pipeline for virtual screening.

CHAPTER 1: Druggable Protein Interaction Sites are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface

David K. Johnson[1] and John Karanicolas[1,2*]

[1] Center for Bioinformatics and [2]Department of Molecular Biosciences,

University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534

*To whom correspondence should be addressed. E-mail: **johnk@ku.edu**, 785-864-8298

1.1 Abstract

Despite intense interest and considerable effort via high-throughput screening, there are few examples of small molecules that directly inhibit protein-protein interactions. This suggests that many protein interaction surfaces may not be intrinsically "druggable" by small molecules, and elevates in importance the few successful examples as model systems for improving our fundamental understanding of druggability. Here we describe an approach for exploring protein fluctuations enriched in conformations containing surface pockets suitable for small molecule binding. Starting from a set of seven unbound protein structures, we find that the presence of low-energy pocket-containing conformations is indeed a signature of druggable protein interaction sites and that analogous surface pockets are not formed elsewhere on the protein. We further find that ensembles of conformations generated with this biased approach structurally resemble known inhibitor-bound structures more closely than equivalent ensembles of unbiased conformations. Collectively these results suggest that "druggability" is a property encoded on a protein surface through its propensity to form pockets, and inspire a model in which the crude features of the predisposed pocket(s) restrict the range of complementary ligands; additional smaller conformational changes then respond to details of a particular ligand. We anticipate that the insights described here will prove useful in selecting protein targets for therapeutic intervention.

1.2 Author Summary

Identifying small-molecule inhibitors of protein interactions has traditionally presented a challenge for modern screening methods, despite interest stemming from the fact that such interactions comprise the underlying mechanisms for cell proliferation, differentiation, and survival. This suggests that many protein interaction surfaces may not be intrinsically "druggable" by small molecules, and elevates in importance the few successful examples as model systems for improving our understanding of factors contributing to druggability. Here we describe a new approach for exploring protein fluctuations leading to surface pockets suitable for small molecule binding. We find that the presence of such pockets is indeed a signature of druggable protein interaction sites, suggesting that "druggability" is a property encoded on a protein surface through its propensity to form pockets. We anticipate that the insights described here will prove useful in selecting protein targets for therapeutic intervention.

1.3 Introduction

Manipulating the interactions between proteins represents a promising avenue for therapeutic intervention in a variety of settings. Given the ubiquitous nature of protein interactions, selectively manipulating such interactions could serve as a means to treat conditions including viral and bacterial infections, cancer, and autoimmune disorders (1-7). In spite of recent ongoing efforts that have provided cause for optimism, protein interactions continue to be viewed as a challenging class of therapeutic target (8-12). While high-throughput screening efforts that fail to yield extensive hits are typically not reported in the literature, hit rates as low as 0.01% in a large pharmaceutical library have been described (13).

This dearth of successful representatives to study has given increased importance to the several cases in which a protein structure has been solved in complex with a biological protein partner and also in complex with a small molecule inhibitor. Wells and McClendon (8) compared six such cases and observed that binding was not associated with a large conformational change in any of these examples; and yet, the concave pocket on the protein surface at which the small molecule binds was typically smaller or not present in the unbound protein structure. In order for inhibitor binding to occur, the surface of the unbound structure therefore had to undergo local rearrangement to reveal a small molecule binding site that would not necessarily be evident from the unbound structure (8).

Given the limited success in identifying modulators of protein-protein interactions, it has proven helpful at an early stage to validate a protein surface site by evaluating its "druggability". As such, fragment-based methods have been developed to experimentally assess the druggability of a protein interaction site by determining which members of a small molecule probe set bind to a target protein, and where on the protein surface these bind. This experiment can be conducted using "SAR by NMR" (14), which tracks chemical shift differences to identify binding sites on the protein surface, or by the "multiple solvent crystal structures" method (15), in which independent structures of the target protein are solved after soaking with a collection of organic solvents. Both of these approaches aim to probe the regions of a

protein surface that can accommodate small-molecule binding, with a preference for sites that are not uniquely disposed to bind a particular pre-selected ligand. In either technique, specific regions of the protein surface that interact with a variety of probe molecules – albeit weakly – are inferred to be a putative site for more potent binding by some yet unidentified compound.

Inspired by these methods, we hypothesize that the ability to form a binding pocket may be the limiting factor for druggability of a protein surface site. We further propose that compounds identified in biochemical screens as inhibitors of protein interactions result from natural shape complementarity to specific surface pockets that form with little energetic cost to the protein. Together these hypotheses imply that druggable sites differ from the rest of the protein surface, in that fluctuations under physiological conditions at druggable sites include a special subset of "pocket-containing" conformations.

To test these hypotheses, we have developed computational methodology to explore protein fluctuations in a biased way, by providing a driving force towards conformations in which a surface pocket is present. Several other studies have generated ensembles of protein conformations reflecting fluctuations around the native state and used these either to assess druggability (38) or as a starting point for docking studies (39-46). Because these ensembles are generated in an unbiased manner, however, a large fraction of the resulting ensembles correspond to protein conformations that are dissimilar to the corresponding bound structure in both pocket size and hydrophobicity (40). One of these studies found that carrying out simulations in methanol led to formation of surface pockets which could accommodate small molecules (47), but the use of this non-biological solvent may lead to unphysical artifacts in the resulting models. Molecular dynamics has also been used in a computational analog of "SAR by NMR" in which simulations are carried out using an explicit mixed solvent, allowing druggable sites to be identified by locating accumulation of probe molecules (39, 42, 48). Though these methods proved effective for binding sites, the unbiased nature of the underlying simulations make these approaches very computationally intensive. Recently Kozakov et al. have developed a computational analog of the multiple solvent crystal structures method, by using docking to identify "consensus" sites at which several
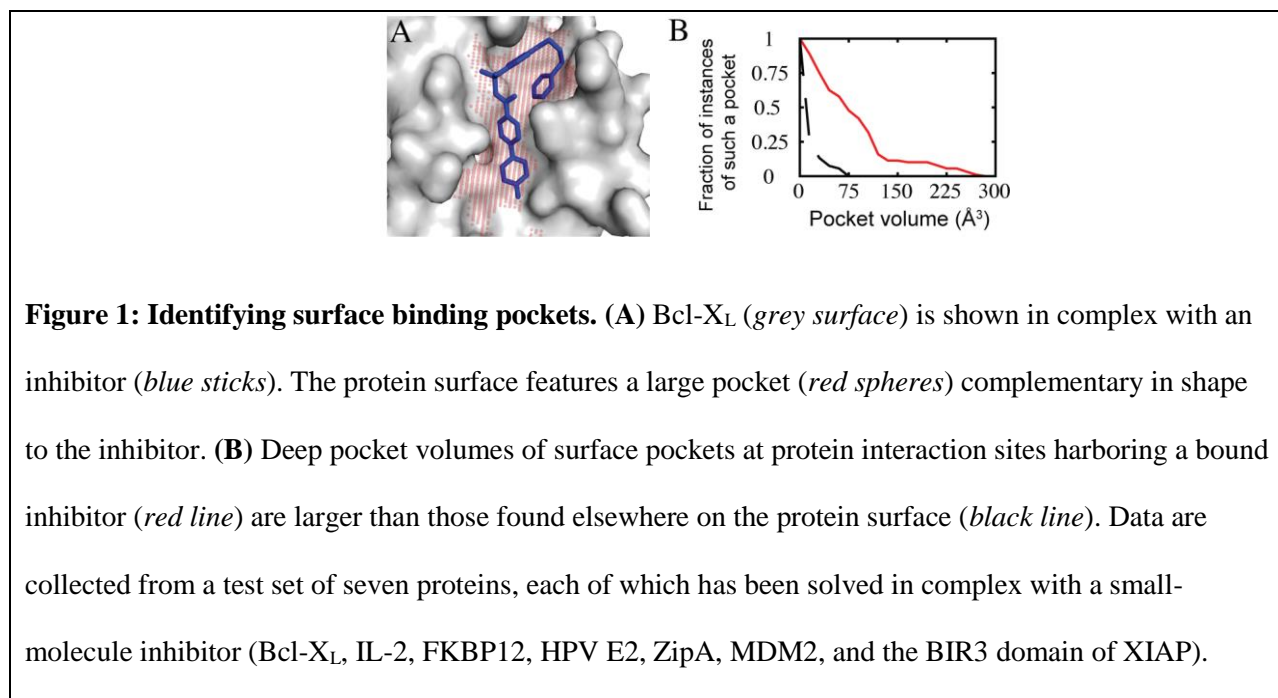
probe molecules cluster (49). They confirmed that these molecular probes indeed cluster at established

druggable sites and that known inhibitors often occupy these consensus sites. Such an approach, however,

cannot efficiently explore surface pockets that form via concerted motions involving the protein backbone

due to the computational expense associated with repeatedly docking multiple small molecule probes. The

biasing potential we describe here avoids this limitation by not needing to dock probe molecules, and

therefore can be used in the course of a simulation that samples a broader range of conformational

fluctuations.

1.4 Results

Quantitative analysis of surface pockets

　　Because a wide, shallow pocket that is unsuitable for small molecule binding can have the same

volume as a deep pocket that is more suitable for small molecule binding, we introduce the concept of

"deep" volume of a pocket: the volume of the pocket that is well-sequestered away from solvent. To

quantitatively identify small molecule binding pockets and measure their "deep" volume, we

implemented a modified version of the LIGSITE^{cs} algorithm (50). This approach starts by creating a grid

around a protein and marking each grid point as occupied by protein, surface, or solvent. Next, the

algorithm performs linear searches on the grid to find "Surface-Solvent-Surface" events: lines drawn

between two surface points that pass through only solvent (**Figure S1**), indicating a concave region on the

protein surface. To distinguish between total pocket volume and deep volume, pocket points that fall

within 2.5 Å of solvent are marked as "surface pocket" points and are excluded from the "deep volume"

calculation. Finally, the remaining contiguous points involved in these events are clustered into "deep

pockets". As expected, the deep pocket volumes we use here are correlated to, but smaller than, pocket

volumes found by other pocket detection methods, such as Q-SiteFinder (41) (**Figure S2**). Our

implementation differs from the original LIGSITE^{cs} algorithm (50) in that our search is restricted to the

region around a specific "target" residue on the protein surface, allowing us to rapidly test for pockets at a specific surface site. Additional minor differences are described in **Text S1**.

Figure 1: Identifying surface binding pockets. (A) Bcl-X_L (*grey surface*) is shown in complex with an inhibitor (*blue sticks*). The protein surface features a large pocket (*red spheres*) complementary in shape to the inhibitor. **(B)** Deep pocket volumes of surface pockets at protein interaction sites harboring a bound inhibitor (*red line*) are larger than those found elsewhere on the protein surface (*black line*). Data are collected from a test set of seven proteins, each of which has been solved in complex with a small-molecule inhibitor (Bcl-X_L, IL-2, FKBP12, HPV E2, ZipA, MDM2, and the BIR3 domain of XIAP).

A demonstration of this method is shown in **Figure 1A**. Bcl-X_L is an anti-apoptotic protein in the Bcl-2 family whose over-expression has been implicated in the survival of cancer cells. A series of acyl-sulfonamide-based ligands have been shown to inhibit Bcl-X_L activity by competing for its peptide-binding groove. Here, we have removed one such inhibitor from a co-crystal structure and applied our modified implementation of the LIGSITE^{cs} algorithm at this surface site. The resulting pocket has intuitive shape complementarity to the ligand even though it was generated from the protein structure without the ligand present. This is unsurprising, given that the ligand occupying this pocket is complementary in shape to the protein surface.

We compiled a test set of all seven proteins for which structures have been solved both alone and in complex with a small-molecule inhibitor bound to a protein interaction site (Bcl-X_L, IL-2, FKBP12, HPV E2, ZipA, MDM2, and the BIR3 domain of XIAP) (**Table 1**). We compared the deep pocket volume at randomly selected regions of the protein surface (see **Text S1**) to the deep volume of the inhibitor-

bound pocket. Since the pocket definition can depend somewhat on the particular "target" residue used, the deep volume of the inhibitor-bound pocket was measured several times using different target residues. As shown in **Figure 1B**, across all seven proteins, most pockets identified by this algorithm at randomly selected sites (*black line*) on the protein surface have a deep volume smaller than 25 Å^3, and all are smaller than 75 Å^3. In contrast, the distribution of inhibitor-bound pocket volumes (*red line*) is significantly shifted, with about half of the inhibitor-bound deep pocket volumes larger than 75 Å^3. This observation is consistent with results generated using other pocket detection methods (51-55), although those other studies were not focused on inhibitors of protein-protein interactions.

Protein	Partner	Number of residues	Number of residues used in simulations	Deep pocket volume from unbound protein structure (Å^3)	Deep pocket volume from protein-protein complex (Å^3)	Deep pocket volume from inhibitor-bound complex (Å^3)	Deep pocket volume from conformations generated via biased simulations (the 5% with lowest energy) (Å^3)	Affinity of most potent known inhibitor (nM)
MDM2	p53	492	119	129	150	201	194	80
ZipA	FtsZ	328	139	91	166	61	92	12,000
Bcl-xL	Bak	233	141	200	377	342	210	0.5
XIAP	SMAC	497	117	158	27	40	323	67
IL-2	IL-2R	153	128	38	38	38	177	60
HPV E2	HPV E1	367	193	93	92	132	129	40
FKBP12	Tbeta1-R	108	107	90	124	145	116	7

Table 1: Our test set is comprised of proteins for which structures have been solved both alone and in complex with a small-molecule **inhibitor** bound to the protein interaction site.

Druggability of protein surface sites

The results presented above demonstrate that inhibitor binding occurs at surface sites containing a pocket, and that these sites are distinct from the remainder of the protein surface. We therefore formulated

13

the hypothesis that the ability of the protein surface to form such pockets may be the limiting determinant of the inherent druggability at this site.

Unlike previous standalone methods for pocket detection, we instead implemented our algorithm as a term in the Rosetta (56) energy function alongside the canonical energetic determinants of protein structure such as packing, hydrogen bonding, and solvation (see Methods section). By including this biasing term in the energy function, we may use any of the standard functionalities provided in Rosetta; inclusion of this term, meanwhile, will lead to simultaneous optimization of both "pocket score" *and* the traditional energy terms. In essence, the contribution from the "pocket" term serves as a proxy for the energy associated with binding of some (unspecified) small molecule partner.

To test the hypothesis that pocket formation may be the limiting determinant of druggability, we performed biased and unbiased simulations on the unbound conformations of Bcl-X_L, targeting residues at the protein interaction site as well as at equivalent randomly selected residues elsewhere on the protein surface. Surface sites included in the random set were matched to those at the protein interaction site on the basis of their secondary structure, and further that a contacting pocket of equivalent size to that of the protein interaction site (evaluated by Q-SiteFinder (41)) was present in the unbound conformation (see **Text S1** and **Figure S3**); further, the random sites were each at least 12 Å from one another. Both backbone and sidechain degrees of freedom were allowed to move during simulations (see Methods section). The deep pocket volumes from each of 1,000 conformations generated via each method are shown as cumulative histograms in **Figure 2A**. Pockets at the protein interaction site (*solid red lines*) form more often and are significantly larger than those formed elsewhere on the protein surface (*dashed black lines*). The largest pockets in the biased simulations are sampled with much higher frequency than in the corresponding unbiased simulations (**Figure 2A**), demonstrating that the biasing potential drives sampling towards these conformations. These observations further hold for each of the other six additional proteins comprising our test set (**Figures 2B-G**), and also after inclusion of additional random sites (**Figure S4**) or starting from the protein-bound conformation (**Figure S5**).

To examine the physiological relevance of the conformations generated in biased simulations, we compared their energies to those obtained in equivalent unbiased simulations. For Bcl-X_L, we used Rosetta to evaluate the (unbiased) energy for each of 1000 conformations generated from an unbiased simulation, a simulation with the biasing term applied to either the protein interaction site or a random surface residue, and equivalent simulations in which the weight of the biasing term was increased tenfold. In all cases we evaluated energies *without* contribution from the biasing term; a histogram of these energies is shown in **Figure 3A**. Conformations from the unbiased simulation (*green solid line*) have a very similar distribution of energies as conformations from a simulation in which the biasing potential was applied to a randomly selected target residue (*black dashed line*); this is unsurprising given that few of these conformations contain pockets, implying that a similar ensemble of conformations are sampled. Applying a stronger weight to the biasing potential using the same randomly selected target residue (*solid black line*) leads to conformations that are far less energetically favorable, indicating that pocket opening to satisfy the biasing potential could not be achieved without extensive energetic cost to the protein. In contrast, applying the biasing potential to a residue at the protein interaction site led to conformations that were only slightly higher in energy (and with overlapping distributions) than those conformations sampled in the unbiased simulation, for either weight of the biasing potential (*red lines*). A scatterplot showing the deep pocket volume for each of these conformations highlights the fact that conformations containing large pockets are sampled with the moderate biasing potential *only* if it is applied at the protein interaction site (**Figure 3B**, *cyan vs. red points*). Application of the stronger biasing potential to random (carefully matched) surface sites leads to generations of low-energy conformations without large pockets, and also pocket-containing conformations with much higher energy (**Figure 3B**, *blue points*).
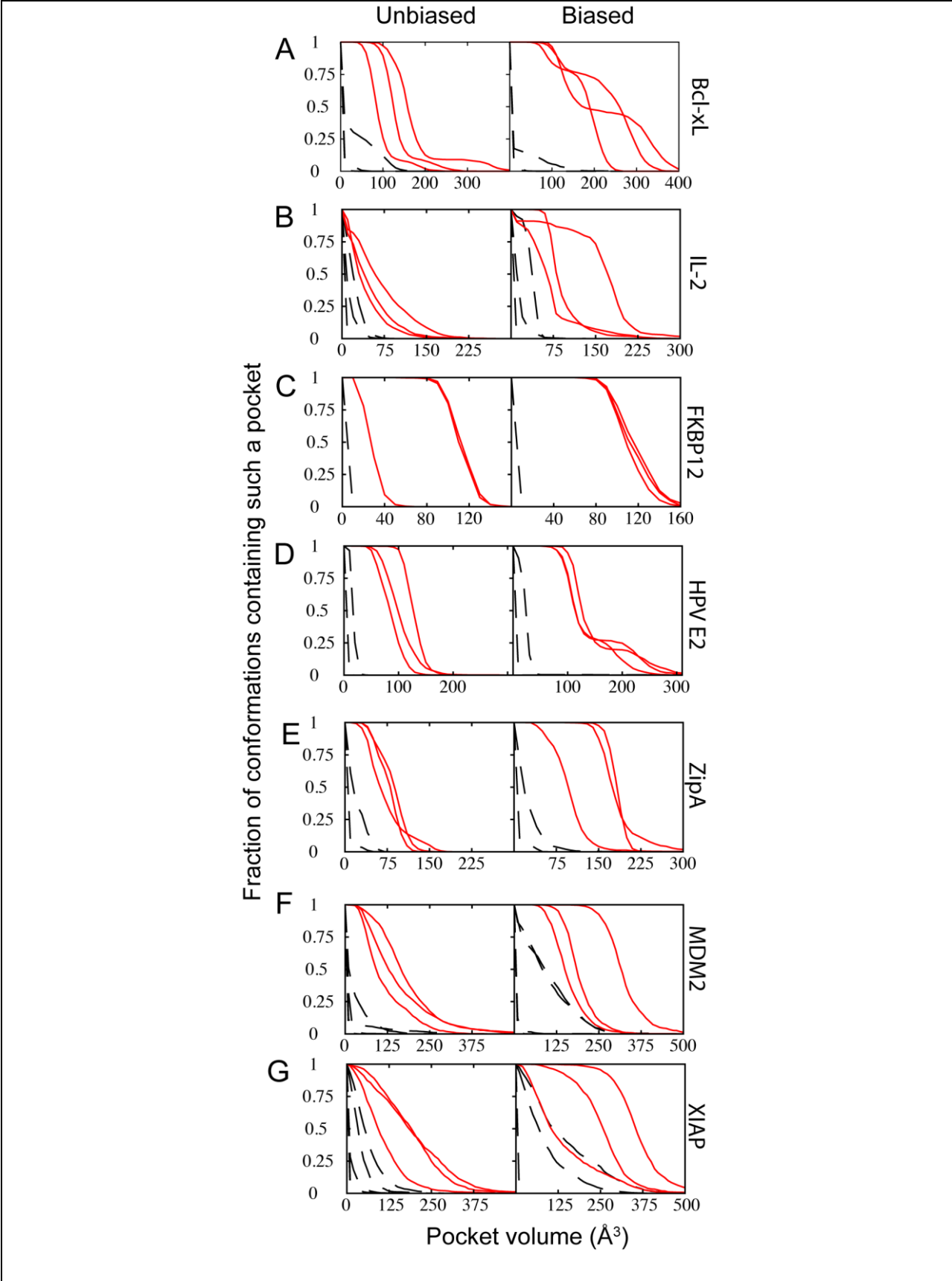
Unbiased   Biased

A — Bcl-xL

B — IL-2

C — FKBP12

D — HPV E2

E — ZipA

F — MDM2

G — XIAP

Fraction of conformations containing such a pocket

Pocket volume (Å³)

Figure 2: Surface pockets emerge only at druggable sites. Volumes of surface pockets are shown from conformations generated with no biasing potential (*left*) and upon inclusion of a "pocket opening" biasing potential (*right*) for each of the seven proteins that comprise our test set. Surface pockets occur at druggable protein interaction sites (*solid red lines*) more frequently than elsewhere on the protein surface (*dashed black lines*). **(A)** Bcl-X$_L$. **(B)** IL-2. **(C)** FKBP12. **(D)** HPV E2. **(E)** ZipA. **(F)** MDM2. **(G)** BIR3 domain of XIAP.

The same observations also hold for each of the other six proteins comprising our test set (**Figure S6**, **Table 1**). Collectively, these results demonstrate that pocket opening at the druggable site can occur with little energetic cost to the protein, while pocket opening elsewhere on the protein surface requires that the protein adopt a highly unfavorable conformation.

It is notable that in each of these seven examples surface pockets were identified at the protein interaction site. It is equally notable, however, that similar surface pockets were *not* observed elsewhere on the protein surface (**Figure 2**). This comparison highlights the qualitative difference between the protein interaction site – already demonstrated to be druggable in a practical sense for each of these examples – and the remainder of the protein surface, at which high-affinity interactions with small molecules have not been observed.
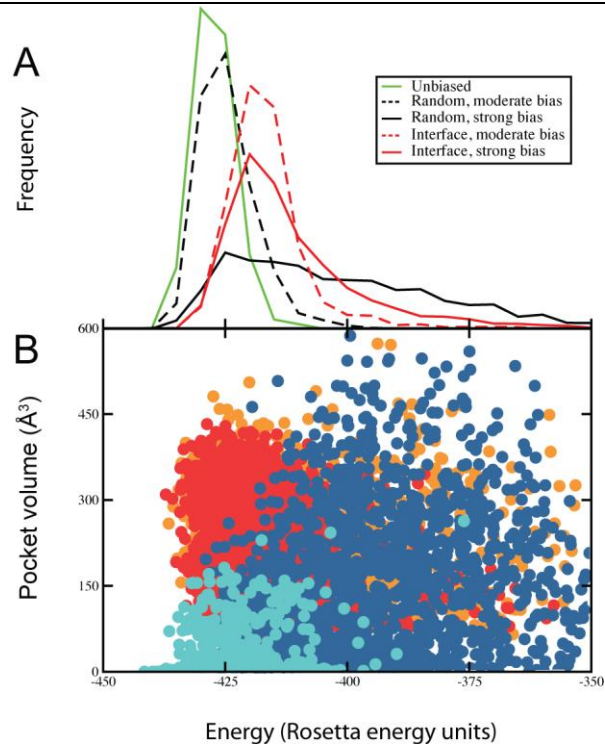
Figure 3: Energetic analysis of Bcl-X$_L$ **pocket opening. (A)** Conformations generated without the use of a biasing potential (*solid green line*) show a similar distribution of energies to those generated with the biasing potential at a randomly selected target residue (*dashed black line*); increasing the strength of the biasing potential here leads to conformations with higher energies (*solid black line*). In contrast, application of the biasing potential at the protein interaction site (*red lines*) leads to conformations with a distribution of energies that strongly overlaps with those energies of conformations sampled in the unbiased simulations, suggesting that these conformations represent low-energy states accessible to the unbound protein. **(B)** A scatterplot showing the deep pocket volume for conformations generated with the biasing potential applied to one of the random sites (*moderate bias in cyan, strong bias in blue*) or to the protein interaction site (*moderate bias in red, strong bias in orange*). Low-energy conformations containing large pockets are sampled only if the biasing potential it is applied at the protein interaction site; while large pockets are sampled using the strong bias at random sites, these conformations have considerably higher energy. All energies shown here were evaluated in the absence of the biasing potential, for fair comparison.

Pocket shapes are encoded on the protein surface

The results presented above demonstrate that there is a natural predisposition towards pocket formation in certain druggable regions on the protein surface. We next asked whether these preferred pocket-containing conformations dictate the range of potential small-molecule inhibitors suitable for binding at this site. Should this be the case, protein conformations generated using a biasing potential to induce pocket formation should resemble inhibitor-bound conformations more than conformations generated using an equivalent unbiased protocol.

Turning first to Bcl-X$_L$, we aligned the lowest-energy conformation produced from the biased simulations described earlier (which were started from the unbound protein structure) to both unbound and inhibitor-bound crystal structures (**Figure 4A**). Direct superposition of the inhibitor from the bound structure onto the unbound crystal structure reveals extensive steric clashes (**Figure 4B**), highlighting the local protein conformational changes that must take place in order for inhibitor binding to occur. Remarkably, examination of the conformation from the biased simulation shows that part of the protein surface has adopted a shape highly complementary to the inhibitor (**Figure 4C, right side**) – even though no information about the identity of the inhibitor was included in the simulation that produced this conformation. Further, this conformational change occurred without direct involvement of the "target" residue at which the biasing term was applied.

On the other hand, the conformation from the biased simulation did not recapitulate the bound conformation exactly: the latter contains an extension to the surface pocket (left side in the orientation shown) to accommodate the "tail" of the inhibitor (**Figure 4D**). Comparison of the protein backbone in all three structures (**Figure 4A**) reveals the protein structural reorganization required to accommodate this "tail": the helix below the inhibitor (**Figure 4A, foreground**) unwinds on the left side, and tightens on the right side (in the orientation shown). Because the deep pocket volume is larger in the inhibitor-bound crystal structure than in the conformation from the biased simulation, we anticipate that sampling in these simulations was insufficient to fully recapitulate this highly concerted conformational change.

Figure 4: Representative conformations of Bcl-X$_L$**.** An unbound crystal structure (*pink*), an inhibitor-bound crystal structure (*green, with inhibitor shown in sticks*), and a low-energy conformation generated from the unbound crystal structure using the biasing potential (*cyan, with target residue in red*) are shown. (A) The overall protein architecture is preserved amongst all three; movement of the helix in the foreground upon binding is not recapitulated in the pocket-opened conformation. (B-D) The pocket revealed in this low-energy conformation nonetheless strongly resembles the surface pocket in the bound crystal structure, and even bears shape-complementarity to the inhibitor. The identity of the inhibitor was not used in generating this conformation, but was added retrospectively for visual comparison.

The diversity of conformational changes associated with inhibitor binding in different proteins is highlighted in **Figure S7**. In IL-2, slight rearrangement of a helix (at top of the orientation shown) allows a sidechain rotamer change that both reveals the binding pocket and creates new interactions with inhibitor shown (**Figure S7A**). In contrast FKBP12, HPV E2 and ZipA each contain a pre-formed pocket on the unbound protein surface that strongly complements the inhibitor, indicating that inhibitor binding occurs primarily via a "lock-and-key" mechanism in these cases (**Figure S7B-D**). Expansion of the ZipA pocket observed in the biased simulations (**Figure 2E**) resulted from sidechain rearrangement on the surface to the left side of the binding site (in the orientation shown), without disrupting the pre-ordered
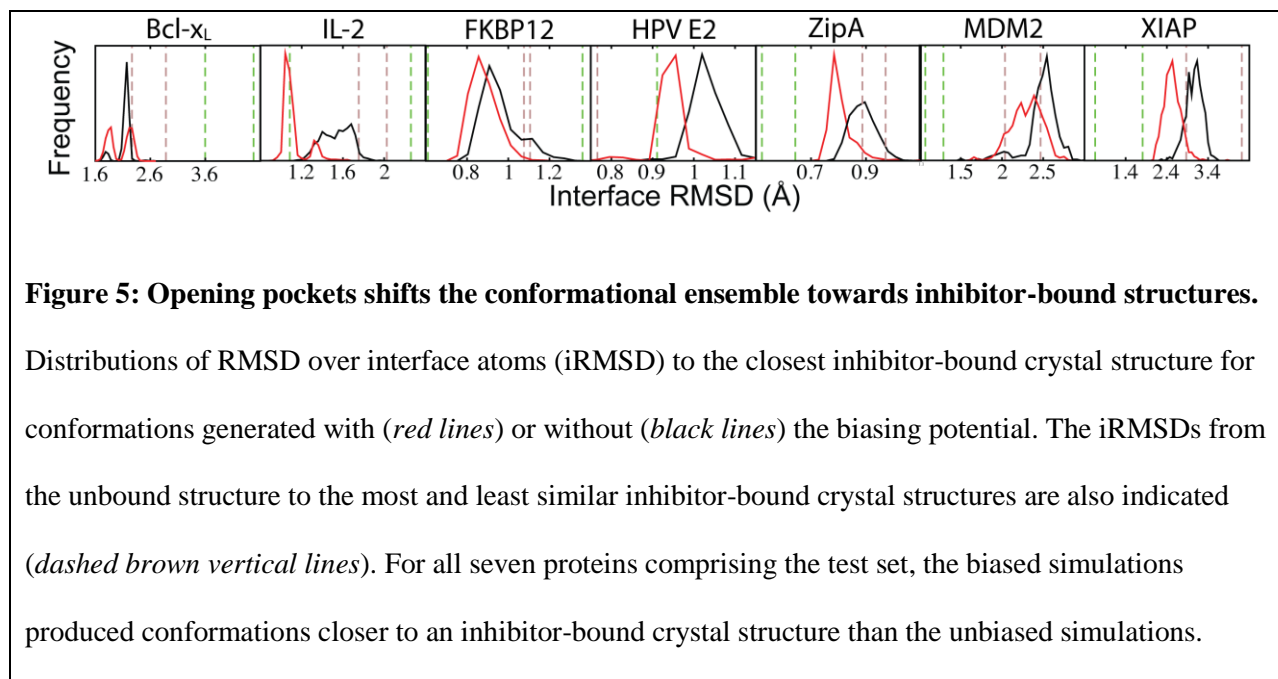
portions of the binding site. Nutlin binding to MDM2, on the other hand, requires splaying apart of two surface helices to create the binding pocket: this conformational change is dramatically recapitulated in the lowest-energy individual conformation from the biased simulations (**Figure S7E**). Finally, unlike these previous examples showing modest conformational changes, binding of a Smac-mimetic to the BIR3 domain of XIAP is associated with extensive rearrangement of long surface loops (**Figure S7F**).

To assess whether the protein conformations generated using the biasing potential resemble inhibitor-bound conformations more closely than conformations generated using an equivalent unbiased protocol, we quantitatively compared the ensembles of conformations produced by each method. We note that neither method uses any knowledge of any particular inhibitor when generating an ensemble of conformations. Therefore, rather than report the protein RMSD in reference to one pre-selected bound conformation we instead individually computed the interface RMSD (iRMSD) relative to every available inhibitor-bound crystal structure (see **Text S1**), and took the lowest of this set of iRMSD values to reflect the suitability of this conformation for binding some (known) unspecified inhibitor.

For each of the seven proteins in our test set described earlier, we show the iRMSD for each member of the biased and unbiased ensembles to its closest inhibitor-bound crystal structure (**Figure 5**). We note that the iRMSD values of conformations sampled vary across these seven different proteins; this stems from the fact discussed earlier that in some cases the unbound starting structure strongly resembles the inhibitor-bound structure (FKBP12, HPV E2, ZipA), while in other cases a more extensive conformational change accompanies binding (MDM2, XIAP) (**Figure S7**). For reference, we also indicate the iRMSD of the unbound starting structure relative to the closest and most distant inhibitor-bound structures (**Figure 5**, *dashed brown lines*) and the protein-bound structure relative to the closest and most distal inhibitor-bound structures (**Figure 5**, *dashed green lines*).

Though there is overlap between the resulting distributions, for every one of the seven proteins we examined the conformations sampled in the biased simulations are closer to an inhibitor-bound

conformation than the conformations sampled in the corresponding unbiased simulations. This observation applies not just to the iRMSD of atoms directly involved in binding (**Figure 5**), but to the backbone atoms of the corresponding residues as well (**Figure S8**). Further, in five of the seven cases (Bcl-X$_L$, IL-2, FKBP12, ZipA, and XIAP) the conformations generated from the biased simulations are closer than the starting unbound conformation to an inhibitor-bound structure (**Figure 5**). Conformations sampled in simulations of HPV E2 and MDM2 that moved further than the unbound conformation from the inhibitor-bound structures may be due to slight inaccuracies in the Rosetta energy function, or alternatively may be sampling novel truly "druggable" conformations for which complementary inhibitors have not yet been identified.

Figure 5: Opening pockets shifts the conformational ensemble towards inhibitor-bound structures. Distributions of RMSD over interface atoms (iRMSD) to the closest inhibitor-bound crystal structure for conformations generated with (*red lines*) or without (*black lines*) the biasing potential. The iRMSDs from the unbound structure to the most and least similar inhibitor-bound crystal structures are also indicated (*dashed brown vertical lines*). For all seven proteins comprising the test set, the biased simulations produced conformations closer to an inhibitor-bound crystal structure than the unbiased simulations.

To further characterize the pockets generated using the biasing potential, we evaluated the percentage of solvent accessible surface area that is hydrophobic (hSASA) for each pocket (**Figure S9**). While the hydrophobicity of the inhibitor-bound pockets vary amongst the different protein test cases, in each case the pockets generated using the biasing potential exhibit similar hydrophobicity to the corresponding inhibitor-bound conformation. Like the shape of these pockets, then, their hydrophobicity appears to be an intrinsic property resulting from details of the surface geometry and composition.

Collectively these results demonstrate that biasing simulations towards conformations in which a surface pocket is present drives the resulting ensemble towards the conformations observed in inhibitor-bound crystal structures. Because no information about the identity of any particular inhibitory compound was included in the biasing potential, this suggests that the general shape (and hydrophobicity) of surface pockets available to a potential inhibitor is an inherent property of the druggable interface itself.
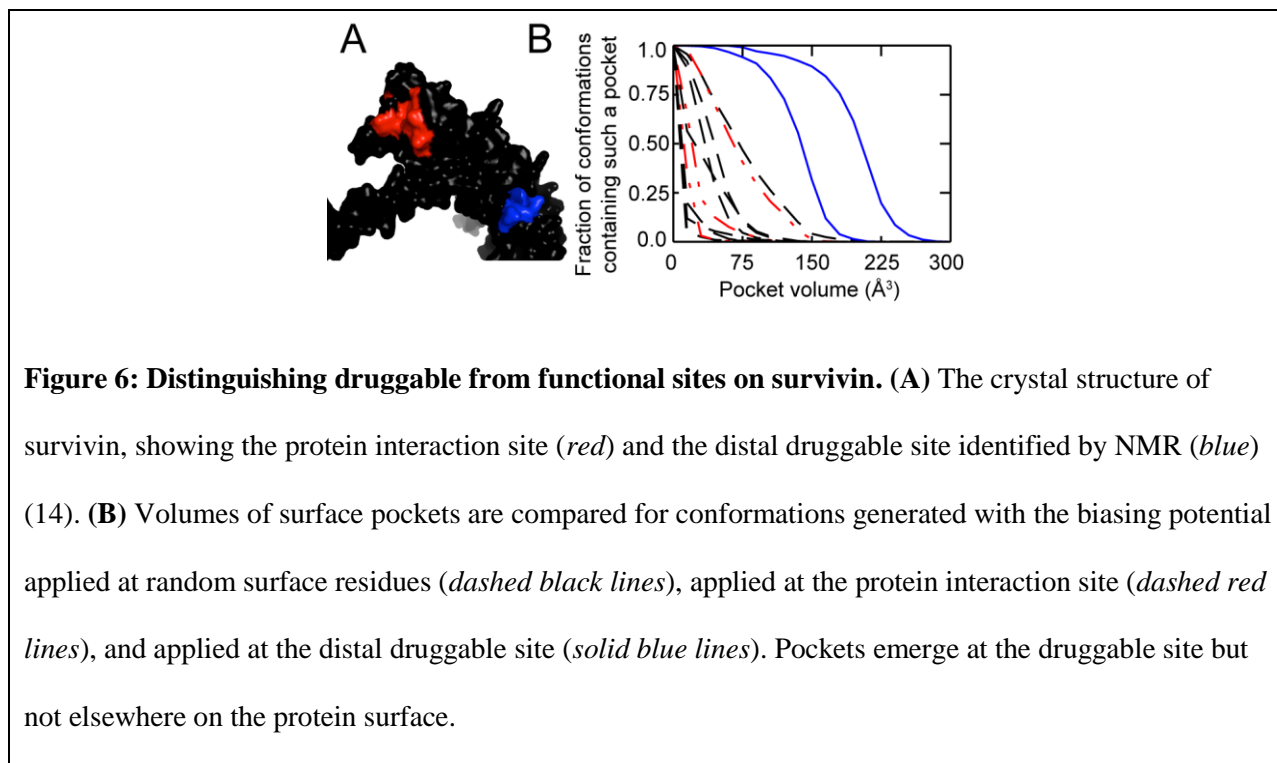
Druggability of survivin

The examples comprising our test set in the studies above were selected on the basis of a known compound directly inhibiting a protein interaction. For each of these cases, then, the druggable site is coincident with the protein functional site. We next turned to survivin, an example of a protein in which the two sites are non-overlapping.

Survivin is among the most strongly tumor-specific proteins known (57), is a notable signature of unfavorable disease outcome (58), and has been well-validated as a therapeutic target using an antisense oligonucleotide (59) and a transcriptional repressor (60, 61). Despite this intense interest, however, no direct inhibitors of survivin function have been identified. Survivin carries dual functions which together explain its important role in cancer: it serves as an inhibitor of apoptosis and is also required for cell division (62). The anti-apoptotic activity of survivin derives from binding the Smac/DIABLO peptide via a BIR domain. While small-molecule inhibitors of other Smac-binding BIR domains have been identified (63-66) (we included the XIAP BIR domain in the test set described in previous sections), efforts in this vein using survivin have not yet proven fruitful (62, 67). Consistent with a hypothesis that the peptide-binding surface of survivin may be intrinsically undruggable, an SAR by NMR approach used by Hajduk et al. found only a single probe that interacted with this site (of 9,370 probes tested) – a full order of magnitude fewer than the number of interacting probes found using proteins for which potent inhibitors have been identified (14). Intriguingly, this study revealed a separate distal site (**Figure 6A**) at which 33

probe compounds bound, well within the frequency observed for druggable sites on other proteins but not useful as a starting point for modulating survivin function.

We carried out a series of biased simulations as described earlier, using the unbound crystal structure of survivin as a starting point. In agreement with results from SAR by NMR (14), we find that pockets at the "undruggable" peptide-binding site of the BIR domain do not form with any higher propensity than at randomly selected equivalent surface sites (**Figure 6B**). This result is particularly notable as it stands in contrast to our observations from the XIAP BIR domain, in which surface pockets *are* formed at the peptide-binding site (**Figure 2G**). Though beyond the scope of the current work, we anticipate that further detailed studies may reveal the precise origin of the differing propensity for binding small-molecules exhibited by the peptide-binding groove of these two highly similar BIR domains.

Though large surface pockets do not form at the peptide-binding site of survivin, they *do* form readily at a specific alternate site: the same distal surface identified by Hajduk et al. (14) (**Figure 6B**, *blue lines*). Taken together, these studies of survivin suggest that low-energy fluctuations that induce pocket formation on the protein surface are indeed a primary determinant of druggability at protein interaction sites rather than a signature of the evolved protein-binding function of these sites.

Figure 6: Distinguishing druggable from functional sites on survivin. (A) The crystal structure of survivin, showing the protein interaction site (*red*) and the distal druggable site identified by NMR (*blue*) (14). **(B)** Volumes of surface pockets are compared for conformations generated with the biasing potential applied at random surface residues (*dashed black lines*), applied at the protein interaction site (*dashed red lines*), and applied at the distal druggable site (*solid blue lines*). Pockets emerge at the druggable site but not elsewhere on the protein surface.

1.5 Discussion

The results presented above support the hypothesis that the conformational transitions that open pockets at a druggable site occur with little energetic cost to the protein. This has two important consequences.

First, pocket formation within the ensemble of physiological protein fluctuations occurs only at a limited subset of surface sites. Because pocket-containing conformations are rare, it is primarily the lack of a suitable surface pocket – and not chemical composition – that renders most protein surface sites undruggable. The highly restricted druggability of protein surfaces supported by both NMR (14) and crystallographic (15) observations that probe molecules interact with a very limited subset of the protein surface. Sites at which the unbound protein crystal structure includes a pre-formed pocket (HPV E2, FKPB12, ZipA) typically exhibit only minor conformational changes that change the size or shape of this

pocket, and are driven primarily by sidechain reorganization. In contrast, for the examples in which the unbound protein crystal structure does not have a pre-formed pocket or has only a small surface pocket (XIAP, MDM2, Bcl-X$_L$, IL-2) concerted motions are needed that couple backbone conformational changes to the sidechain reorganization that reveals the binding pocket. In both cases, the biasing potential we describe here serves as a proxy for the binding energy associated with some (unspecified) complementary compound, and thus drives sampling towards these conformations.

Second, these results suggest that a small number of low-resolution, low energy inhibitor shapes are encoded on the protein surface through intrinsic structural and dynamic features of the protein. We propose that low-energy fluctuations produce protein conformations displaying one of these preferred shapes, at which point a complementary small-molecule ligand may be accommodated. The protein then responds to the particular steric and chemical details of the ligand via subsequent smaller conformational changes, without changing the gross features of the pocket. This model implies that exploring the pocket shapes may give clues at low-resolution to the shapes of complementary ligands, and may form the basis for a computational screening approach that matches pocket shapes to potential inhibitory ligands.

As pointed out by Cheng and colleagues (68), pocket formation is necessary but may not be sufficient for a protein surface to be druggable: the curvature and hydrophobicity of the pocket are also important. By applying their methodology to the pocket conformations generated using the biasing potential described here, it may be possible to filter for the most druggable conformations from among these pocket-containing ensembles. These conformations may not necessarily correspond to the protein-bound conformation, which has been used in some cases as a starting point for mimicry by small-molecules (31). Generating an ensemble of druggable conformations for a given protein target may prove valuable in identifying new inhibitory compounds, even in cases where one or more inhibitors are already known. Rather than aiming to identify additional inhibitors by analogy to these known compounds, or even using the protein structure solved in complex with one of these known compounds (or the natural protein partner), matching directly against an ensemble of pocket-containing conformations removes bias

towards this parent compound (or protein partner). We anticipate that the surface pockets to which the protein is most predisposed, as revealed by this approach and therefore identified without artifacts of conformational changes in response to any particular binding partner, may serve as an optimal starting point for computational screening and may facilitate identification of chemotypes unrelated to those of known inhibitors ("scaffold hopping" (3, 69)). These new scaffolds may in turn yield potent novel small molecule inhibitors through subsequent chemical elaboration.

In principle, fragment-based approaches allow multiple probe molecules to be subsequently linked together to create a single compound presenting these probe moieties in the appropriate orientation (70). Computational solvent mapping (49), while suitable for assessing druggability in certain cases, may prove limiting for early drug discovery because of its inability to explore the concerted backbone motions necessary for thoroughly sampling the "pocket ensemble". While FKBP sidechain motions within 6 Å of the binding site are sufficient for recognizing small probe molecules (70), the further necessity of backbone reorganization in other cases described here (XIAP, MDM2, Bcl-X_L, IL-2) underscores the importance of backbone motions for conformational fluctuations that reveal surface pockets. Exploring these alternate conformations may further prove useful in guiding efforts to improve potency of known inhibitors, by identifying sites at which substituents can make additional strong contacts with the protein surface.

Finally, we note that the generality of our approach is conducive to its large-scale application for comparing druggability across many protein interaction sites. In the future, we expect this approach may be used to address the outstanding question of whether these few successful examples of small-molecules disrupting protein interaction sites are outliers, or whether they instead represent the first step towards an important class of new tools for therapeutic intervention.

27

1.6 Methods

Identifying Pockets on Protein Surfaces

We implemented our modified version of the LIGSITE^{cs} (50) algorithm in the Rosetta software suite (56). Briefly, a grid is centered at the residue of interest on the protein surface and grid points are marked as occupied by either protein (P) or solvent (S). The algorithm performs linear searches in the X, Y and Z directions as well as in each diagonal for "P-nS-P" events: cases where a line no longer than 12 Å can be drawn between two points on the protein surface that pass through only solvent, establishing this solvent region as part of a surface pocket. Subsequent additional criteria were used to eliminate spurious definition of pockets and reduce the effect of grid-based artifacts (described in **Text S1**). Adjoining grid points defined as "pocket" were clustered to determine the deep pocket volume of the largest contiguous single pocket in contact with the target residue.

Simulation protocol

Simulations were carried out using the "relax" protocol (71) in Rosetta, which incorporates both backbone and sidechain degrees of freedom in a Monte Carlo search. To bias the simulations towards pocket-containing conformations we added to the standard energy function an additional energy term corresponding to the current deep pocket volume multiplied by a proportionality constant of -0.25 Rosetta energy units per Å^3 ("moderate" bias). Simulations carried out with the "strong" bias used a proportionality constant of -0.25 Rosetta energy units per Å^3. Data for each histogram was collected from 1,000 independent simulations. The simulation protocol is described in complete detail in **Text S1**.

1.7 Acknowledgements

CHAPTER 2: Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations

David K. Johnson[1] and John Karanicolas[1,2*]

[1] Center for Computational Biology and [2] Department of Molecular Biosciences,

University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534

*To whom correspondence should be addressed. E-mail: johnk@ku.edu, 785-864-8298

2.1 Abstract

Small-molecules that inhibit interactions between specific pairs of proteins have long represented a promising avenue for therapeutic intervention in a variety of settings. Structural studies have shown that in many cases, the inhibitor-bound protein adopts a conformation that is distinct from its unbound and its protein-bound conformations. This plasticity of the protein surface presents a major challenge in predicting which members of a protein family will be inhibited by a given ligand. Here, we use biased simulations of Bcl-2-family proteins to generate ensembles of low-energy conformations that contain surface pockets suitable for small molecule binding. We find that the resulting conformational ensembles include surface pockets that mimic those observed in inhibitor-bound crystal structures. Next, we find that the ensembles generated using different members of this protein family are overlapping but distinct, and that the activity of a given compound against a particular family member (ligand selectivity) can be predicted from whether the corresponding ensemble samples a complementary surface pocket. Finally, we find that each ensemble includes certain surface pockets that are not shared by any other family member: while no inhibitors have yet been identified to take advantage of these pockets, we expect that chemical scaffolds complementing these "distinct" pockets will prove highly selective for their targets. The opportunity to achieve target selectivity within a protein family by exploiting differences in surface fluctuations represents a new paradigm that may facilitate design of family-selective small-molecule inhibitors of protein-protein interactions.

2.2 Author Summary

Despite intense interest and considerable effort, there are few examples of small molecules that directly inhibit protein-protein interactions. Crystal structures of early successes have highlighted the plasticity of the protein surface, as some inhibitor-bound proteins are captured in conformations that are distinct from both their unbound and their protein-bound conformations. The lack of a single well-defined structure presents a challenge for predicting the members of a protein family to which a given compound will show activity (ligand selectivity). Here we generate ensembles of conformations from simulation, and show that we can predict ligand selectivity based on which family members sample conformations complementary to the ligand. This approach may present a new avenue for designing highly-selective inhibitors of protein-protein interactions.

2.3 Introduction

Selectivity of a compound for its desired protein target – or targets – is an important property optimized in the course of small-molecule drug discovery (12). Some diseases, such as chronic myeloid leukemia, can be traced to dysfunction of a single protein target (BCR-ABL); in such cases, drugs such as imatinib are sought to act selectively against that target (72). Conversely, a number of drugs (such as sunitinib and chlorpromazine) have proven exceedingly successful because they act on multiple targets (73, 74); this has led to increased interest in "polypharmacology" to address complex disease states such as cancer and psychiatric conditions (16). A clear downside of compound promiscuity, however, is the potential for adverse effects (toxicity) due to interactions with unrelated proteins, or even interactions with other proteins in the same family as the target (17). One recent cautionary example is ABT-263 (navitoclax), a Bcl-2 inhibitor that exhibited a dose-limiting adverse effect (thrombocytopenia) stemming from its inhibition of Bcl-xL (18, 19).

Tuning the selectivity of a lead compound can sometimes be carried out by exploiting differences in shape and electrostatics between target and off-target proteins, using insights from structure-activity relationships (SAR) (20) or structural biology (21-23) to focus on features that will prevent specific undesirable interactions. While determinants of selectivity have been carefully mapped in a number of "traditional" drug targets, such as kinases, this has not yet been the case for emerging classes of "non-traditional" drug targets, most notably small-molecule inhibitors of protein-protein interactions.

Though extensive efforts to identify inhibitors of protein interactions have only recently begun to bear fruit, structural analysis of available examples has revealed that often the inhibitor-bound protein is captured in a conformation that is distinct from both the unbound and protein-bound conformations (75). In such cases, the unbound and protein-bound conformations could not have served to rationalize binding of the inhibitor its desired target, let alone explain selectivity against other members of the protein family.

We recently described an approach for biased exploration of protein fluctuations, in order to better sample pocket-containing conformations at protein interaction sites (76). We found that when starting from an unbound protein structure, we observe low-energy conformations that include deep surface pockets at druggable sites but not elsewhere on the protein surface. These findings supported a "conformational selection" model (77), whereby inhibitors recognize low-lying excited states of the protein that are naturally visited under physiological conditions.

A natural implication of this conformational selection model is that the particular variety of pocket shapes visited by the protein surface will dictate the regions of chemical space in which complementary inhibitors can be found: this would have clear implications for designing new inhibitors. Conversely, one may instead view this complementarity from the perspective of the *ligand*: under this model, an inhibitor is expected to be active against a given protein if, and only if, the protein surface includes a suitably complementary pocket among those that are sampled under physiological conditions. Within a protein family, then, only the subset of family members that sample the corresponding pocket will be inhibited by this compound.

The Bcl-2 protein family in humans is comprised of about 25 members that together regulate apoptosis through a series of protein-protein interactions that induce either pro-death (Bax, Bak, and others), pro-survival (Bcl-xL, Bcl-2, Mcl-1, Bcl-w, and others), or derepression/sensitizing (Bid, Bim, and others) activity (78). The critical step of disregulating apoptosis in tumor maintenance and chemoresistance has made certain members of this family exceptionally attractive targets for therapeutic intervention for many years, and in a variety of cancers (79). Despite the overall dearth of examples of small molecules that inhibit protein interactions, a number of such compounds have been reported against assorted members of the Bcl-2 family. Further, selectivity of these inhibitors against a panel of Bcl-2 family members have also been reported in some cases. These data, together with the availability of multiple experimentally-derived structures of inhibitor-bound complexes, position the Bcl-2 family as a

rich model system in which we can explore the determinants of selectivity for small-molecule inhibitors of protein-protein interactions.

Here, we use the Bcl-2 family to explore whether the ensemble of pocket shapes sampled on protein surfaces can be used to explain ligand selectivity. We will use simulation to generate ensembles of pocket-containing conformations (76), then directly probe whether a complementary pocket for a given ligand is present in the ensemble. If successful, this approach may allow us ultimately to predict, and by extension *design*, inhibitors with a desired selectivity profile; such compounds could serve both as a starting point for development of new therapeutics, and also as "tool compounds" to probe the underlying biological mechanism of disease.

2.4 Methods

Computational methods are implemented in the Rosetta software suite (56); Rosetta is freely available for academic use (www.rosettacommons.org), with the new features described here included in the 3.6 release. Computational methods are summarized below, with further detail (including Rosetta command-lines to access these tools) is available in the *Supporting Methods* section.

Identifying surface pockets

Pockets were identified using the "pocket" protocol in Rosetta (76). This is a local pocket detection approach adapted from the Ligsite algorithm (52), and it uses geometric criteria to identify concave regions on a protein surface suitable for small-molecule binding. The approach entails building a small grid in the vicinity of one or two "target" residues, then mapping the van der Waals surface for the whole protein onto the grid; the exposed grid points are marked as "solvent." Next, the grid is screened to identify linear segments of solvent points that are bounded by protein points; the grid points along these segments are marked as "pocket" points. Pocket grid points are clustered into "pockets," and then any clustered pockets not in direct contact with the target residue(s) are discarded. The "deep volume" of a

pocket is defined as the volume of the pocket that is well-sequestered from the solvent (more than 2.5 Å from any solvent grid point).

Generation of a representative protein surface pocket is demonstrated in **Figure 1**. A number of related approaches have been described for identifying potential small molecule binding sites on protein surfaces in this way (51, 52, 80-84); the primary difference between these methods relative to the method described here is our use of "deep" pocket volume, which leads to improved discrimination when distinguishing known ligand binding sites from other shallow concave regions on the protein surface (76). These "deep" pocket volumes are correlated to, but smaller than, pocket volumes identified by other representative methods (76); in the context of this study, then, we expect that focusing our pocket description on regions in direct contact with the protein will allow the most critical features to be captured with minimal extraneous information.
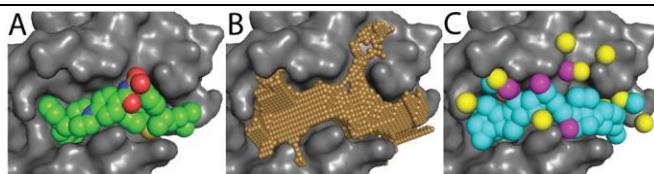

Figure 1: Building "exemplars" from surface pockets. (A) Bcl-xL (*grey surface*) is shown in complex with an inhibitor (*spheres*). **(B)** The protein surface features a large pocket (*small white spheres*) that is complementary in shape to the inhibitor. **(C)** From this surface pocket, an "exemplar" is built: a map of the "perfect" ligand to complement this protein surface, without considerations of atom connectivity. The exemplar is comprised of hydrogen bond donors (*yellow*) and acceptors (*magenta*) that complement surface polar groups on the protein, and hydrophobic atoms that fill the remainder of the surface pocket (*cyan*). In the studies we report here, we will use the shape and chemical features of the exemplar as a proxy for the shape and chemical features of the protein surface pocket.

Target residue selection

The selection of the target residue pairs is important in determining the location of the grid, and in turn may affect the resulting pocket shapes. To avoid biasing results towards any particular inhibitor – and to allow the method to be applied in cases for which no inhibitor has yet been identified – we did not use the structure of the protein in complex with any small-molecule ligand when defining the target residues.

Instead, we developed an algorithm to select suitable target residues starting from the structure of the protein-protein complex. First, the "Robetta" server (85, 86) is used to estimate the contribution to the binding free-energy of each interfacial sidechain (ΔΔGres) via computational alanine scanning. All pairs of interfacial residues are then exhaustively tested by building a 24 Å candidate grid placed at the residue pair's center of mass, and summing ΔΔGres for each residue that falls on this grid. To ensure that the grid placement captures the key energetic contributors to the protein-protein interaction, the residue pair that captures the largest cumulative ΔΔGres is used in subsequent studies. In essence, this approach aims to align the pocket grid such that it is optimally close to the energetically dominant residues of the protein-protein interaction. This tool is implemented in the Rosetta software suite, and its use it demonstrated in the *Supporting Methods* section (56).

The effect of the particular target residues used will be examined further in the *Results* section.

Building ensembles of pocket-containing conformations

Ensembles of pocket containing conformations were generated using the "relax" protocol in Rosetta, which incorporates both backbone and side chain degrees of freedom in a Monte Carlo search. Starting from the standard Rosetta energy function, we added a term corresponding to the "deep pocket" volume multiplied by a proportionality constant. This biasing potential favors pocket-containing conformations that have more deep pocket volume, and thus drives sampling towards such conformations.

In our previous work, we found that applying a proportionality constant of -0.25 Rosetta energy units per $\mathring{A}^3$ allowed identification of pocket-containing conformations with similar energies to those observed in unbiased simulations (76). In the present study, therefore, we reused the same proportionality constant. A separate independent trajectory was used to generate each of the 1000 output conformations used for each protein included in our studies. It took about 5 minutes to generate each conformation on a modern CPU, though the independence of trajectories made it trivially scalable to multiple processors.

Generating "exemplars" to represent pockets

To compare the shape and chemical similarity of surface pockets, we introduce the concept of an exemplar: a map of the "perfect" ligand that could complement a surface pocket, without the natural constraints of bond connectivity or chemistry. The lack of such physical requirements means that the exemplar does not correspond to any particular chemical entity, but instead simply consists of a collection of atoms in space. This collection occupies volume, and may include hydrogen bond donors and acceptors, but by construction does not imply any particular connectivity between the atoms.

Exemplars are built from the "deep volume" that defines a protein surface pocket. First, the ideal location for a hydrogen bond partner of every donor/acceptor on the protein surface is determined; if this location lies within the pocket, then this region of the pocket volume is reserved for the polar group that will complement the protein surface. After placing these polar groups in the pocket, the remaining unoccupied volume is filled with hydrophobic (carbon) atoms using a greedy algorithm, such that the center of two atoms are no less than 1.7 $\mathring{A}$ apart. Exemplar points are then clustered together based on a proximity threshold of 5 $\mathring{A}$, so that cases in which two small pockets flank the target residues are represented by a single exemplar. Generation of an exemplar from a representative protein surface pocket is demonstrated in **Figure 1**; complete details of our implementation are included in *Supplementary Methods*.

These exemplars are analogous in philosophy to the popular pharmacophore maps used in medicinal chemistry to reflect consensus properties of known ligands (87, 88); while the latter approaches rely on mimicry of the existing binding partners (31, 32, 89, 90), however, exemplars are instead built purely from features of the protein surface. Recently, the utility of protein-based pharmacophores has been explored for pose prediction and virtual screening (91-93), but such approaches have not been used for the comparison of pocket shapes as described below.

Comparing exemplars

Having represented the shape and chemical features of a protein surface pocket by its "perfectly complementary ligand" (the exemplar), we are now positioned to quantitatively compare pockets using standard tools developed for comparison of chemical entities in 3D, provided that these tools do not require (or assume) knowledge of bond connectivity. A variety of ligand-based shape comparison methods have been developed for aligning pairs of molecules on the basis of volume overlap (94, 95); a convenient modern implementation of this approach is the ROCS software (OpenEye Scientific Software, Santa Fe, NM) (3, 96). ROCS represents molecular shape as the sum of Gaussians centered at each atomic position, and can rapidly calculate the near-optimal alignment that maximizes volume overlap between two molecules. Chemical groups are also included as separate Gaussians in the scoring/optimization step, to include electrostatic effects.

The results we present below are built upon quantitative comparisons of the similarity in shape and hydrogen bonding patterns for pairs of protein surface pockets. In all cases we carry out this analysis by generating an exemplar for each of the two pockets, then using ROCS to align and score the overlap between these exemplars.

We note that method to define pockets, and thus exemplars, makes use of a grid centered at a pair of target residues. To quantitatively examine the effect of the grid orientation on the exemplar similarity, we used four pocket-containing crystal structures and rotated each to 100 random orientations. Using one

of these structures as a reference (ABT-737 bound to Bcl-xL), we found that rotated variants of this structure produced similar exemplars, as did rotated variants of a related structure (Bcl-xL in complex with a related ligand). In contrast, rotated variants of Bcl-xL bound to unrelated ligands, or rotated variants of Mcl-1, produced highly dissimilar exemplars (**Figure S1**). Collectively, these observations demonstrate that the orientation dependence of generating exemplars on a grid makes a negligible contribution to the exemplar distances we report below.

2.5 Results

The principal goal of the studies we report here is to understand how conformational changes at a protein interaction site can drive selectivity of the small-molecule inhibitors that bind at these sites. We have developed an approach for building an "exemplar" – a map of an idealized ligand – that describes a protein surface pocket and enables quantitative comparisons between pockets (**Figure 1**, *Methods* section). In each of the sections below, we use exemplar comparisons to examine the similarity of pockets observed on the surface of proteins comprising the Bcl-2 family.

Plasticity of the protein surface allows recognition of diverse chemical scaffolds

There have been dozens of reported inhibitors of Bcl-2 family members (18, 19, 97-115), spanning a broad range of chemotypes. Unfortunately, experimentally-derived structures are not available for the majority of these inhibitors in complex with their cognate protein partner(s); this makes it difficult to gain insight into the detailed basis for molecular recognition in these cases. Instead, we began by compiling a comprehensive collection of all structures in the Protein Data Bank containing a Bcl-2 family member in complex with a non-fragment small-molecule inhibitor bound at the protein interaction site: these are listed in **Table S1**. There are 28 such complexes, covering 26 unique inhibitors, and 3 different proteins are represented: Bcl-xL (14 structures), Bcl-2 (9 structures), and Mcl-1 (5 structures).

The diversity of the inhibitors in this set is readily apparent from the Tanimoto similarity of fingerprints describing each chemical structure (**Figure 2a**). Unsurprisingly, groups of compounds that are similar by this measure typically represent a chemical series designed by a single research group (e.g. compounds *1-3* from WEHI (116)). Given this collection of chemical scaffolds, we sought to ask how such diverse compounds could be recognized on the surface of a single protein family. There are two possibilities: either these compounds might adopt a shared three-dimensional structure not evident from their chemical structure (i.e. a non-obvious example of "scaffold hopping" (117)), or else the protein surface must be sufficiently malleable to adopt different conformations when binding different ligands.

To test whether these distinct compounds present a common structure to complement the protein surface, we carried out comparisons of the inhibitors' shape and chemical features in their active (bound) conformations (**Figure 2b**). To examine the pattern of similarity between chemical structure and three-dimensional structure we sorted the all-vs-all Tanimoto scores for each set, and found a statistically significant non-zero Spearman correlation coefficient between these rankings ($p < 10^{-35}$). This observation confirms that dissimilar chemical structures do *not* somehow adopt a shared three-dimensional structure. As expected, a given shape and pattern of hydrogen bond donors/acceptors is conserved within a chemical series, but not across different chemical scaffolds.

To directly evaluate similarity of the conformations adopted by the protein to bind each ligand, we next carried out an analogous comparison using the exemplar derived from each inhibitor-bound pocket (**Figure 2c**). Here again we observe the same pattern of similarity, mimicking both that of the chemical structures ($p < 10^{-26}$) and their corresponding three-dimensional structures ($p < 10^{-20}$). While each of the inhibitors within a given chemical class bind to a very similar pocket on the surface of the cognate protein, different chemical classes each take advantage of a dramatically different pocket on the protein surface. Moreover, the surfaces of different proteins bound to similar ligands (e.g. Bcl-xL complex *7* vs Bcl-2 complex *19*) resemble one another more closely than the surface of a single protein bound to chemically distinct ligands (e.g. Bcl-xL complexes *3* vs *7*).
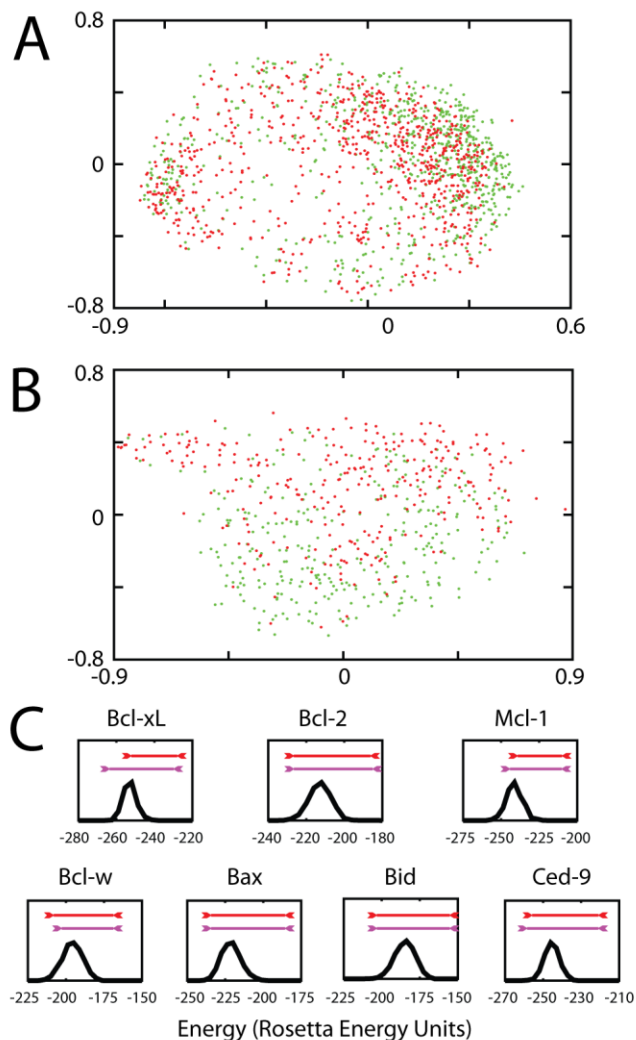
Figure 2: Bcl-2 family members recognize different inhibitors using distinct surface pockets. In all cases color gradient indicates the similarity between complexes, expressed as Z-scores (*green* are most similar, *red* are most dissimilar). **(A)** Chemical similarity of the inhibitors. **(B)** Three-dimensional similarity of the inhibitors' active conformation. **(C)** Similarity of protein surface pockets, measured using exemplar similarity. Numbering in all cases corresponds to complexes in **Table S1**. A representative subset of the complexes are included in this figure; a corresponding figure containing all available complexes is available as **Figure S2**.

41

These observations highlight the plasticity of this protein surface: multiple members of the Bcl-2 family can adopt similar conformations to bind a given ligand, yet a given protein can also form radically different surface pockets to accommodate different ligands. Together with the fact that the unbound structures of these Bcl-2 family members lack suitably deep surface pockets for inhibitor binding (**Figure S3**), the observations presented here underscore the fact that molecular recognition in this protein family cannot be explained using a single conformation, but rather an explanation of ligand selectivity will instead require consideration of the many available conformations that this surface can adopt.

Ensembles of low-energy pocket-containing conformations

To explore pocket-containing conformations of these proteins, we generated conformational ensembles using simulations in the presence of a biasing potential (76). In essence, energy associated with the biasing potential in these simulations serves as a proxy for the binding energy of some (unspecified) ligand, which in turn serves to stabilize alternate conformations of the protein surface. The biasing potential takes account of the protein surface geometry ("deep pocket volume"), but does not encode any information about the identity or features of any particular ligand. Thus, this approach allows efficient sampling of the conformational space available for protein reorganization in response to ligand binding. While other methods could have been employed to sample alternative conformations (such as unbiased molecular dynamics simulations with retrospective analysis to identify pocket-containing conformations (118)), we chose this approach because it allowed us to rapidly generate a large ensemble of pocket containing conformations.

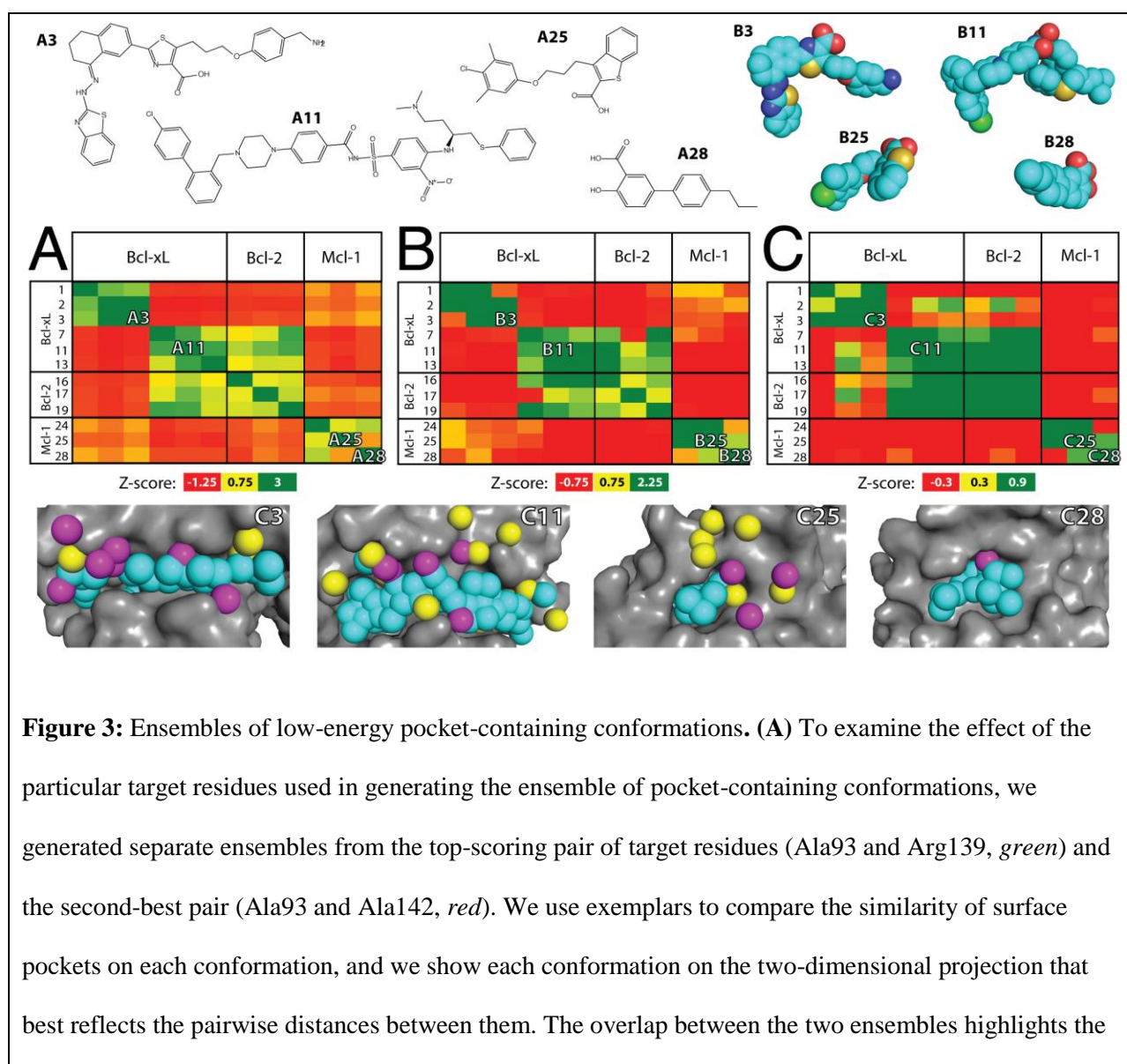As noted earlier, our approach defines the relevant protein surface on the basis of the protein/peptide interaction site, and not based on any known small-molecule inhibitor (*Methods* section). In light of the fact that particular "target" residues are needed to produce the ensembles of pocket-containing conformations, we begin by examining the effect of the target residues on the resulting "pocket ensemble."

Starting with the Bcl-2 family member Bcl-xL, we selected the two top-scoring pairs of target residues resulting from analysis of its peptide-bound structure. We then used each pair of target residues to generate separate ensembles of 2000 pocket-containing conformations, with each trajectory initiated from the unbound structure of Bcl-xL. To visually compare the conformational "pocket space" sampled by these two ensembles we built an exemplar from each pocket, and used multidimensional scaling analysis (MDS) to construct the two-dimensional projection that best reflects the pairwise distance between every pair of exemplars. The resulting map demonstrates that these two ensembles are strongly overlapping, and points to the robustness of the conformational space sampled to the particular target residues used (**Figure 3a**). In light of the similarity between the ensembles, we used conformations generated using the second-best scoring pair of target residues (rather than the top-scoring pair), for consistency with our previous studies of Bcl-xL (76).

While application of the same approach for selecting target residues led to similar residues for most members of the Bcl-2 family, the residue pair selected on Mcl-1 was notably different. While this derived exclusively from different energetic contributions to the protein-protein interaction from each protein surface, it is interesting to note that known inhibitors of Bcl-xL and Mcl-1 bind at slightly different regions on the surface of these two proteins (**Figure S4**). To ensure that the range of pockets across the Bcl-2 family would be fully captured in our studies, we used *both* the pair of residues derived from Bcl-xL and those derived from Mcl-1 to generate ensembles of pocket-containing conformations; in the analyses presented below, all such conformations are combined into a single ensemble regardless of the target residues used to generate them.

To ensure that only physiologically relevant conformations were included in the subsequent analysis, we compared their energies to those obtained in equivalent unbiased simulations of the corresponding protein. Previously, we found that this method produced ensembles of pocket-containing conformations a slightly higher but overlapping distribution of energies than the unbiased ensemble (76). With the caveats that conformations generated in our "pocket opening" protocol are not drawn from a

Boltzmann distribution and the energy differences do not necessarily capture real differences in free energy, we instead simply collect pocket-containing conformations from the biased simulations that are within 15 Rosetta energy units of the unbiased ensemble. Thus, these represent conformations that are in principle available within the unbiased ensemble, but are simply not observed due to limitations of sampling. The distribution of energies observed in an unbiased simulation of each Bcl-2 family member is presented in **Figure 3b**, along with the range of energies spanned by conformations from biased simulations.

Figure 3: Ensembles of low-energy pocket-containing conformations. **(A)** To examine the effect of the particular target residues used in generating the ensemble of pocket-containing conformations, we generated separate ensembles from the top-scoring pair of target residues (Ala93 and Arg139, *green*) and the second-best pair (Ala93 and Ala142, *red*). We use exemplars to compare the similarity of surface pockets on each conformation, and we show each conformation on the two-dimensional projection that best reflects the pairwise distances between them. The overlap between the two ensembles highlights the

robustness of the conformations to the particular target residues. **(B)** For each member of the Bcl-2 family, we generated an ensemble of conformations from unbiased simulations; the distribution of these energies is shown (*black*). The range of energies for pocket-containing conformations generated using a biasing potential target residues derived from the Bcl-xL protein interaction site (*magenta*) or the Mcl-1 protein interaction site (*red*) suggest that many of these conformations are energetically accessible to these proteins under physiological conditions. All energies shown here were evaluated in the absence of the biasing potential, for fair comparison. Each simulation is started from the structure of the unbound protein; a corresponding figure starting from the peptide-bound structures containing all available complexes is available as Figure S5.

Ensembles of conformations sampled by Bcl-2 family members include the inhibitor-bound pockets

We next sought to compare the surface pockets on conformations comprising these ensembles to those pockets observed on experimentally-derived inhibitor-bound structures. We again turned to MDS analysis, this time including exemplars built from the experimentally-derived structures of the unbound, peptide-bound, and inhibitor-bound protein. For each of Bcl-xL, Bcl-2, and Mcl-1 (all family members for which structures of inhibitor-bound complexes are available), the resulting maps show that the pocket-containing conformational space sampled via simulation includes thorough coverage of experimentally-derived inhibitor-bound structures (**Figure 4**). Though each simulation was initiated from either the unbound or the peptide-bound crystal structure, the resulting ensemble is notably distinct from these starting states; instead, the biasing potential drove sampling towards pocket-containing conformations that include examples very similar to the inhibitor-bound structures. In addition, each protein also samples a collection of conformations (marked "*D*") with exemplars that differ from those observed in any available inhibitor-bound conformations: these will be discussed in more detail in the following section.

In all three cases we find that these ensembles, generated without any prior information about the inhibitors, span the space of known inhibitors. This indicates that each protein is predisposed to adopt the

particular pocket shapes observed in the corresponding inhibitor-bound structures: these are not

conformational changes that are "induced" by the inhibitor, but rather these are among a suite of available

conformations from which the inhibitor may select. Upon binding, meanwhile, smaller changes to the

protein surface may then occur in response to particular features of the ligand (such as reorientation of
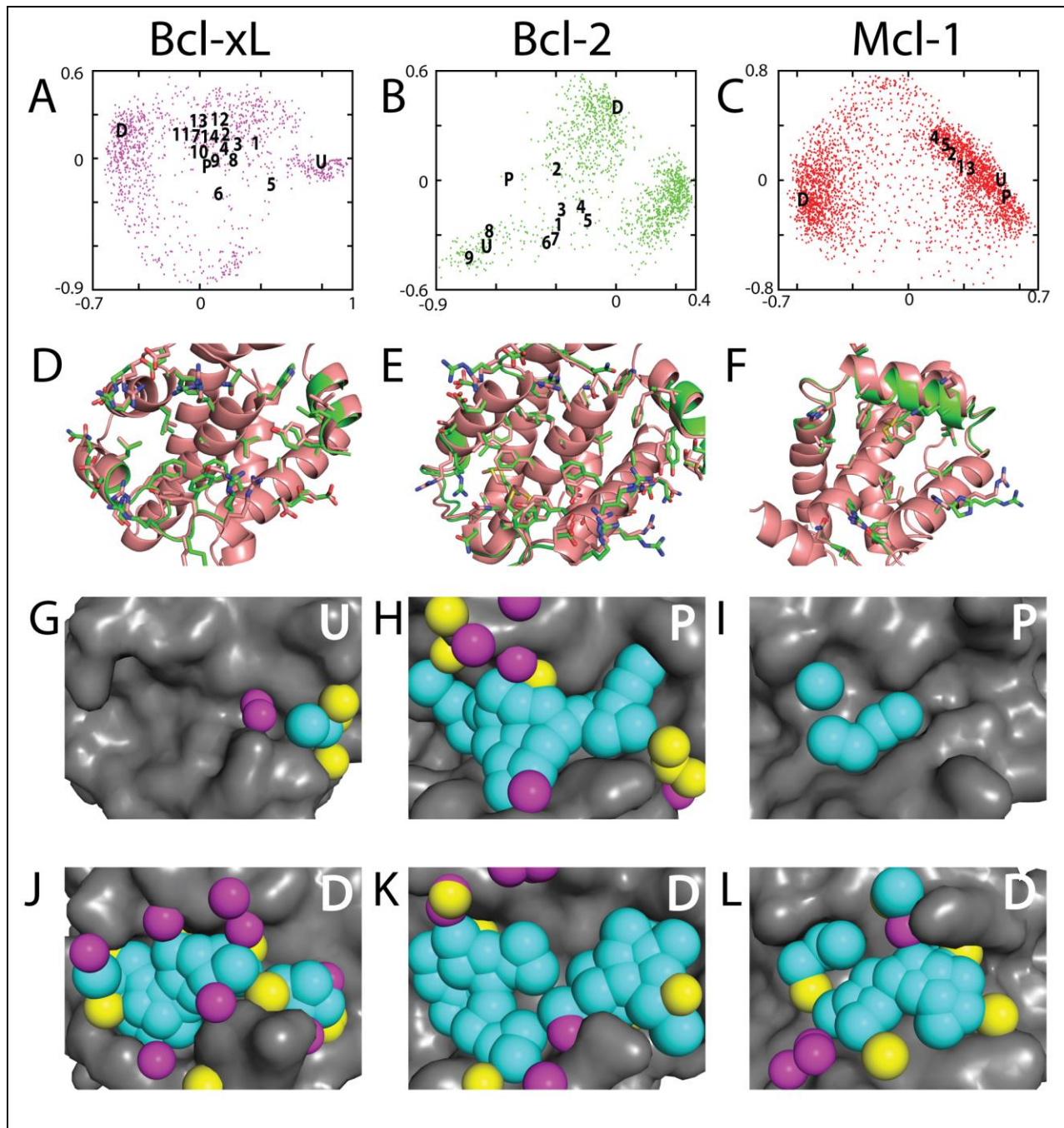
hydrogen bond donors and acceptors).

Figure 4: Maps of "pocket space" sampled by individual Bcl-2 family members. The ensemble of pockets observed from simulation: individual conformations are represented as points on a two-dimensional projection that reflects the pairwise distances between their exemplars. The relative position of exemplars from experimentally-derived Bcl-xL unbound ("*U*") and peptide-bound ("*P*") structures are indicated, as are the positions of exemplars from Bcl-xL structures solved in complex with various inhibitors (numbers correspond to complexes listed in Table S1)**.** Exemplars marked "*D*" correspond to the same "distinct" conformations described in Figure 5.

Different Bcl-2 family members sample different ensembles of surface pockets

To explore differences between ensembles of pocket-containing conformations of Bcl-2 family members, we carried out an analogous MDS analysis comparing exemplars from multiple family members; an important aspect of exemplar generation and comparison is that the description of a surface pockets is not tied to the sequence of the protein, allowing one to compare exemplars on the surfaces of different proteins. We compiled conformations generated from simulations of each Bcl-2 family member (Bcl-xL, Bcl-2, Mcl-1, Bcl-w, Bax, Bid, and Ced-9), and carried out MDS analysis using the complete set of exemplars (**Figure 5a**). Unsurprisingly, we observe that different family members sample different surface pockets; however, we do not observe such differences between the (closely related) human and mouse Mcl-1 sequences (**Figure S6**).

In these projections we again note that the ensemble generated for a given protein spans the majority of known inhibitors of that protein. Unsurprisingly, given the similarity we observed between inhibitor-bound Bcl-2 pockets and inhibitor-bound Bcl-xL pockets (**Figure 2c**), we find that most pockets observed in the inhibitor-bound structures occupy a similar region on this projection (all except *1*, *5*, *6*, and *23*). Notably, many known inhibitors fall in a region of "pocket space" that is sampled by more than one protein. In the case of Bcl-xL, for example, many known inhibitors bind to pockets that are not only observed in simulations of Bcl-xL, but also in simulations of Bcl-2. On this projection, these inhibitors

47

(all except *1*, *5*, and *6*) reside in a region sampled by both Bcl-xL and Bcl-2. Similarly, most of the Bcl-2 inhibitors (all except *23*) are in the same region sampled by both Bcl-xL and Bcl-2; in contrast, this other Bcl-2 inhibitor (*23*) overlaps with a regions sampled by Mcl-1 but not Bcl-xL.

In addition to regions shared by more than one family member, each map contains a "distinct" region that is sampled exclusively by a single protein (**Figure 5a**, "*D*"). As expected, exemplars corresponding to these conformations are very different from those of the (unbound or peptide-bound) conformations from which the corresponding simulations were initiated (**Figure 5b**). Further comparison of the conformations themselves show that these alternate conformations are accessed primarily through concerted reorganization of the sidechains that comprise the surface pocket, though corresponding changes to the protein backbone – especially in the case of Mcl-1 – are also required to enable this reorganization (**Figure 5c**). In summary, relatively modest structural changes to the protein conformation can produce radically different exemplars, and none of the inhibitors described to date bind to these particular protein conformations. To facilitate further study of these alternate protein conformations, we have made several such representatives publicly available on Proteopedia (119, 120) (http://proteopedia.org/wiki/index.php/User:John_Karanicolas/Selectivity_by_small-molecule_inhibitors_of_protein_interactions_can_be_driven_by_protein_surface_fluctuations , with the first model for each protein corresponding to the conformation we highlight here).

The regions that we have described as "distinct" on these maps correspond to pockets that are only sampled by a single member of the Bcl-2 family. In other words, these "distinct" pockets are far from each of the pockets sampled by other family members. So far, we have identified these regions visually on the basis of a two-dimensional projection of "pocket space"; to avoid the loss of information associated with reduction of dimensionality, however, we can instead identify "distinct" pockets using exemplar distances directly.
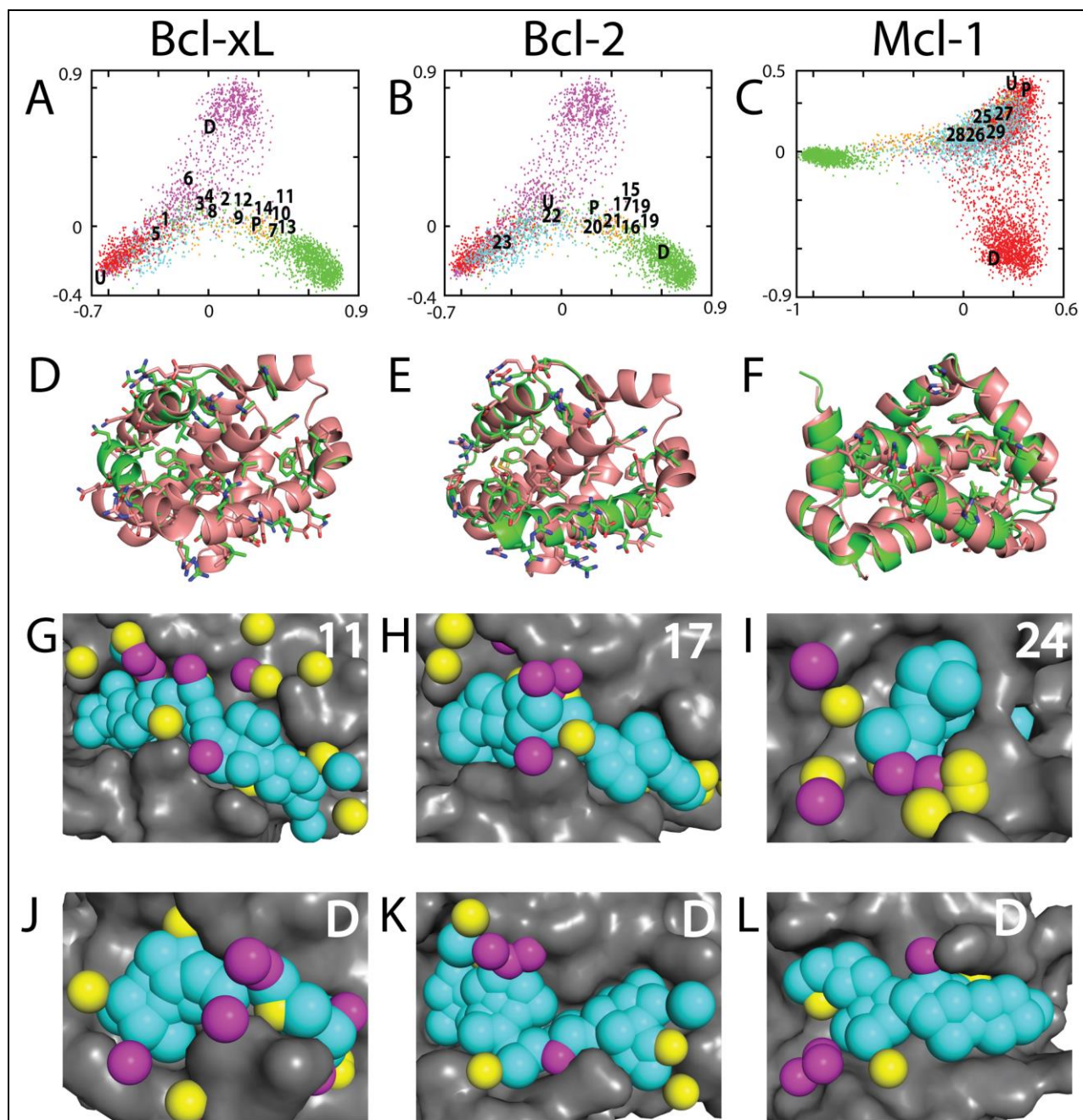
Figure 5: Comparison of "pocket space" sampled by each Bcl-2 family member. **(A)** A projection built using ensembles collected from simulations of several Bcl-2 family members: Bcl-xL (*magenta*), Bcl-2 (*green*), Mcl-1 (*red*), Bcl-w (*orange*), and Bax (*cyan*). Bid and Ced-9 were used in generating the projection, but for clarity are not included on this map. Target residues derived from the Bcl-xL protein interaction site were used in generating exemplars shown on the Bcl-xL and Bcl-2 MDS plots, whereas

target residues derived from the Mcl-1 protein interaction site were used in generating exemplars shown on the Mcl-1 MDS plots. For each of Bcl-xL, Bcl-2, and Mcl-1 we observe a distinct region of conformational space ("*D*") that is not sampled by any other Bcl-2 family member. **(B)** Comparison of an exemplar from each "distinct" region to the corresponding unbound ("*U*") or peptide-bound ("*P*") protein structure from which the simulation was initiated. **(C)** Comparison of the conformation harboring the "distinct" pocket to the corresponding unbound/peptide-bound protein structure from which the simulation was initiated.

For each conformation comprising the ensembles used above, we found the exemplar distance of the closest pocket sampled by a different family member. To avoid describing rarely-sampled (outlying) conformations as distinct, we subtracted from this the exemplar distance of the closest pocket from one's own ensemble. This measure, that we will call "distinctness," is largest for conformations taken from regions of pocket space that are well-sampled within a given ensemble, but not visited in the ensembles of the other family members.

We evaluated the "distinctness" of each conformation in the ensembles presented above, and compiled the results for each family member into a histogram (**Figure 6**). As expected, these results are consistent with our observations from the MDS analysis presented above: all of the pockets sampled by Bcl-w strongly resemble pockets from other family members and are therefore not "distinct," whereas Bcl-xL and Mcl-1 each sample certain highly distinct regions. Particularly striking from this analysis is the "distinctness" of experimentally-derived inhibitor-bound structures of Bcl-2 family members: none of these inhibitors take advantage of the "highly distinct" pockets available on the surfaces of Bcl-xL or Mcl-1. Rather, each of these compounds targets a pocket that is sampled not only by the cognate binding partner, but also by at least one other member of the Bcl-2 family.
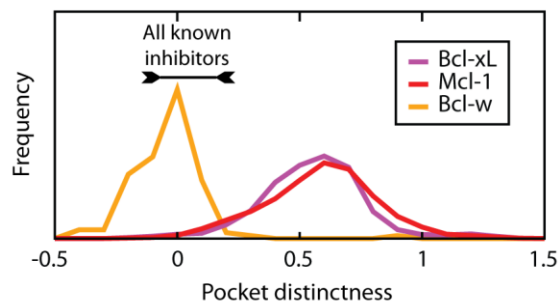
Figure 6: Ensembles of available pocket shapes contain distinct pocket shapes. We define the "distinctness" of a pocket as the difference in exemplar distances of the closest conformation from a different family member, and the closest conformation from one's own ensemble. Histograms are shown over conformations that comprise the ensembles used above. By this measure, *all* known inhibitors of *all* Bcl-2 family members bind to pockets that are not unique to their cognate target protein (i.e. low "distinctness"). Data are shown for three representative Bcl-2 family members, the complete set are included as **Figure S7**.

Pocket shape similarity as a predictor of ligand selectivity

As noted in the MDS analysis, the majority of known Bcl-xL inhibitors fall in a region of "pocket space" that is sampled by both Bcl-xL and Bcl-2 (**Figure 5a**). Given that Bcl-2 is found to form pockets that are similar to these inhibitor-bound Bcl-xL pockets, one might expect that these compounds would inhibit Bcl-2 in addition to Bcl-xL. Conversely, in light of the conformational selection model presented earlier, one would expect that the *lack* of similar pockets in the ensembles generated for other Bcl-2 family members might suggest that these proteins would *not* be inhibited by these compounds.

To explore this idea, we measured the exemplar distance between each known inhibitor bound pocket and each pocket observed in a given ensemble (starting from an unbound or peptide-bound structure). We then compared the most similar exemplar distances for each protein/inhibitor pair to

51

experimentally-derived binding data, to evaluate whether the pockets sampled in these ensembles dictate the Bcl-2 family members that will be inhibited by a given compound.

The results from this analysis are presented in **Figure 7a**. Using pockets observed in inhibitor-bound crystal structures of Bcl-xL (*1-14*), we find in many cases that highly similar pockets are sampled in ensembles generated from simulation of Bcl-xL and Bcl-2 (*green*), but not in the corresponding ensembles from Mcl-1 or Bcl-w (*yellow/red*): this represents a quantitative recapitulation of our observation that these ligands occupy surface pockets accessible only to Bcl-xL and Bcl-2 (**Figure 5a**). In light of this finding such compounds would be expected to bind Bcl-xL and Bcl-2, but not Mcl-1 or Bcl-w; available experimental binding data confirm that indeed this is generally the case.
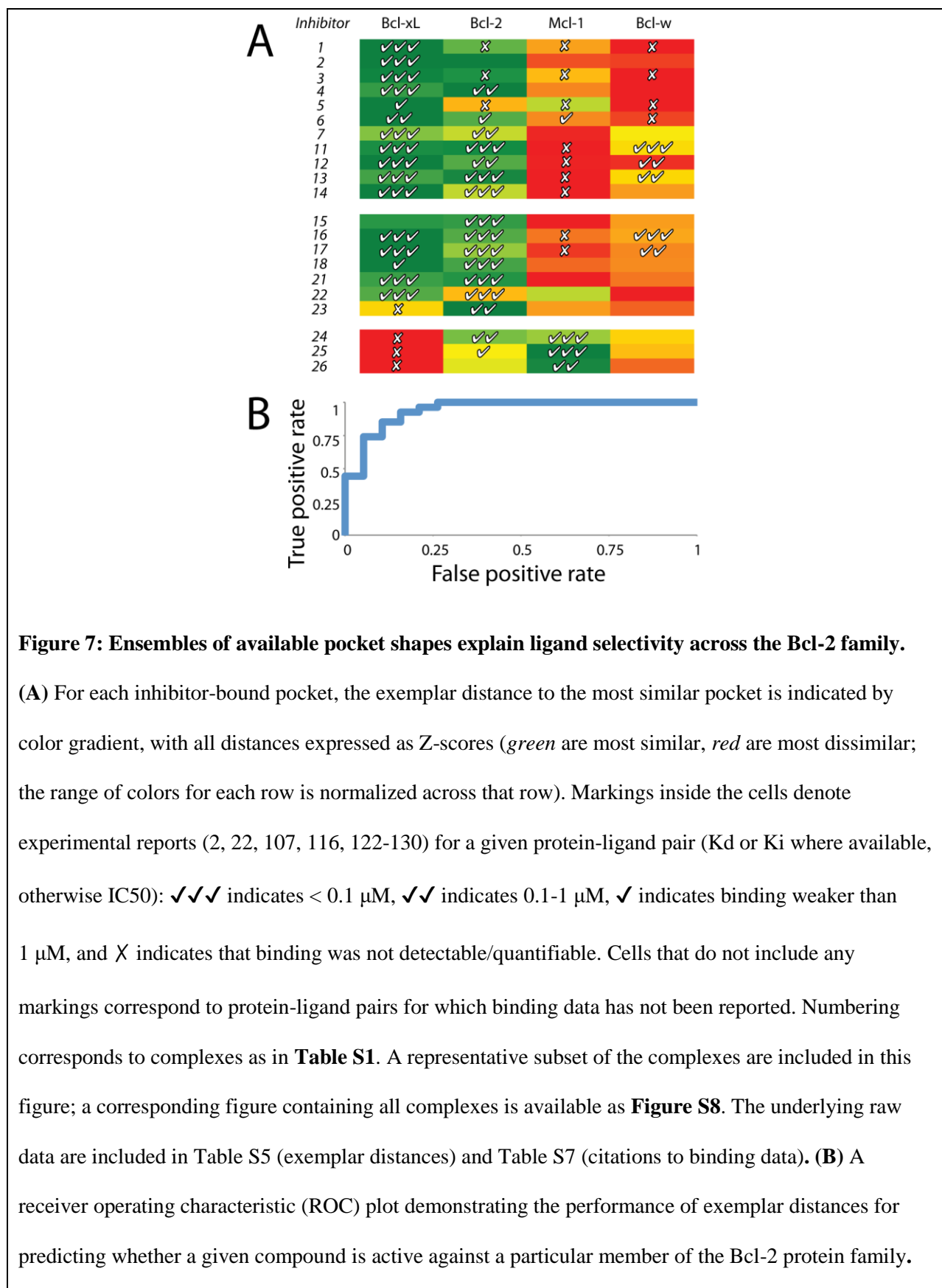
We observe a similar pattern for most inhibitor-bound crystal structures of Bcl-2 (*15-21*), which is again unsurprising given the similarity of these inhibitor-bound pockets to those of Bcl-xL (**Figure 2c**). The sole exceptions to this pattern are complexes *22* and *23*, for which corresponding pockets are observed in either the Bcl-2 or Bcl-xL ensembles but not both. The former (*22*) is indeed a dual inhibitor, but binds in a shallow pocket that is not well-described by any Bcl-2 exemplar in our ensemble. The latter (*23*) is indeed selective for Bcl-2 over Bcl-xL, as anticipated from the lack of overlap with the Bcl-xL ensemble from simulation, and therefore represents successful recapitulation of the binding data.

While we noted earlier that Mcl-1 can adopt conformations with highly distinct surface pockets, we also noted that inhibitor-bound crystal structures do not make use of these pockets (**Figure 6**). Rather, we find that these compounds (*24-26*) instead bind to pockets that are very similar to those included in the Bcl-2 ensemble (**Figure 7a**), and indeed experimental observations confirm that these compounds also inhibit Bcl-2, but not Bcl-xL.

Among the most notable incorrect predictions are those involving Bcl-w: a number of these compounds inhibit Bcl-w (*11, 12, 13, 16, 17*), but corresponding pocket shapes were not included in our sampling. In retrospect, this may have arisen due to structural features of the starting conformation from

which this conformational ensemble was generated: when evaluated by Molprobity (121), this member of the NMR ensemble contained only 72% of residues in favorable regions of Ramachandran space. These unfavorable structural features may in turn have led to a lack of convergence of our simulations.

Overall, however, there is a striking relationship between the pockets visited by a given ensemble and the experimentally-derived ligand selectivity. To quantitatively examine the ability of the pockets sampled in these ensembles to recapitulate the selectivity profile of a given ligand, we asked how well this approach could be used to distinguish the most tightly binding protein-ligand pairs (those marked with   ) from those pairs that bind too weakly to be detected/quantified (those marked with   ). For each inhibitor-bound crystal structure, we normalized the exemplar distances to the variation across the corresponding row; this essentially expressed ligand selectivity as a Z-score indicating how closely a given ensemble approached the inhibitor-bound pocket. Using these Z-scores to rank the likelihood of interaction for a given protein-ligand pair, we used a receiver operating characteristic (ROC) plot to show performance at this binary prediction problem (**Figure 7b**); we find that the predictions from this method far outperform those of a random classifier ($p < 4x10^{-7}$).

Figure 7: Ensembles of available pocket shapes explain ligand selectivity across the Bcl-2 family.

(A) For each inhibitor-bound pocket, the exemplar distance to the most similar pocket is indicated by color gradient, with all distances expressed as Z-scores (*green* are most similar, *red* are most dissimilar; the range of colors for each row is normalized across that row). Markings inside the cells denote experimental reports (2, 22, 107, 116, 122-130) for a given protein-ligand pair (Kd or Ki where available, otherwise IC50): ✓✓✓ indicates < 0.1 μM, ✓✓ indicates 0.1-1 μM, ✓ indicates binding weaker than 1 μM, and ✗ indicates that binding was not detectable/quantifiable. Cells that do not include any markings correspond to protein-ligand pairs for which binding data has not been reported. Numbering corresponds to complexes as in **Table S1**. A representative subset of the complexes are included in this figure; a corresponding figure containing all complexes is available as **Figure S8**. The underlying raw data are included in Table S5 (exemplar distances) and Table S7 (citations to binding data)**. (B)** A receiver operating characteristic (ROC) plot demonstrating the performance of exemplar distances for predicting whether a given compound is active against a particular member of the Bcl-2 protein family**.**
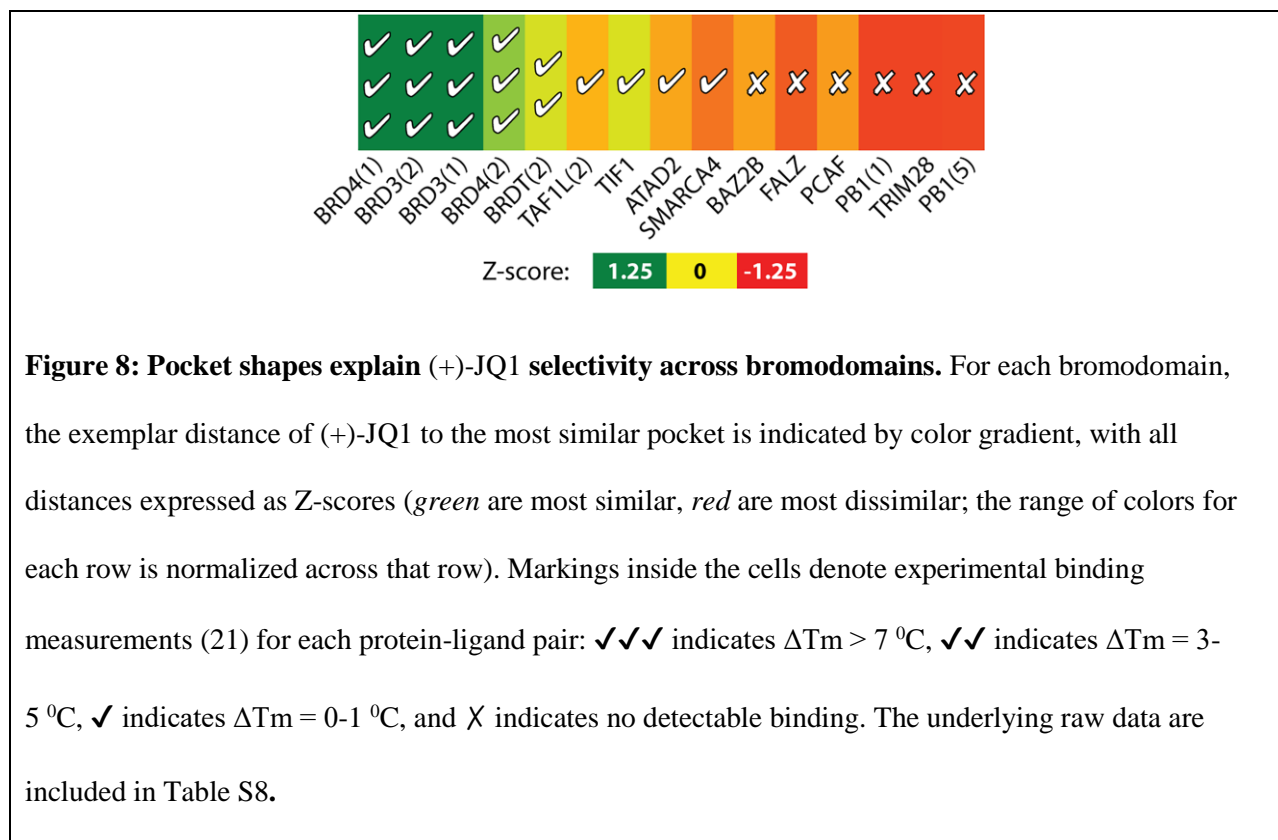
The absence of experimental binding data for many protein-ligand pairs is a natural shortcoming associated with culling this information from available reports in the literature. The complete maps of ligand selectivity for each compound, as inferred from ensembles generated by simulation, thus stand as completely new predictions in many cases (**Figure S8**).

Because the pockets should match the shape of the ligand, a reasonable assumption would be that similar results would be found by comparing the shape of the inhibitor to the ensemble of pocket shapes. To explore this idea we measured the exemplar distance between each native inhibitor conformer and the most similar pocket observed in a given ensemble of Bcl-2 family members, and then compared these exemplar distances to experimentally-derived binding data. As expected, the predictions (**Figure S9**) are very similar when comparing to the ligand directly, as opposed to comparing against the ligand-bound pocket.

Beyond the Bcl-2 family

The abundance of reported inhibitors of Bcl-2 family members, including their selectivity across the Bcl-2 family, enabled the detailed comparison presented above. Due to the challenges encountered to date in identifying small-molecule inhibitors of protein interactions, however, there do not yet exist any further examples that we know of in which multiple compounds target different members of a protein family. Nonetheless, we were able to identify a separate example of a small-molecule inhibitor of a protein interaction that has been shown to inhibit select members of a protein family (for which at least one ligand-bound crystal structure is available): this is (+)-JQ1, a compound shown to selectively inhibit a subset of human bromodomains (21). While only one compound, binding data are available against many family members. Here, we have generated ensembles of pocket-containing conformations for 16 bromodomains, and measured the extent to which each family member samples a pocket similar to that observed in the crystal structure of (+)-JQ1 bound to the first bromodomain of BRD4. In this single additional example, here again we find that the presence of complementary surface pockets in the

ensemble generated by simulation can accurately predict the ligand selectivity across a protein family (**Figure 8**), whereas we do not observe complementary pockets for family members that do not tightly bind (+)-JQ1.

Figure 8: Pocket shapes explain (+)-JQ1 **selectivity across bromodomains.** For each bromodomain, the exemplar distance of (+)-JQ1 to the most similar pocket is indicated by color gradient, with all distances expressed as Z-scores (*green* are most similar, *red* are most dissimilar; the range of colors for each row is normalized across that row). Markings inside the cells denote experimental binding measurements (21) for each protein-ligand pair: ✓✓✓ indicates $\Delta Tm > 7\ ^0C$, ✓✓ indicates $\Delta Tm = 3$-$5\ ^0C$, ✓ indicates $\Delta Tm = 0$-$1\ ^0C$, and ✗ indicates no detectable binding. The underlying raw data are included in Table S8**.**

2.6 Discussion

Identification of small-molecule inhibitors of protein interactions immediately raised the question of how these compounds might interact with proteins that appeared to lack complementary surface pockets; the answer came through structural studies of an interleukin-2 complex that showed the ligand can occupy a hydrophobic groove not present on the unbound protein surface (75). These structural studies, together with analysis of binding thermodynamics, first pointed to the "adaptivity" of this protein surface: the protein can adopt multiple conformations, one of which presents a surface complementary to

the ligand. Our analysis of Bcl-2 family complexes supports this view, and extends it further: we find that the plasticity of the protein surface allows multiple distinct surface pockets to be presented, and these different pockets can be recognized by inhibitors with dramatically different chemotypes.

Upon generating ensembles of pocket-containing conformations by simulation, we find that these ensembles span all the pockets used by known inhibitors – even though no information about any inhibitor was used to influence the simulations in any way. This observation provides strong and direct evidence for an underlying model of conformational selection (77): the protein surface is predisposed to adopt certain pocket shapes, and these shapes in turn restrict the range of complementary ligands. Upon binding, the protein may then undergo further smaller changes in response to particular features of the ligand.

By comparing the regions of "pocket space" explored by several Bcl-2 family members, we find that certain shapes are available to all family members: we expect that compounds complementing such pockets will show very broad specificity across this family. We further find that most inhibitors reported in the literature bind to pockets that are shared by more than one family member, and accordingly most are found to be active against more than one family member.

Conversely, we also find that many Bcl-2 family members sample pockets that are not accessible to any other family member: here lies a tremendous opportunity, since we expect that a compound built to complement such a pocket will prove highly selective for its target. Development of compounds that target these highly "distinct" pockets represents a tantalizing new strategy for drug discovery: by building target selectivity into the broad features of the chemical scaffold itself, selectivity may be more robustly preserved in the course of optimization of the compound for other orthogonal desirable properties (bioavailability, pharmacokinetics and pharmacology). Despite the existence of these pockets on the surface of Bcl-2 family members and extensive interest in identifying selective inhibitors, however, not a

single crystal structure reported to date includes a compound that targets any of these highly "distinct" pockets.

How then can we identify compounds that achieve target selectivity by explicitly targeting these "distinct" pockets? We anticipate the solution may lie with the exemplars themselves. As noted earlier, the exemplar is essentially a map of the "perfect" ligand to complement a given pocket, albeit a ligand that is not physically realizable. Accordingly, we expect that the exemplar will serve as an ideal template for ligand-based screening of (virtual) compound libraries; tools such as ROCS (3, 96) that evaluate volume and chemical overlap may be used to find compounds that closely mimic the shape and chemical features of the exemplar. Indeed, the ability to assess the selectivity profile by comparing the native ligand conformer to the exemplars derived from the ensembles implies that ROCS is capable of identifying compounds with the desired shape and chemical features needed to strongly interact with a member of a protein's ensemble. Together, this set of tools may provide both a means to identify pockets that "encode" a desired selectivity profile within a protein family, and also a means to connect the resulting pockets to specific compounds that exhibit this selectivity profile.

The overall paucity of examples of small-molecule inhibitors of protein interactions necessitated our focus for this study be largely restricted to the Bcl-2 family. As selective inhibitors of other protein families involved in protein interactions are reported, it will be exciting to refine the insights presented here. In light of the fact that many of the other small-molecule inhibitors of protein interactions described to date also bind to similarly "adaptable" binding sites (8), meanwhile, we are optimistic that the perspectives presented here will prove extensible to these therapeutic targets as well.

2.7 Acknowledgements

CHAPTER 3: Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein-Protein Interactions

David K. Johnson[1] and John Karanicolas[1,2*]

[1] Center for Computational Biology, [2] Department of Molecular Biosciences, and

University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534

*To whom correspondence should be addressed. E-mail: johnk@ku.edu, 785-864-8298

3.1 Abstract

Protein-protein interactions play important roles in virtually all cellular processes, making them enticing targets for modulation by small-molecule therapeutics: specific examples have been well validated in diseases ranging from cancer and autoimmune disorders, to bacterial and viral infections. Despite several notable successes, however, overall these remain a very challenging target class. Protein interaction sites are especially challenging for computational approaches, because the target protein surface often undergoes a conformational change to enable ligand binding: this confounds traditional approaches for virtual screening. Through previous studies, we demonstrated that biased "pocket optimization" simulations could be used to build collections of low-energy pocket-containing conformations, starting from an unbound protein structure. Here, we demonstrate that these pockets can further be used to identify ligands that complement the protein surface. To do so, we first build from a given pocket its "exemplar": a perfect, but non-physical, pseudo-ligand that would optimally match the shape and chemical features of the pocket. In our previous studies, we used these exemplars to quantitatively compare protein surface pockets to one another. Here, we now introduce this exemplar as a template for pharmacophore-based screening of chemical libraries. Through a series of benchmark experiments, we demonstrate that this approach exhibits comparable performance as traditional docking methods for identifying known inhibitors acting at protein interaction sites. However, because this approach is predicated on ligand/exemplar overlays, and thus does not require explicit calculation of protein-ligand interactions, exemplar screening provides a tremendous speed advantage over docking: 6 million compounds can be screened in about 15 minutes on a single 16-core, dual-GPU computer. The extreme speed at which large compound libraries can be traversed easily enables screening against a "pocket-optimized" ensemble of protein conformations, which in turn facilitates identification of more diverse classes of active compounds for a given protein target.

3.2 Introduction

The concept of a pharmacophore dates back at least a century: it is traditionally attributed to Paul Ehrlich, who recognized that certain parts of molecules were responsible for their biological activity (131). This concept was modernized fifty years later, shifting away from chemical groups and towards a more abstract notion of chemical forces in three-dimensional space (132). The IUPAC now defines a pharmacophore as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" (24).

Pharmacophores enable design of small molecules capable of presenting specific functional moieties to elicit a desired biological response, and for decades they have been used to inspire medicinal chemists' development of new analogues (25-27). Because they describe the spatial arrangement of critical interactions with a receptor, pharmacophores can also be used as templates for computational screens seeking to identify ligands containing functional groups positioned to recapitulate these interactions.

The first computed example of a modern pharmacophore is attributed to Lemont Kier, who recognized the spatial similarity of (modeled) three-dimensional geometries of various muscarinic receptor agonists (133). Presently, a broad assortment of computational tools can be used to define pharmacophores in distinct ways (28, 29, 91-93, 134-137). The first pharmacophore-building algorithms drew information from the ligand alone: such approaches begin by finding a consensus structural alignment of multiple active compounds, then seek to identify shared functional groups in this set (28). More recently, development of tools such as LigandScout (29) allow key interactions to instead be defined from one or more crystal structures of a receptor with assorted ligands bound – here again, identifying features shared by multiple ligands to build a consensus pharmacophore.

More recent efforts have focused on building pharmacophore models from protein structures alone, solved without any bound ligand in the active site. These approaches typically begin by docking an assortment of small (chemically diverse) probe molecules into the active site, then evaluating the interactions with the protein that these probes make (91-93). Individual interactions presented by different probe molecules are then combined into a consensus pharmacophore, and used as a template to identify larger compounds that simultaneously recapitulate the interactions from multiple probes. As an alternative, other approaches instead define desirable three-dimensional properties of candidate ligands using the "negative image" of the binding pocket (135, 136).

Pharmacophores have been applied extensively to many diverse targets, including enzymes (138-141), G protein-coupled receptors (142-144), and transporters (145-147). In each of these cases, the protein target has evolved to bind some natural small-molecule partner: already this suggests that the chemical space of prospective hits may be similar to that of the natural binding partner(s) (148). In most such cases one or more of the natural ligand(s) are known, so the task that remains entails identifying alternate molecules that recapitulate the key interactions of these natural ligand(s).

In some cases, however, an important biological target is not evolved to bind *any* natural small-molecule ligand: these include protein-protein interactions, protein-RNA interactions, and others. Moreover, representatives from this target class include well-validated targets for cancer, viral and bacterial interactions, and autoimmune disorders (1, 2, 4-7, 149). Given the lack of a natural small-molecule binding partner, it is not evident *a priori* in which regions of chemical space prospective inhibitors may reside – or even whether the protein surface is indeed druggable (or rather, "ligandable" (150)) with *any* small molecule at all (151). Until one or more small-molecule inhibitors have been identified, there are (of course) no active compounds to use as templates for building a pharmacophore model. As a further additional challenge, crystal structures of inhibitors bound to protein-protein inhibitors have shown that in some cases the binding pockets on the protein surface are transient, appearing neither in the unbound structure nor in that of the protein-bound complex (8).

Despite these challenges, pharmacophore models have still successfully been applied to protein-protein interactions, in certain cases. Most notably, several groups have recently developed related computational approaches that select some of the side chains from the natural protein partner (the "hotspot" or "anchor" residues), and then use this collection of functional groups to build a pharmacophore (30-37). However, this approach comes with several important limitations. First, the inhibitors designed to match such pharmacophores may be limited in the chemical space they sample, given that they are built from mimicry of canonical amino acid side chains. Second, their binding affinity may also be limited, given that they make use of only a subset of the natural binding partner's interactions. Finally, this approach cannot take advantage of conformational changes at the protein surface that might allow binding of small molecules that do not resemble the natural binding partner; as noted earlier, such conformational changes are frequently observed in crystal structures of inhibitors bound at protein interaction sites (8).

In our previous work, we developed a biasing potential that can be used to create ensembles of low-energy pocket-containing conformations, starting from an unbound protein structure that lacks a surface pocket suitable for small-molecule binding (76). Others have shown that docking-based virtual screens can be improved by docking to an ensemble of receptor conformations, rather than to a single static protein structure (152-160). Thus, we surmised that the inclusion of many surface pockets – derived from "pocket optimization" biased simulations (76) – would facilitate identification of inhibitors beyond those that simply mimic the natural binding partner.

In order to quantitatively compare the shape and chemical features between pairs of surface pockets, we then developed the concept of pocket "exemplars" and used them to compare pockets to other pockets in order to predict the selectivity of active compounds between related protein family members (161). Briefly, an exemplar is built from a protein structure by determining the locations of hydrogen bond donors/acceptors within the pocket that would form idealized interactions with the receptor, then filling the remainder of the pocket with a series of spheres that simply occupy volume. In essence, an

exemplar is the negative image of the protein pocket: it represents the "ideal" ligand that would complement a given pocket, albeit in the absence of physical chemical constraints (i.e. the bond connectivity is neither specified nor implied).

By representing pockets through their "idealized" complementary ligands, we found that we could make accurate comparisons of pocket similarity (161). Here, we extend this approach further: instead of simply comparing exemplars to other exemplars, as done in previous work, we demonstrate that these exemplars can also be used directly as templates for very rapid pharmacophore-based virtual screening. The speed also enables us to screen against *ensembles* of conformations from the "pocket optimized" simulations, which had previously only been used to assess druggability and selectivity. Because exemplars are built directly from the protein structure, this approach is particularly well suited for regimes in which no natural binding partner for the protein target is available, or, as in the case of protein-protein interactions, the natural binding partner does not provide a suitable template from which to build a conventional pharmacophore.

3.3 Computational approach

Conformational sampling of the protein and exemplar-generation are implemented in the Rosetta software suite (56); Rosetta is freely available for academic use (www.rosettacommons.org). Virtual screening using the exemplar was carried out using the FastROCS (162, 163) and ROCS (163, 164) software. The relevant command lines used to run both Rosetta and FastROCS/ROCS are included as *Supplementary Methods*.

In addition to the command-line program included in Rosetta, we have provided a tool to generate an exemplar from an input PDB file through the ROSIE web interface (165). This tool is available at: http://rosie.rosettacommons.org/make_exemplar .

Our approach for screening a chemical library using exemplars is summarized schematically in **Figure 1**; each step is described in detail below.
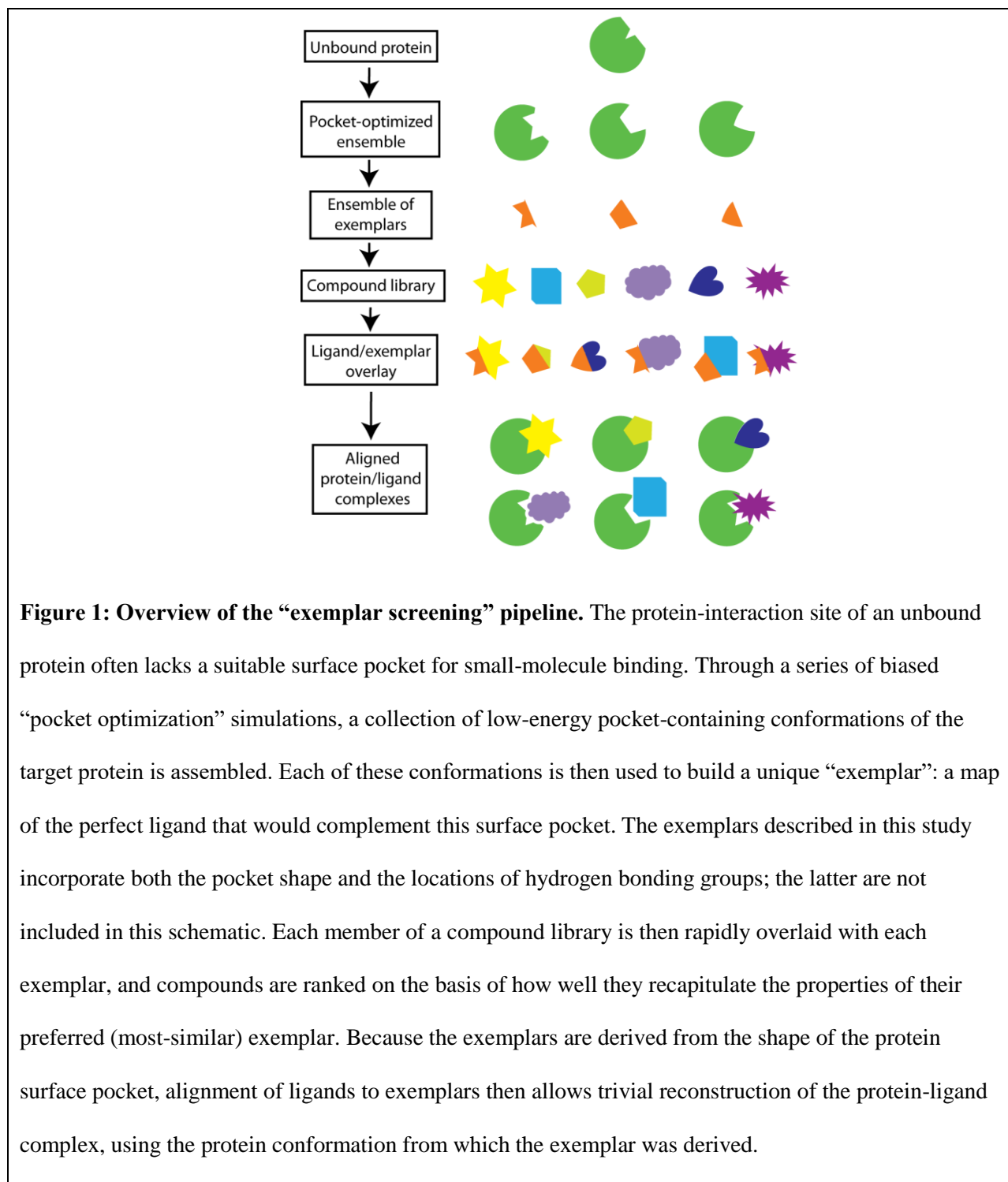

Figure 1: Overview of the "exemplar screening" pipeline. The protein-interaction site of an unbound protein often lacks a suitable surface pocket for small-molecule binding. Through a series of biased "pocket optimization" simulations, a collection of low-energy pocket-containing conformations of the target protein is assembled. Each of these conformations is then used to build a unique "exemplar": a map of the perfect ligand that would complement this surface pocket. The exemplars described in this study incorporate both the pocket shape and the locations of hydrogen bonding groups; the latter are not included in this schematic. Each member of a compound library is then rapidly overlaid with each exemplar, and compounds are ranked on the basis of how well they recapitulate the properties of their preferred (most-similar) exemplar. Because the exemplars are derived from the shape of the protein surface pocket, alignment of ligands to exemplars then allows trivial reconstruction of the protein-ligand complex, using the protein conformation from which the exemplar was derived.

Using pocket optimization to build ensembles of pocket-containing conformations

To generate ensembles of pocket containing conformations, we use a Monte Carlo sampling technique implemented in the Rosetta software suite (56); this local move set was originally designed for refining templates in comparative modeling applications (166), and includes both sidechain and backbone moves among its degrees of freedom. In order to enrich our sampling with low-energy conformations that include surface pockets, we include a biasing term in the energy function that favors conformations in which a surface pocket is present (76). We have previously shown that this approach creates ensembles of pocket-containing conformations with energies that overlap those from the corresponding unbiased simulations, implying low energetic cost to open the particular pockets we observe (76, 161). For each protein of interest we carried out biased simulations, collected the final output structure from each of 1,000 independent trajectories, and retained only those with energies comparable to those observed in analogous unbiased simulations (76, 161).

In order to focus the resulting pockets on the intended surface (the protein interaction site, in this case), the energetic "bonus" associated with pocket formation is awarded only if the pocket is in direct contact with one or two pre-determined "target" residues. For this study we used an automated tool to determine the target residues (161), provided that a crystal structure of the protein-protein complex was available. In one case (Brd4) the crystal structure of the protein-protein complex included non-canonical amino acids at the protein interaction site, so for this protein we chose the target residues manually.

Generating exemplars

Starting from the conformation generated by the "pocket optimization" simulation, we defined the surface pocket as described previously (76). Briefly, the protein and its inflated van der Waals surface were mapped to a local grid centered on the target residues, and the pocket was identified by searching for linear segments of solvent bounded by the protein. The "deep" pocket volume was defined by removing the most solvent-exposed portion of the pocket.

We then generated an exemplar from each pocket, as described previously (161). Briefly, we used the protein conformation to identify the location of idealized hydrogen bond donors and acceptors that overlap with any of the pocket, but do not clash with the protein. The remainder of the deep pocket volume was then filled in using carbon atoms, placing these such that they were completely contained in the pocket with minimal overlap, and requiring no less than 1.5 Å between the centers of atom pairs.

For virtual screening, we sought to use several representative exemplars from the ensemble of "pocket-optimized" output structures. To select these representatives, we used Tanimoto scores computed with the ROCS software (163, 167) as a measure of exemplar similarity (161), and computed the pairwise similarity of all the exemplars. We then grouped together similar exemplars using hierarchical clustering, and excluded from consideration any clusters with less than 10 members (in cases for which there were fewer than 200 exemplars being clustered, we relaxed this criterion to require only 5 members). From each cluster, we advanced two representative exemplars: the cluster centroid, and the exemplar from the lowest-energy conformation in the cluster.

Building small-molecule libraries for screening

In the benchmark experiments described here, we compiled collections of active and decoy compounds (as described in the *Results* section). We then used the OMEGA software (168, 169) (with default parameters) to generate up to 100 low-energy three-dimensional conformations for each compound. The active compounds were identified using the TIMBAL database (170, 171) when available and the CHEMBL database (172) when the target was not present in TIMBAL. In the cases where the targets were not present in TIMBAL nor CHEMBL, the identity of the crystallographic ligand was used (but not the crystallographic conformation).

Active compounds were retained only if they had K_i or K_d less than 10 μM, and molecular weight between 200 and 800 Da. In the cases where these criteria were met by multiple members of a chemical series (2D Fingerprint Tanimoto similarity > 0.7), only the most potent compound was retained. In order

to make use of compounds for which only IC50 values were reported (instead of K_i or K_d), these values were scaled by a factor of 0.5 so that they could be compared on equal footing with K_i or K_d values (173).

Exemplar screening

Starting from an exemplar, FastROCS (162, 163) was used to rapidly overlay each compound in the library onto the exemplar. To generate full receiver-operating characteristic curves, the top-scoring conformer overlay of each compound was then rescored (without optimization) using ROCS (163, 164). For the large-scale virtual screening application presented later, only the top 0.75% scoring compounds for a particular exemplar were rescored using ROCS: in a virtual screen, often one re-ranks only the top-scoring compounds from the initial ranking.

Using ROCS, we computed the Tversky score for each overlay: this asymmetric comparison evaluates the extent to which the ligand subsumes/matches the exemplar features (shape and hydrogen bond donors/acceptors) *without* penalizing the ligand for including "extra" features (for example, if the ligand volume extends outside that of the exemplar, or if the ligand includes additional hydrogen bond donors/acceptors not matched to any in the exemplar) (174). We note that Tversky scores depend on properties of the exemplar; for this reason, one cannot simply compare raw Tversky scores for ligands overlaid with different exemplars. When screening with multiple exemplars, therefore, we expressed each ligand overlay as a Z-score for the exemplar used (i.e. normalized to the mean and standard deviation of the scores from all compounds overlaid to this exemplar). Each compound was assigned a final score corresponding to the best Z-score drawn from any of the exemplars in the set.

3.4 Results

Exemplars map the binding pocket of known ligands

In this study, we use an "exemplar" to describe the shape and chemical composition of a protein surface pocket. The exemplar is a collection of atoms that captures the ideal shape and chemical properties that a ligand would exhibit if it were to perfectly complement the pocket: thus, we expect that exemplars may naturally serve as templates for pharmacophore screening.

The exemplars we will use here employ the same three atom types defined in our original conception (161): an atom type to define locations at which hydrogen bond donors should be placed, an atom type to define locations at which hydrogen bond acceptors should be placed, and a non-polar atom type used to fill the remainder of the pocket volume. By construction, then, the exemplar maps out the shape of the pocket and also indicates the key locations for moieties that will participate in hydrogen bonding interactions.

As a starting point, we used the crystal structure of Bcl-xL bound to ABT-737, a compound that inhibits binding of BH3 peptides at this site. In this protein-inhibitor complex, ABT-737 occupies a long, deep channel along the protein surface (**Figure 2a**). We removed the ligand from this structure, and used the protein structure to generate its corresponding exemplar; indeed, we find that the corresponding exemplar has similar size and shape as the inhibitor itself (**Figure 2b**). We reiterate that the exemplar is built from the protein structure alone, without direct knowledge of the ligand: here, the size and shape of the exemplar matches that of the ligand, because this ligand is a natural match for the size and shape of the protein surface pocket. We note that the exemplar includes both a hydrogen bond acceptor and a hydrogen bond donor in close proximity, in the region occupied by the inhibitor's sulfonamide group; this arises from the fact that Bcl-xL includes a tyrosine hydroxyl group at this location on the protein surface, allowing complementary ligands to present either a donor or an acceptor.

To examine whether this exemplar indeed unambiguously encodes the location and orientation of this inhibitor, we built a series of 200 low-energy conformations of ABT-737, and "hid" the crystallographic conformation among these. We then overlaid each conformation with our exemplar, and evaluated the extent to which the ligand filled the exemplar volume and matched the locations of hydrogen bonding groups (see *Computational Approach*). The top-scoring model from this "pose prediction" experiment makes use of the crystallographic ligand conformation, and places the ligand in essentially the same position and orientation as in the crystal structure (**Figure 2c**).

We turned next to the crystal structure of XIAP bound to a bicyclic peptidomimetic compound that disrupts binding of the Smac peptide (**Figure 2d**). We find that the exemplar built from this protein structure is much smaller (**Figure 2e**), consistent with the fact that the XIAP inhibitor is much smaller than ABT-737, and binds at a much shallower surface pocket. We further note that the exemplar is expected only to capture features that interact with the protein: thus, the exemplar excludes the solvent-facing aromatic ring and carbonyl group found in this inhibitor.

Once again we built a series of 200 low-energy conformations for this inhibitor, and included among these the crystallographic conformation. Upon overlaying each of these with the exemplar derived from the XIAP, we find that the top-scoring conformation does not precisely match the crystal structure (**Figure 2f**). The left side of the molecule includes two hydrogen bonds to the protein; the exemplar successfully includes these features, and indeed this part of the molecule is faithfully recapitulated in the predicted pose. The right side of the molecule includes a pair of aromatic rings, and one of these faces solvent. Here, the exemplar includes a secondary pocket (at the top left) that is not filled by the inhibitor. Because of this, the top-scoring pose instead uses an alternate ligand conformation that allows this secondary pocket to be accessed. The lack of agreement between the crystallographic pose and the predicted pose, then, derives not from how the overlaid structures are scored, but rather from the fact that the exemplar included features of the protein surface that are not used in binding this particular ligand.

Encouraged by these results, we next proceeded to explore the utility of this exemplar-based method for virtual screening.
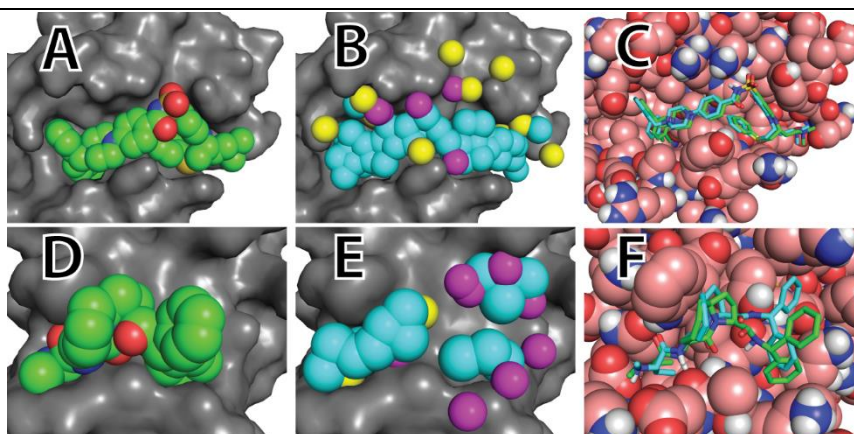

Figure 2: Defining exemplars. (A) Bcl-xL (*grey surface*) is shown in complex with small-molecule inhibitor ABT-737 (*spheres*). **(B)** This conformation of Bcl-xL was used to construct an exemplar comprised of hydrogen bond donors (*yellow*), hydrogen bond acceptors (*magenta*), and non-polar atoms that fill the remainder of the pocket volume (*cyan*). The exemplar is in essence a map of the ligand that would perfectly complement the pocket. **(C)** Upon alignment to the exemplar, the predicted pose of ABT-737 (*cyan*) very closely matches the pose observed in the crystal structure (*green*) (ligand RMSD 0.93 Å). **(D)** XIAP (*grey surface*) is shown in complex with a small-molecule inhibitor (*spheres*). **(E)** This conformation of XIAP was used to construct an exemplar. **(F)** Upon alignment to the exemplar, the predicted pose of this inhibitor (*cyan*) recapitulates many interactions from the pose observed in the crystal structure (*green*), including specific hydrogen bonds (ligand RMSD 2.61 Å).

Virtual screening using a ligand-bound protein structure

The goal of a virtual screen is to distinguish the active compounds in a chemical library (true positives) from among the inactive compounds in the library. Previously we compiled a virtual screening benchmark for small-molecule inhibitors of Bcl-xL and XIAP (175); the benchmark evaluates the ability of various methods to distinguish known inhibitors of these protein interactions from "decoy" compounds

that are presumed to be inactive. Compounds active against either of these proteins were drawn from the

TIMBAL database (170, 171). To prevent bias from inclusion of many members of a chemical series, we

removed any compound with 2D Fingerprint Tanimoto similarity (176) closer than 0.70 to any other

compound in the dataset. This led to a total of 27 compounds active against Bcl-xL, and 14 compounds

active against XIAP. To generate physicochemically-matched decoy compounds, we drew all compounds

included in the TIMBAL database described as inhibitors of *other* protein-protein interactions. We

removed all decoy compounds with 2D Fingerprint Tanimoto similarity closer than 0.70 to any active

compound, or to any other decoy compound. This led to a total of 328 decoy compounds for Bcl-xL, and

425 decoy compounds for XIAP. While the decoy compounds are not explicitly known to be inactive

against these protein targets, we rely on their lack of similarity to known targets to minimize the

likelihood of inadvertently including in the decoy set any compounds that are active.

We started from the crystal structures of Bcl-xL and XIAP each bound to a known inhibitor; to

avoid biasing our benchmark, we excluded this particular inhibitor from among the active compounds in

our library. We then used the exemplars generated from these two protein structures (**Figure 2**) to overlay

and rank each compound in the corresponding compound library: the compounds that most closely match

the exemplar are inferred to be the active compounds in the set. We also used three other docking tools to

rank these compounds: DARC (175), DOCK 6 (177, 178), and AutoDock4 (179). Parameters used to run

each of these three programs are described fully in our previous work (175). Finally, in addition to these

docking methods we used a ligand-based method, ROCS (149, 163, 167), to evaluate the three-

dimensional similarity of each compound to the conformation of the known inhibitor in the crystal

structure.

From the rank order of each compound, we generated Receiver Operator Characteristic (ROC)

curves for each of these methods, for both Bcl-xL and for XIAP (**Figure 3**). For a given method, each

point on the plot corresponds to a different ranking at which compounds are be considered active/inactive

(i.e. what fraction of the library will be classified as active). For a given cutoff, the y-axis defines the

percentage of active compounds that have been (correctly) classified as active (true positives) and the x-axis defines the percentage of inactive compounds that have been (incorrectly) classified as active (false positives). For a purely random classifier, the percentage of both active and inactive compounds will rise at an equal rate (*grey curve*).

We found that all five of these methods outperformed the random classifier, in both the Bcl-xL benchmark and the XIAP benchmark. The performance of ROCS in the XIAP benchmark was particularly impressive: we expect that this arises because many of the active compounds in the XIAP set are peptidomimetics, and thus can be recognized by virtue of the pattern of hydrogen bonds that is shared with the template ligand. Meanwhile, the performance of ROCS when applied to Bcl-xL is not quite as impressive, since this set of inhibitors uses a more diverse range of interactions.

Overall, we find that the exemplar screening method performed at a similar level in these benchmarks as the three other receptor-based methods to which we compared it.
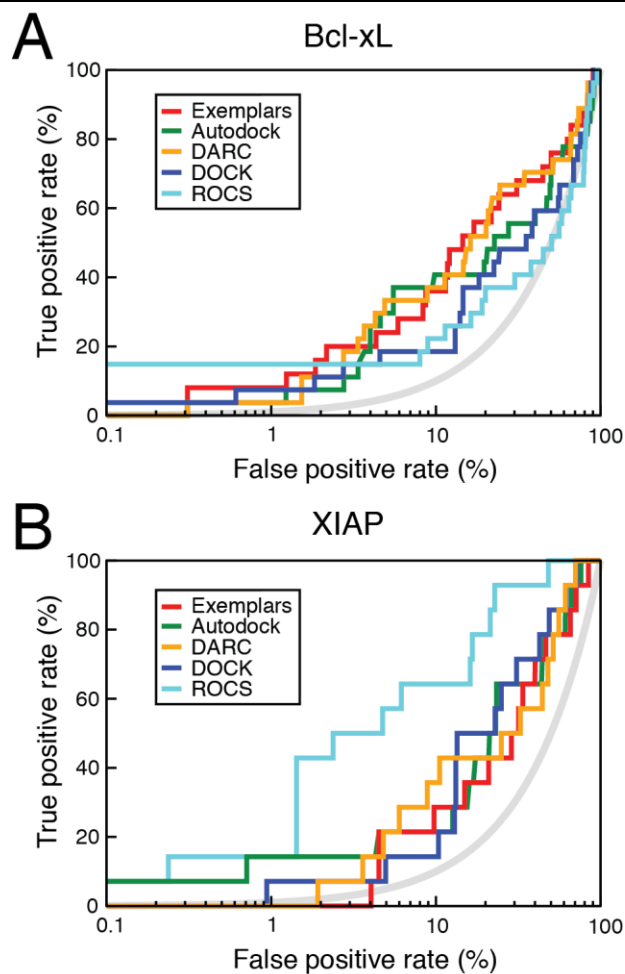
Figure 3: Exemplar screening using a ligand-bound protein structure. This receiver operating characteristic (ROC) plot compares the performance of exemplar screening to three structure-based docking methods (DARC, DOCK 6, and AutoDock4), along with a ligand-based method (ROCS). Data are presented on a semi-log plot, to emphasize "early" performance of each method; the *grey curve* indicates performance of a random predictor. This benchmark experiment involves distinguishing **(A)** 27 compounds active against Bcl-xL from among 328 decoy compounds, and **(B)** 14 compounds active against XIAP from among 425 decoy compounds. In both cases, decoy compounds were drawn from the TIMBAL database (170, 171).

Virtual screening using a "pocket-optimized" unbound structure

Though one was provided for each of the previous benchmarks, at the outset of many real virtual screening scenarios there is no available inhibitor-bound crystal structure from which to draw the protein conformation. While screening against an unbound conformation leads to diminished performance for most target classes (180), this is a particularly acute problem when seeking inhibitors of protein-protein interactions since these protein surfaces are particularly likely to require conformational changes (relative to the unbound protein) in order to bind the inhibitor (8). As noted earlier, we have developed a "pocket optimization" method that – starting from the unbound protein structure – uses biased sampling to efficiently explore low-energy conformations that contain surface pockets suitable for small molecule binding (76, 161). We therefore sought to extend this virtual screening benchmark to a regime in which no inhibitor-bound protein conformation is available, and one must instead screen using a "pocket-optimized" conformation of the target protein.

From "pocket optimization" simulations of Bcl-xL and XIAP, we therefore extracted the lowest-energy pocket-containing conformations (832 conformations for Bcl-xL and 354 conformations for XIAP), and used these to carry out the same screening experiment described earlier. We omitted ROCS from this comparison: given that we approach this experiment with no presumed knowledge of any other inhibitors, one would not have a template for use with a ligand-based screening method. Among the four receptor-based methods, we find that both exemplar screening and DARC exhibit similar performance as in the previous experiment (**Figure 4**). In contrast, both DOCK 6 and AutoDock4 show dramatically diminished performance in this experiment.
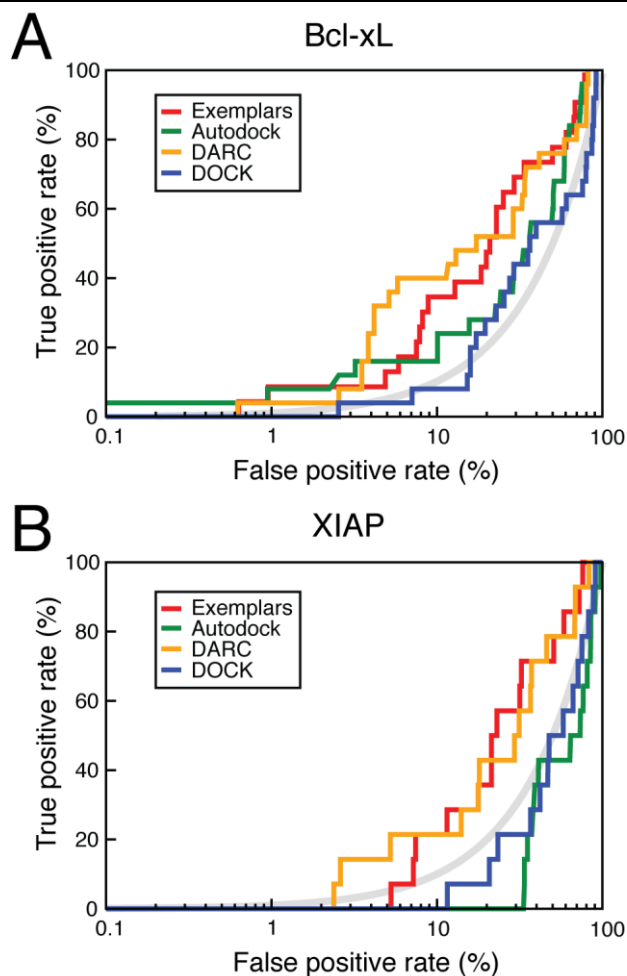
Figure 4: Exemplar screening using a "pocket-optimized" unbound structure. This benchmark experiment entails discrimination of the same active versus "decoy" compounds as in **Figure 3**. This time, however, the protein conformation used for screening was taken from "pocket optimization" simulations. Data are again presented on a semi-log plot, to emphasize "early" performance of each method; the *grey curve* indicates performance of a random predictor. This benchmark experiment is again applied to **(A)** Bcl-xL and **(B)** XIAP.

Previously, we found that "pocket-optimized" conformations often include small structural differences relative to the corresponding ligand-bound conformations, but nonetheless often capture features of the binding pocket with high fidelity (161): this, in turn, may explain the performance observed in this benchmark. At high resolution, the pocket-optimized conformations do not fit the active

compounds quite as well, because the detailed positioning of certain protein groups is not optimal since the simulations were performed without knowledge of any particular inhibitor: this presents a problem for methods such as DOCK 6 and AutoDock4, which evaluate the complementarity between individual intermolecular contacts. In contrast, the faithful recapitulation of the crude features of the binding pocket allows methods that focus instead on identifying ligands that match overall shape and chemical complementary of the pocket itself (DARC and exemplar screening) to perform just as well in this regime.

Virtual screening using an ensemble of "pocket-optimized" unbound structures

An important advantage of the exemplar screening method presented here is that the underlying ligand/exemplar overlays (using ROCS) prove much faster than any of the explicit docking approaches included in our study. In fact, OpenEye recently developed a newer implementation ("FastROCS") that takes advantage of multiple CPU and GPU cores, and is even faster than ROCS (163, 181). Using this new implementation, we find that screening 6 million compounds against a single exemplar, depending on the number of hit compounds written to disk, can be completed in as few as 15 minutes on a single 16-core, dual-GPU computer.

Given the speed of this approach, it now becomes feasible to screen a large compound library against *every* exemplar derived from an ensemble of pocket-optimized protein conformations, not just the exemplar from the lowest-energy conformation. Nonetheless, for computational efficiency we instead elected to identify a representative subset of exemplars for screening. To this end we used the same overlay approach (in ROCS) to evaluate the pairwise similarity of the exemplars, and then used this measure as the basis for hierarchical clustering. For Bcl-xL, we used 10 distinct exemplar clusters, and for XIAP we used 6 distinct exemplar clusters. From each cluster we then extracted the exemplar from the lowest-energy conformation, and the exemplar corresponding to the cluster centroid. This led to a collection of 15 representative exemplars for Bcl-xL, and 9 representative exemplars for XIAP.
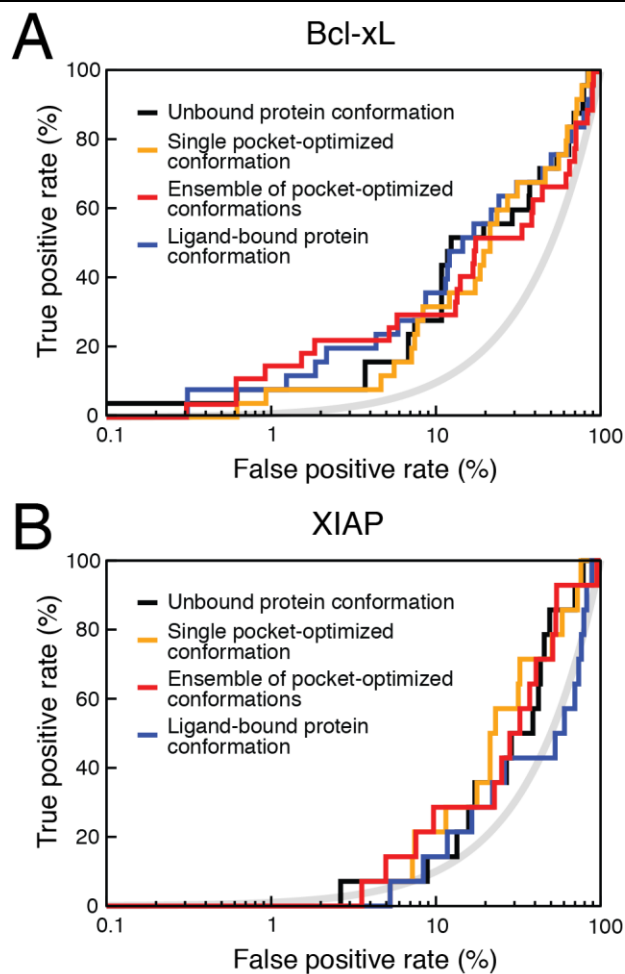
Figure 5: Exemplar screening using an *ensemble* of "pocket-optimized" unbound structures.
Screens were carried out using the same benchmark sets as **Figures 3** and **4**, for both **(A)** Bcl-xL and **(B)** XIAP. Here we compare performance when screening using exemplars generated from the unbound protein structure, from a single pocket-optimized conformation, from an ensemble of pocket-optimized conformations, or from an inhibitor-bound protein structure. The *grey curve* indicates performance of a random predictor.

We then used each of these exemplars to once again screen our benchmark libraries for Bcl-xL and XIAP. For a given exemplar, we computed the mean and standard deviation of the scores from all overlaid ligands, and used this to express the ligand/exemplar overlap as a Z-score. We then searched over all the exemplars to find the best Z-score for each compound, and used these as the basis for ranking

the compounds. Overall, we find that this "ensemble" exemplar approach yields similar performance in this experiment as using an exemplar from the unbound structure, or using an exemplar from the single lowest-energy pocket-optimized protein conformation (**Figure 5**).

Ultra-high-throughput virtual screening using an ensemble of "pocket-optimized" unbound structures

As noted above, a key advantage of the exemplar screening method is its speed. This is important, because it allows one to carry out screens using multiple receptor conformations, and/or screen against very large compound libraries. As a demonstration of this capability, we compiled from the ZINC database (182) a collection of ~8.33 million drug-like compounds to use as "decoys" for a large screening experiment. Once again, we hid among these decoy compounds either the 24 compounds active against Bcl-xL, or the 14 compounds active against XIAP.

We screened the compound library using each of the 15 representative exemplars from pocket-optimized conformations of Bcl-xL, or the 9 representative exemplars from pocket-optimized conformations of XIAP. The scale of this screening experiment would make it extremely time-consuming using any of the other three receptor-based (docking) methods included in our previous benchmark experiments. Using exemplar screening, however, screening more than 8 million compounds against each of these two protein targets – with multiple receptor conformations for each – required a wall time (total execution time) of only 7 hours on a single (dual GPU) computer.

Results from this large-scale screen are compiled into **Figure 6**. In both cases, exemplar docking against an ensemble of protein conformations identifies active compounds far more effectively than a random predictor (**Figure 6a**), with the area under the curve of each ROC plot (AUC) at 0.793 and 0.848 for Bcl-xL and XIAP, respectively. Screening against the Bcl-xL ensemble identified 50% of the (chemically distinct) active compounds within the top 1% of the library, while screening against the XIAP ensemble identified 29% of the active compounds within the top 1% of the library. The early enrichment – a key measure for virtual screening because only a small fraction of the total library is expected to

79

actually be tested experimentally – is also impressive: in both cases the top-ranking 0.01% of the library has more than 750-fold enrichment of active compounds relative to the complete library (**Figure 6b**). This exciting result suggests that – with very modest computational expense – much of this large screening database can be eliminated from further consideration, leaving a much smaller library that is considerably enriched with active compounds. Notably, performance in this screen is worse when the exemplar from the single lowest-energy pocket-containing conformation is utilized in this screen, instead of exemplars from a small ensemble of low-energy conformations (**Figure 6a**): this observation highlights the benefit of screening against multiple receptor conformations, which in turn is enabled by the very fast exemplar screening approach.

Examination of the top-ranking complexes involving compounds known to be active reveal compelling binding modes for these compounds (**Figure 6c**), the structures of which have not (to date) been solved in complex with their protein partners. Meanwhile, the top-ranking complexes involving decoy compounds also exhibit compelling binding modes (**Figure 6d**); as expected from their match to the underlying exemplars, each of these combine exquisite shape complementarity with multiple intermolecular hydrogen bonds. It is useful to recall that the decoy compounds in this experiment are merely presumed inactive, and have not been explicitly tested against Bcl-xL or XIAP: it is quite possible that the top-ranking compounds from this experiment may indeed exhibit activity their respective protein targets.
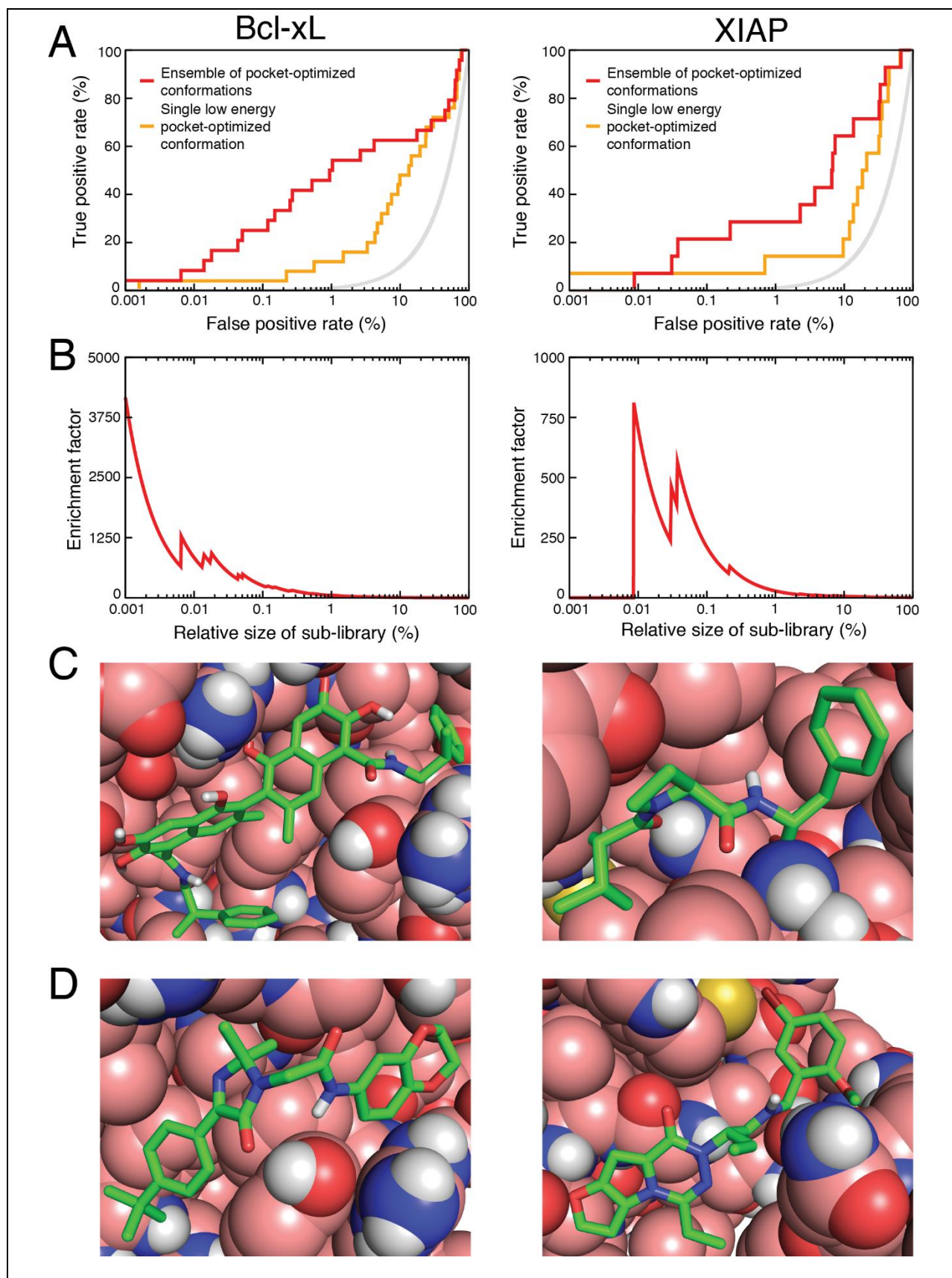
Figure 6: Ultra-high-throughput screening using an ensemble of "pocket-optimized" unbound structures. The same 27 compounds active against Bcl-xL and 14 compounds active against XIAP were now hidden among a library of 8.33 million drug-like compounds. This collection was screened using exemplars generated from an ensemble of pocket-optimized conformations. (**A**) This receiver operating characteristic (ROC) plot compares the performance of exemplar screening when using an ensemble of "pocket optimized" conformations (*red*) or the single lowest energy, "pocket optimized" conformation (*orange*). The *grey curve* indicates performance of a random predictor. The BEDROC metric with α=20 (183), a key measure designed specifically to assess performance for "early recognition", is 0.559 for Bcl-xL when screening against an ensemble of conformations, but drops to 0.249 when screening against the single lowest-energy conformation. Similarly for XIAP, the BEDROC value drops from 0.419 to 0.058 when screening against a single conformation instead of an ensemble. (**B**) The enrichment factor indicates the extent to which active compounds are over-represented in "sub-libraries" generated by extracting the top-scoring compounds from the original compound library, when screening against an ensemble of "pocket optimized" conformations (*red*). Because this value corresponds to enrichment over the fraction of active compounds in the original compound library, a random predictor would have a value of 1 for all "sub-libraries". (**C**) Representative models of top-scoring complexes involving compounds known to be active against each of these protein targets, for which no structural information is currently available. (**D**) Representative models of top-scoring complexes involving "decoy" (presumed inactive) compounds. To date, these compounds have not been explicitly tested for activity against the corresponding protein targets.

Ultra-high-throughput virtual screening for a variety of protein-protein interaction targets

Previously we compiled a set of 18 complexes, each corresponding to a crystal structure of a small-molecule inhibitor bound at a protein interaction site (184); in 11 of these cases the binding site is only composed of the biological unit (i.e. the stoichiometry of binding is clearly 1:1), and a structure of the unbound protein is also available. None of the proteins in the set are homologs of one another; for

proteins in which more than one inhibitor-bound structure was available, we retained only the structure involving the most potent inhibitor. These 11 complexes comprise the majority of available non-redundant structures of small-molecule inhibitors of protein-protein inhibitors, and are presented in **Table S1**. For each of the inhibitor-bound structures, we removed the ligand and generated an exemplar from the inhibitor-bound conformation. For each of the unbound protein structures, we built an ensemble of pocket-optimized conformations; upon clustering, we retained between 9 and 27 distinct conformations for each (depending on the number of high-population clusters that were identified), and used these exemplars for our subsequent screening benchmark. As described above, we collected a non-redundant and diverse set of inhibitors for each target protein, drawing from the TIMBAL (170, 171) and CHEMBL (172) databases in addition to the PDB. These known inhibitors were added to the collection of ~8.33 million drug-like compounds from the ZINC database (182), which served as "decoy" compounds in this experiment.

Exemplar screens were carried out using either a crystallographic inhibitor-bound conformation, the single lowest-energy pocket-optimized conformation, or the ensemble of distinct pocket-optimized conformations after clustering. The rank of each active compound, relative to the set of decoy compounds, was then determined for each case: a summary of these data are presented in **Table 1**, **Table S2**, and **Table S3**.

Focusing first on the results from screening against an ensemble of pocket-optimized conformations (**Table 1**), the enrichment of active compounds is evident. Remarkably, for 4 of the 11 protein target, a known active compound is ranked in the top 10 (of more than 8 million compounds). Further, known active compounds are included in the top 1% of scored compounds for 9 of the 11 protein targets. It is once again worth noting that not all of the decoy compounds that score more favorably than these active compounds are necessarily inactive: they are simply presumed inactive for the purposes of this experiment, since they have typically not yet been tested for activity against these protein targets.

Protein name	Best rank of known active	Number of known actives in top 100	Fraction of known actives in top 1%
HIV-1 integrase (87)	2	4	0.16
Mdm2 (34)	175	0	0.24
Bcl-xL (25)	10	1	0.08
BRD4 (20)	208209	0	0.00
XIAP-BIR3 (12)	45043	0	0.08
Grb2-SH2 (9)	3	3	0.78
Interleukin-2 (6)	1573	0	0.17
HPV E2 (4)	20111	0	0.50
Menin (3)	1	1	0.33
WDR5 (3)	613564	0	0.00
VHL (1)	25376	0	1.00

Table 1: Ultra-high-throughput virtual screening for a variety of protein-protein interaction targets, using an ensemble of pocket-optimized conformations. A collection of diverse known inhibitors for each protein target were hidden amongst a set of 8.33 million drug-like "decoy" compounds (the number of active compounds for each protein target is in parentheses). The compound library was then screened using multiple exemplars, corresponding to an ensemble of pocket-optimization simulations started from the *unbound* protein structure in each case. Corresponding data are also presented for screens carried out using an inhibitor-bound protein structure (**Table S2**), or using only the single top-scoring pocket-optimized conformation (**Table S3**).

It must also be pointed out that screening against the ensemble of conformations was necessary, in most cases, to achieve this level of performance. While screening against an inhibitor-bound conformation did yield a top-ranked active compound in 3 cases, this approach did not identify a known active compound in the top 1% for *any* of the other 8 protein targets (**Table S2**). Meanwhile, the use of a single pocket-optimized conformation found known active compounds in the top 1% for 4 of the 11 protein targets, but these were typically not ranked as highly as those obtained by screening using the other approaches (**Table S3**).

3.5 Discussion

Identifying small-molecule inhibitors of protein-protein interactions remains a challenging task. Because these protein surfaces are typically not evolved to bind small molecules, they often lack the deep binding pockets to bind small molecules with high affinity (8, 185). In the structurally-characterized examples to date, some small-molecule inhibitors are found to bind using pre-existing pockets whereas others take advantage of transient pockets that are not observed in crystal structures of the unbound protein. This presents a natural problem for structure-based virtual screening approaches, since they require knowledge of the inhibitor-bound protein conformation (or at least something close to it): at protein interaction sites the unbound protein conformation may not bear sufficient resemblance to the inhibitor-bound conformation – in terms of available surface pockets – to allow successful virtual screening. Ligand-based methods for virtual screening are even less applicable, if very few known inhibitors are available.

In this study, we use biased simulations to predict low-energy pocket-containing conformations, starting from the unbound protein structures. We then use "exemplars" derived from these conformations as templates for rapidly screening very large compound libraries. By targeting protein conformations that would not otherwise be available without prior ligand-bound crystal structures, we enable discovery of potential inhibitors that would not be identified using other methodologies.

Comparison to traditional docking

For small-molecule inhibitors of protein interactions, the exemplar-based screening approach presented here offers two primary advantages over traditional, docking-based methods. The first advantage is accuracy: this is a notably challenging regime for traditional docking methods (185). The screening benchmark we use in this study is intentionally difficult, since the TIMBAL dataset is composed entirely of compounds that inhibit various protein-protein interactions and is thus devoid of decoy compounds that can be easily ruled out as unsuitable. When screening against a ligand-bound protein structure we find that exemplar screening performed with similar accuracy as DOCK 6 and AutoDock4. However, screening against a "pocket optimized" conformation revealed the sensitivity of these methods to details of the protein structure, which in turn led to diminished performance. In contrast to this, overlaying ligands with exemplars strongly favors matching of the *overall* shape and chemical features, which in turn provides robustness to slight variations in the protein structure.

The second advantage, more dramatic than the first, is the fact that the exemplar-based screening approach is orders of magnitude faster than docking. Because large-scale virtual screens that involve docking are often constrained by computational requirements, compound libraries are typically filtered to remove redundancy while at the same time maximizing the diversity within the dataset (186, 187). However, small changes to a compound can have profound effects on binding affinity (188, 189); as such, screening a complete library may yield hits from chemical series that may otherwise have been excluded based on the score of a single representative member included in the non-redundant set.

The speed advantage of exemplar-based screening approach relative to traditional docking approaches, moreover, also enables screening against multiple receptor conformations; this would not necessarily be feasible for a large compound library using traditional docking, given the computational expense of screening against even a single receptor conformation.

Comparison to sidechain mimicry

An emerging approach for designing of inhibitors of protein-protein interactions without docking entails mimicry of certain functional groups presented by the partner protein (or peptide). This approach has recently been demonstrated by Koes and colleagues: their approach identifies "anchor residues" important for a protein-protein interaction, then combines these into pharmacophores to search for compounds that mimic the three-dimensional geometry of these residues' sidechains (30). Given the over-abundance of certain sidechains as "anchor residues" (phenylalanine, tyrosine, tryptophan, valine, and leucine), this approach lends itself naturally to construction of custom screening libraries (accessible through multi-component reaction chemistry) containing compounds mimicking these sidechains (31). Others have also used similar methods to extract specific sidechains – or smaller functional groups – from protein complexes, and defined pharmacophores based on these as templates for virtual screening (32-37).

Because their overlay with the protein's natural binding partner is evident, the use of mimicry produces highly intuitive hit compounds. An important limitation, however, is that these hits represent a very limited subset of potentially active compounds. The interactions of the protein partner may not be optimal (given constraints of the protein geometry and/or lack of evolutionary pressure for tight binding), and this approach cannot be extended to conformations of the target protein other than that observed in the protein-protein complex. As noted earlier, many high affinity inhibitors acting at protein interaction sites take advantage of precisely these alternate conformations, because they contain surface pockets more suitable for small-molecule binding (8).

The exemplar-based screening we describe here is notably different from these approaches, in that it does not rely on mimicry of the protein partner, but instead builds a pharmacophore entirely from the structure of the target protein. Accordingly, it is not subject to the limitations described above: the resulting hits can fill the exemplar without carrying forward artifacts from the protein partner, and can be applied equally well to other conformations of the target protein beyond simply the one used for protein-binding.

Comparison to other receptor-based pharmacophore methods

In the absence of small-molecule inhibitors (or natural substrates), others have also used features of the receptor to build pharmacophores (91, 92, 135, 136, 190-193); like these, the "exemplars" we describe here are essentially receptor-based pharmacophores. In general there are two types of approaches for building receptor-based pharmacophores: either by using the geometry of the receptor to generate the pharmacophore directly (135, 136) (as we have done in this study), or else by docking a series of "probe" compounds against the receptor to identify potential interaction sites at which these probes accumulate (91, 92, 190-193). For the most part, however, the use of receptor-based pharmacophores has thus far been confined to "traditional" drug target classes, and not extended to protein interaction sites.

With the exception of the studies presented here, there has been – to our knowledge – only one other application of receptor-based pharmacophores for targeting a protein interaction site. In order to identify compounds inhibiting assembly of the heterodimeric HIV reverse transcriptase active complex, Grohmann and colleagues developed such an approach (193). To account for receptor flexibility this group carried out a short molecular dynamics simulation, selected six snapshots, and then docked probe molecules against these conformations in order to identify potential interactions; this led to three unique pharmacophores. They then used these pharmacophores to carry out a virtual screen, and used the compounds that overlaid with their pharmacophores as a starting point for docking. The top-scoring compounds in these docking studies were advanced into a suite of biochemical assays, in which one of these compounds was found to exhibit (weak) activity (193).

The computational pipeline we describe here (**Figure 1**) differs from that of Grohmann and colleagues in several key respects. First, applying pocket optimization to generate a collection of receptor structures provides an efficient and effective means for sampling diverse low-energy conformations (76, 161); in contrast, others have shown through molecular dynamics simulations of the same protein interaction sites that up to half of the corresponding pockets – which may be required for small-molecule binding – are very seldom sampled through (unbiased) molecular dynamics simulations (40). The pocket

88

optimization simulations have the additional advantage of requiring relatively modest computational resources, and they scale in a trivially parallel manner to yield a linear speedup with the number of available processors.

The second important difference results from the approach to building and applying the pharmacophore model. The computational pipeline described here does not include docking at any stage: the pharmacophore is built solely from the geometry of the receptor (without docking probes), and then models of the bound complexes are generated directly from the overlays to the pharmacophore (without docking the hit compounds). By avoiding explicitly protein-ligand docking at every stage, we enable rapid screening of large ensembles of receptor conformations, against very large compound libraries.

The approach of Grohmann and colleagues was immediately applied to HIV reverse transcriptase (193), without examining the performance of this approach through any benchmarking studies. Moving forward, it will be interesting and exciting to parse the relative merits of each step in these computational pipelines, and carefully determine the most effective method for identifying compounds that are potent and selective inhibitors of many other targets.

Ensemble docking

Already others have noted that docking against multiple receptor conformations ("ensemble docking") can enhance performance relative to docking against a single static conformation (152-160). Indeed, our "ensemble docking" benchmark confirmed that using a collection of distinct conformations generated from an unbound structure, we obtained similar performance to that observed when using a protein structure solved in complex with an inhibitor (**Figure 5**). Further, when performing virtual screens across a diverse series of targets, screening against an ensemble of structures outperformed the analogous screens against an individual protein conformation, allowing discovery of more diverse active compounds (**Figure 6**, **Table 1, Tables S2 and S3**).

Prior studies have sought to incorporate multiple conformations when using sidechain mimicry (34-36) or pharmacophores built from docking probe compounds (191-193). Typically, these approaches then consolidate pharmacophores into a small number of consensus templates for virtual screening. A potential risk inherent to this step is that the average location of a feature across several conformations may correspond to a position that is not explicitly compatible with any of the individual conformations that contribute to this average. By instead creating and screening with multiple exemplars explicitly, we aim to identify complexes in which the ligand is truly optimal for a particular protein structure – rather than for an averaged protein conformation that may not be reflective of the individual members of the underlying ensemble.

Advantages of pocket optimization

These studies also provide another demonstration of the utility of pocket optimization to guide ligand discovery. Previously we showed that pocket optimization can be used to evaluate the druggability of a given protein surface though its propensity for forming pockets (76). We also showed that the *shapes* of these pockets dictate the space of "allowed" ligands that can interact with a given surface, and thus can dictate inhibitor selectivity (161). This is further underscored through the results presented here, in which we demonstrate directly that features of these pockets can be used to identify complementary ligands, even in the absence of a ligand-bound structure to use as a starting point.

Putting these pieces together, we envision a unified pipeline for ligand discovery. Starting from the unbound protein structure, one could use pocket optimization to gauge druggability of the target surface. To ensure selectivity of the resulting inhibitors, one would then select only those pockets that are sampled by the desired protein target(s), and not by other members of this protein family. Exemplars built from these pockets would serve as templates for virtual screening, to identify inhibitors that recognize pockets unique to the protein target(s). Collectively, this approach offers the prospect of not only identifying compounds that bind to the protein target, but that do so with the desired selectivity.

Finally, we note that this approach may also enable "cryptic" allosteric sites to be addressed through novel ligands. In the absence of a binding partner at this site mimicry would not be possible, and the lack of a pre-formed binding pocket would render traditional docking unsuitable. In light of the multitude of such sites that potentially exist (194), the exemplar screening method presented here may enable identification of new allosteric inhibitors that would be exceedingly challenging to discover using other existing methods.

3.6 Acknowledgements

3.7 Associated Content

Supporting Information Available: Supplementary Methods, Supplementary Tables. This material is available free of charge via the Internet at http://pubs.acs.org.

CONCLUSIONS

Protein-protein interfaces continue to be a challenging class of drug targets. Not having been evolved to bind small molecules, they often lack the deep binding pockets needed to bind small molecules with high affinity (8, 185). Inspired by examples such as interleukin-2, where the ligand occupies a hydrophobic groove that is absent in the unbound and protein bound conformations (75), we sought to enable the task of computationally identifying inhibitors of protein-protein interactions by identifying low energy conformations that contain surface pockets that are more suitable for binding small molecules.

To this end we introduced a grid-based "pocket optimization" protocol to sample pocket containing conformations. Pocket optimization quickly creates ensembles of conformations that contain larger pockets than unbiased simulations, while having energies that overlap with those unbiased simulations, as shown in both Chapters 1 and 2. Using the resulting ensembles of "pocket optimized" conformations, we sought to address several key issues regarding targeting protein-protein interfaces: whether a particular protein interface is druggable, predicting the selectivity of a compound across a protein family, and identifyingof novel inhibitors of protein-protein interactions.

Druggability of protein-protein interaction sites

In Chapter 1 we implemented and improved upon the LIGSITE[CS] pocket detection algorithm (50) to discriminate between pockets observed in unbound conformations and pockets observed in ligand bound conformations at protein-protein interaction sites. With this pocket detection algorithm in place, we introduced the "pocket optimization" protocol. Using the change in pocket volume at a specified location on the protein surface as a proxy for the binding free energy of some (unspecified) small molecule partner at that location, we sample low energy, pocket-containing conformations.

Having compiled a set of seven protein-protein interaction sites with structures solved for unbound, small-molecule bound, and protein bound conformations, we sought to identify what makes

those surfaces druggable. Applying the pocket optimization protocol to target locations at the druggable interfaces of protein-protein interactions as well as random locations on the protein surfaces, we observed that pockets could only form at the druggable interfaces. Indeed, the only way to sample pockets away from the druggable interfaces was through over-weighting the biasing potential, and thus sampling non-physical conformations. That led us to the conclusion that the lack of a surface pocket, and not the chemical composition, is what causes certain protein surface sites to be undruggable.

In the set above, we noted that some pockets are large and pre-formed (HPV E2, FKBP12, and ZipA) with sidechain reorganization driving formation of the small molecule binding pocket. However, other cases (XIAP, MDM2, Bcl-X_L, and IL-2) lack a pre-formed pocket, and formation of the small-molecule binding pocket requires concerted motions that couple backbone conformational changes to sidechain reorganization. Looking at the ensemble of pocket-optimized conformations versus the unbiased ensemble, we observed that, for both cases above, the pocket optimized ensemble sampled more conformations closer in interface RMSD to the known small-molecule bound conformations.

This observation suggests that intrinsic structural and dynamic features of the protein encode the low-resolution inhibitor shapes found in the low-energy ensemble that protein may sample. As one of these preferred shapes is displayed, a complementary small-molecule ligand may be accommodated, followed by a smaller inducement of conformational changes to match the high-resolution steric and chemical details of the ligand. In such a model, the low resolution shape of the pocket may give insight into the low resolution shapes of complementary ligands.

Selectivity of protein-protein inhibitors

Selectivity of an inhibitor towards its target(s) is an important property that is optimized through the drug discovery process. When a small molecule has unintended interactions with unrelated proteins or family members of the target protein, there arises the potential for adverse effects (toxicity) (17). We

drew inspiration from ABT-263 (navitoclax), a Bcl-2 inhibitor whose effective dose was limited due to toxicity (thrombocytopenia) from its interaction with Bcl-xL (18, 19), to try to predict these potential off-target interactions between closely related family members.

Drawing from our conclusions from Chapter 1 that low-resolution inhibitor shapes are encoded on the protein surface and sampled via low-energy fluctuations, we sought to address selectivity by the presence or absence of the low resolution shape in the ensemble of low-energy, pocket containing conformations. We introduced the concept of "exemplars": non-physical "molecules" that define the shape and chemical features of a pocket. We found that comparing the shape and chemical overlap of exemplars derived from small-molecule bound conformations could be used as a proxy to comparing the 3-dimensional ligand conformations themselves, thus allowing protein-sequence independent comparison of pocket shapes.

Because of its clinical significance, we chose the Bcl-2 family as our model system. Using the pocket optimization protocol described in Chapter 1, we generated ensembles of pocket optimized conformations of all Bcl-2 family members with available unbound structures. From that ensemble, we created a set of exemplars to represent the binding pocket from each conformation. We observed that the known small-molecule binding pockets were spanned by the pocket optimized ensemble, adding further evidence towards the conformation selection model described in Chapter 1. Indeed, the presence or absence of a known binding site in a protein's ensemble was shown to be an excellent predictor for activity/inactivity.

Extending this analysis to all Bcl-2 family members, we observed that certain shapes are available to all family members. An inhibitor matching such shapes would be expected to show broad specificity across the family. Indeed, most known inhibitors lay in regions of "pocket space" sampled by more than one family member and have been found to be active against more than one family member. However, we also observed that many Bcl-2 family members sampled pockets that were not accessible to the other family members. Such "distinct" pockets offer a new strategy for drug design: building

94

selectivity into the low resolution details of the chemical scaffold, to more robustly maintain the desired selectivity during medicinal chemistry optimization of bioavailability, pharmacokinetics and pharmacology.

Discovery of novel small-molecule inhibitors

As mentioned above, some small-molecule inhibitors bind pre-existing pockets, while others bind transient pockets that are not observed in the absence of the ligand. This poses a problem for traditional structure-based screening approaches, as they require a conformation bearing resemblance to an inhibitor-bound protein conformation. Without suitable pockets on the protein surface, structure-based virtual screening has a much lower chance of success. In the absence of known inhibitors, of course, ligand-based screening methods are even less useful.

To address these challenges, we sought to screen against members of the ensemble of pocket-containing conformations. By using the exemplars from these conformations as a template for screening, we used GPU computing to rapidly compare multiple conformations against very large compounds libraries. Targeting previously inaccessible conformations may allow discovery of inhibitors that would otherwise not be identified by other methodologies.

This exemplar-based approach has two advantages over traditional, docking-based approaches. The first advantage is accuracy. When screening against the TIMBAL dataset composed of various inhibitors of protein interactions, a difficult dataset devoid of decoy compounds easily ruled out, we observed that exemplar screening performed with similar accuracy to DOCK 6 and AutoDock4 when screening against a known inhibitor bound structure. However, those docking approaches suffered from diminished performance when docking to a low energy, pocket optimized conformation that started from an unbound structure. In contrast, exemplar screening was robust to slight variations in the protein structure, by matching the overall shape and chemical features.

The second advantage over docking-based approaches is speed — exemplar screening is several orders of magnitude faster than docking. In order to accommodate computational limitations, compound libraries are typically filtered to remove redundancy while maximizing diversity, to use in large-scale virtual screens that involve docking (186, 187). However, profound effects on binding affinity have been observed from small changes to a compound (188, 189). It is not unreasonable to conclude that screening a complete library can identify hits that may otherwise be excluded based on the score of a single representative member of the chemical series in a non-redundant set.

Because of the fast speed of the exemplar screen, it is possible to screen more than one receptor conformation. It has previously been observed that performance can be enhanced by docking against multiple receptor conformations ("ensemble docking") as opposed to docking against a single static conformation (152-160). This observation was, indeed, confirmed by our ensemble docking benchmark. When screening against an ensemble of conformations generated using pocket optimization starting with an unbound structure, in the absence of any information of a ligand, we obtained similar performance to a screen that used a ligand-bound protein structure. This exciting observation lends evidence that an exemplar screen can be used to find novel inhibitors truly *de novo*.

Future work

There are several different ways in which this research can be expanded upon in the future. First, the exemplar screening protocol can be combined with the selectivity analysis to screen an ensemble of "unique" conformations for the *de novo* design of selective inhibitors. Indeed, the results of Chapter 2 showed that there are protein-exclusive pocket shapes within the low energy ensemble of pocket-containing conformations. The speed of exemplar screening enables its use with very large compound libraries, increasing the chances of finding an inhibitor to fit a specific pocket shape.

Second, pocket optimization could be performed at a known allosteric site to sample the shapes that surface pockets may adopt at a known allosteric site. Because exemplar screening is not tied to a known ligand or natural protein partner, this screening methodology might then be applied to identify novel allosteric inhibitors. Similarly, this approach can be applied at other druggable sites on the protein surface to explore variations in pocket shapes to identify novel small molecule binders.

Next, because many protein-protein interactions are cancer targets, drug resistance inevitably becomes a concern. Certain mutations can directly change the shape of the binding site (195) while others can shift the population of pocket shapes within the ensemble (196), decreasing the binding affinity and, hence, the potency of a small molecule. For a given protein target, ensembles can be generated for the commonly observed mutants in addition to the wild type protein. Performing a similar selectivity analysis could be used to identify likely conformations that may lead to resistance mutations, and identification of new compounds that do not allow for resistance. Likewise, selectivity analysis between wild type and mutation states could be used to identify pocket shapes not sampled by the wild type protein, to better target diseased states.

Finally, this work can be used to enhance the RNA hotspot mimicry protocol introduced by Ragul Gowthaman and Yan Xia (197). In RNA hotspot mimicry, the rings from the base pairs and nearby hydrogen bonds are used as a pharmacophore to screen against large databases, much like exemplars. First, exemplars can be used during the hotspot pharmacophore generation to enrich the set of hydrogen bond donors and acceptors to take advantage of donors and acceptors on the protein surface that a small molecule may interact with. Increasing the number of hydrogen bonds interactions that a ligand makes with the receptor will increase its binding affinity. Second, RNA hotspot mimicry was used with a single, static structure. When NMR models of protein-RNA interactions exist, then a screen can be performed on the entire NMR ensemble. Indeed, this can be worked into the exemplar screening pipeline to potentially identify more potent compounds that would otherwise have been missed. Finally, the concept of assessing selectivity using ensembles can be expanded to the protein-RNA model by modeling the fluctuations that

97

may be overserved with a protein-RNA complex. Identifying other, low-energy RNA-bound states that are not observed by x-ray crystallography in other complexres could be used to gain more confidence in the selectivity of a given RNA inhibitor, and also help identify unique pharmacophores from an ensembles.

APPENDIX A: Supporting Informations

A.1 Supporting Information for Chapter 2

A1.1 Supplementary Methods

Implementation in Rosetta

Computational methods are implemented in the Rosetta software suite (56). Calculations were carried out using svn revision 48537 of the developer trunk source code. Rosetta is freely available for academic use (www.rosettacommons.org), with the new features described here included in the 3.5 release.

Identifying Pockets on Protein Surfaces

A three-dimensional grid of size 20 Å x 20 Å x 20 Å and grid spacing 0.5 Å is centered at the center of mass of the "target" residue (a residue at the center of the known protein interaction site). Grid points occluded by the protein are identified using the center of mass and van der Waals radius of each heavy (non-hydrogen) atom in the protein. The protein surface is then mapped using an expanded radius that includes a 1.0 Å buffer region, and surface points in direct contact with the target residue are marked as such. Grid points not occluded by the protein (including "surface" points) are labeled "solvent".

For each "surface" point on the grid, a ray is traced in each direction of the six principle axes and four diagonals (corresponding to half the faces and half the corners of a cube centered at this grid point). Solvent points are marked as "pocket" when a ray contains a surface point followed by at least 3 Å, but no more than 12 Å, of consecutive solvent points followed by another surface point (**Figure S1**).

Though appropriately identifying known inhibitor binding sites, this algorithm proved susceptible to defining certain classes of spurious pockets not suitable for interacting with small-molecules. The most common were surface (saddle-shaped) "channels," in which the pocket is bounded on only two sides. To eliminate these, our algorithm carries out a second search along the same 10 directions. Any grid points marked as "pocket" that are bounded on both sides by only solvent are reverted to solvent themselves (**Figure S1**).

Adjacent grid points defined as "pocket" are used to determine the volume of the largest contiguous single pocket in contact with the "target" residue. The pocket volume is taken to be the number of pocket points that are not within 2.5 Å of a solvent grid point, multiplied by 0.125 Å^3 (the volume associated with each point based on the 0.5 Å grid spacing). To eliminate effects of the precise grid orientation with respect to the protein, the grid is randomly rotated and the pocket identification is repeated. The final pocket volume is taken to be the pocket volume averaged over 100 rotations of the grid.

The Rosetta command line used to determine the pocket volume as described above is as follows (for target residue number 108):

```
pocket_measure.linuxgccrelease -s input_pdb -num_angles 100 -pocket_max_spacing 12 -pocket_psp
false -pocket_sps -central_relax_pdb_num 108
```

Selection of random surface residues

Random surfaces were selected using the following criterion: secondary structure in the unbound structure matches those of the target residues, the residue must be at least 12 Å from the protein interaction site, and the residue must be at least 12 Å from all other residues selected so far. The selection process was carried out iteratively until no additional residues met these criteria. Only those with pocket size (evaluated by Q-SiteFinder (41)) matching the protein interaction site (**Figure S3**) were included in **Figures 2** and **3**).

Simulations were carried out using the now-default "fast-relax" variation of the relax protocol

(71) in Rosetta, with an additional biasing term proportional to the volume of the target pocket (in $\mathring{A}^3$)

included in the standard Rosetta energy function. The energetic contribution of this term was capped

when the pocket volume reached 300 $\mathring{A}^3$. In order to reduce simulation time, we computed pocket volume

using only two random grid orientations (as opposed to averaging over 100 rotations of the grid as

above). The fast-relax protocol used in these studies was comprised of five cycles (Monte Carlo steps) per

trajectory, where each cycle entailed repacking of the side chains with the repulsive part of the van der

Waals term incrementally ramped up in weight, followed by a gradient-based energy minimization. The

last of these, the gradient-based energy minimization, formally requires derivatives of the pocket volume

with respect to the protein atomic coordinates that are not solvable analytically. For this reason we set the

derivatives of the "pocket" contribution to zero, meaning that this minimization step is carried out without

regard to the pocket size. Despite this, application of the Metropolis criterion using energies that include

this "pocket" contribution nonetheless serves to bias the resulting trajectory towards pocket-containing

conformations.

The Rosetta command line to carry out the unbiased simulation described above is as follows:

```
relax.linuxgccrelease -relax:fast -s input.pdb -nstruct 1000
```

The equivalent biased simulations were carried out by adding the command-line flags:

```
-pocket_zero_derivatives -pocket_max_spacing 12 -pocket_psp false -pocket_sps -pocket_num_angles
2 -score:weights score12 -score:patch pocket.wts.patch -cst_fa_file constraints.txt
```

The pocket.wts.patch file contains the single line "pocket_constraint = 1.0" and the constraints.txt

file contains a single line of the form "pocket_weight target_residue" where the pocket weight has units

REU per $\mathring{A}^3$ and target_residue can have formats such as 108 or A:108.

Unbiased energetic evaluation of conformations

The pocket-opened structures were minimized with Rosetta (56) using the standard Rosetta energy function to remove any clashes, and the final (unbiased) energy was reported again using the standard energy function.

This minimization step was carried out using the Rosetta command line:

```
minimize.linuxgccrelease -in:file:fullatom -s input.pdb
```

iRMSD calculations

Interface residues were defined as residues that have either an attractive or repulsive van der Waals (as implemented in Rosetta) interaction with the inhibitor of absolute magnitude greater than 0.01 REUs. All protein Cα atoms were used for alignment, then iRMSD was computed over all non-hydrogen atoms of the interface residues.

PDB structures used in calculations

Results were described for seven proteins comprising our test set: Bcl-X$_L$, IL-2, FKBP12, HPV E2, ZipA, MDM2, and the BIR3 domain of XIAP.

All unbound calculations (and starting points for simulations) used structures with the following PDB IDs: 1R2D (Bcl-X$_L$), 1M47 (IL-2), 2PPN (FKBP12), 1R6K (HPV E2), 1F46 (ZipA), 1Z1M (MDM2), and 1F9X (BIR3 domain of XIAP). For NMR structures (1Z1M and 1F9X), we used the first model in the NMR ensemble as a starting point.

The pocket volumes shown in Figure 1B were calculated from bound structures with the following PDB IDs: 2YXJ (Bcl-X$_L$), 1M48 (IL-2), 1BKF (FKBP12), 1R6N (HPV E2), 1S1J (ZipA), 1RV1 (MDM2), 2JK7 (BIR3 domain of XIAP).

The iRMSD calculations to the bound conformations in Figure 5 used the following bound structures with PDB IDs. Bcl-X$_L$: 1YSG, 1YSI, 1YSN, 2O22, 2O2M, 2O2N, and 2YXJ. IL-2: 1M48, 1M49, 1M4A, 1M4B, 1PW6, 1PY2, and 1QVN. FKBP12: 1A7X, 1BKF, 1BL4, 1F40, 1FKB, 1FKD,

1FKF, 1FKG, 1FKH, 1FKI, 1FKJ, 1J4H, 1J4I, 1J4R, 1QPF, 1QPL, 2DG3, 2DG4, 2DG9, and 2FKE.

HPV E2: 1R6N. ZipA: 1S1J, 1S1S, 1Y2F, and 1Y2G. MDM2: 1RV1, 1T4E, and 1TTV. BIR3 domain of

XIAP: 1TFQ, 1TFT, 2JK7, 2OPY, 3CLX, 3CM2, 3CM7, and 3EYL.

Calculations described in Figure 6 were started from the unbound survivin structure, PDB ID

1E31.

Description of pocket-opened conformations

The pocket-opening of Bcl-X$_L$ is described in the main text.

IL-2 (**Figure S7A**) is similar to Bcl-xL in that a large clash is found when superimposing the

ligand onto the unbound structure. Several of the helical regions align well, but there are differences in

the alignment of the helix located behind the left side of the small molecule. This allows a side chain to

move and resolve the clash while at the same time allowing narrowing the left side of the pocket,

allowing more interaction with the inhibitor.

FKBP12 (**Figure S7B**), HPV E2 **Figure S7C**), and ZipA (**Figure S7D**) are each cases where the

pockets are largely pre-formed. The backbones of the bound, unbound, and pocket-opened structure of

FKBP12 all are in very close alignment with the exception of some movement of a loop region to make a

more favorable interaction with the ligand in the bound structure. This movement is partially recapitulated

in the opened structure. Likewise, superimposing the small molecule onto the unbound structure yields

no clashes. This is consistent with the distribution of pocket sizes (**Figure 2C**) where two of the target

residues had only a small increase in the pocket size when the biasing potential was applied. HPV E2 is

also largely pre-formed, though there is a clash at the deepest part of the small molecule pocket. The

backbones differ little, though the clash is resolved in the bound structure by the slight translation of the

helix containing the target residue and the rotation of a histidine side chain. ZipA is also lacking in

clashes and the backbone largely aligns around the interface. When larger pockets were identified

through biased simulations, the expansion was largely confined to the area adjacent to the binding site on the left side of the figures.

MDM2 (**Figure S7E**) has several significant clashes when the small molecule is superimposed onto the binding site. There are essentially two motions going on that open the pocket. The helix to the right wobbles to the right relieving two clashes with the small molecule while the helix below the small molecule tilts away, relieving the other clash. These motions were recapitulated in the biased simulation and yield visually impressive results.

The interface of the BIR3 domain of XIAP (Figure S7F) is somewhat disordered and has a large clash on one side when the small molecule is superimposed onto the unbound structure, and a pre-formed pocket on the other side. The pocket is opened by a loop moving out of the way, along with some compensatory wobble motion with a helix that the loop attaches to.

A1.2 Supplementary Figures
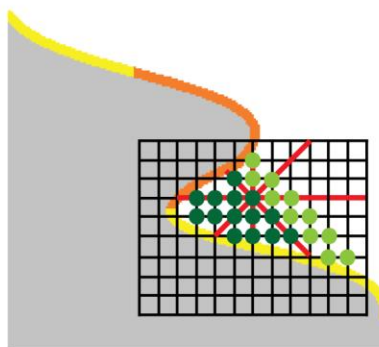
Figure S1: Identifying pockets on protein surfaces (complements Figure 1). A grid is centered at the residue of interest on the protein surface (only partial grid is shown here). Grid points are classified as either "protein" (*grey background*) or "solvent" (*white background*). Linear searches (*red lines*) are used to identify and mark "protein-surface-protein" events (*green gridpoints*). These are further classified based on degree of buried (*light vs dark green*) and adjoining grid points are clustered to yield discrete "pockets". Only a surface pocket in contact with the surface of the "target" residue (*orange*) contributes to the biasing potential.
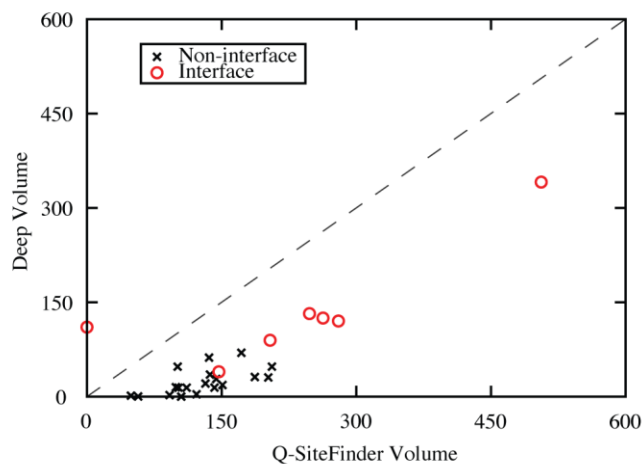
Figure S2: Deep pocket volumes compared to Q-SiteFinder pocket volumes (complements Figure 1). Deep pocket volumes of surface pockets at protein interaction sites harboring a bound inhibitor (*red circles*) and pockets found elsewhere on the protein surface (*black x's*) are plotted against the corresponding pocket volumes identified by Q-SiteFinder, for each of the sites used in Figure 1B. While the two are correlated, Q-SiteFinder volumes are typically larger than the corresponding deep pocket volumes.
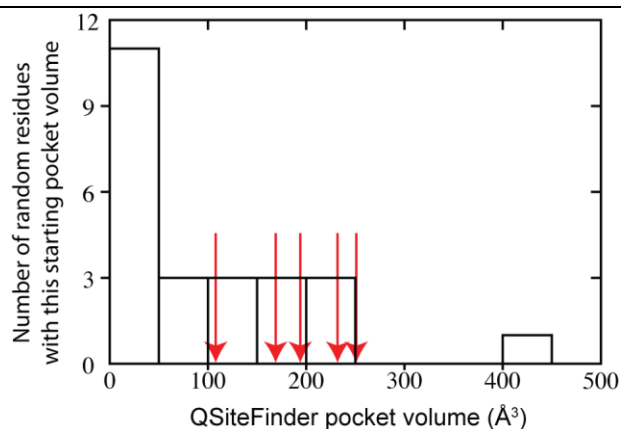
Figure S3: Matching of random surface sites to protein interaction sites (complements Figure 2). The distribution of pocket volumes identified by Q-SiteFinder for the complete set of random surface sites matched to the protein interaction sites on the basis of burial and secondary structure is shown (*black*); the corresponding pocket volumes for the protein interaction sites are also shown (*red arrows*). In order to create a matched set, only random sites with Q-SiteFinder pocket volumes greater than 100 Å^3 were considered for Figures 2 and 3.

Fraction of conformations containing such a pocket

Pocket volume (Å³)

108
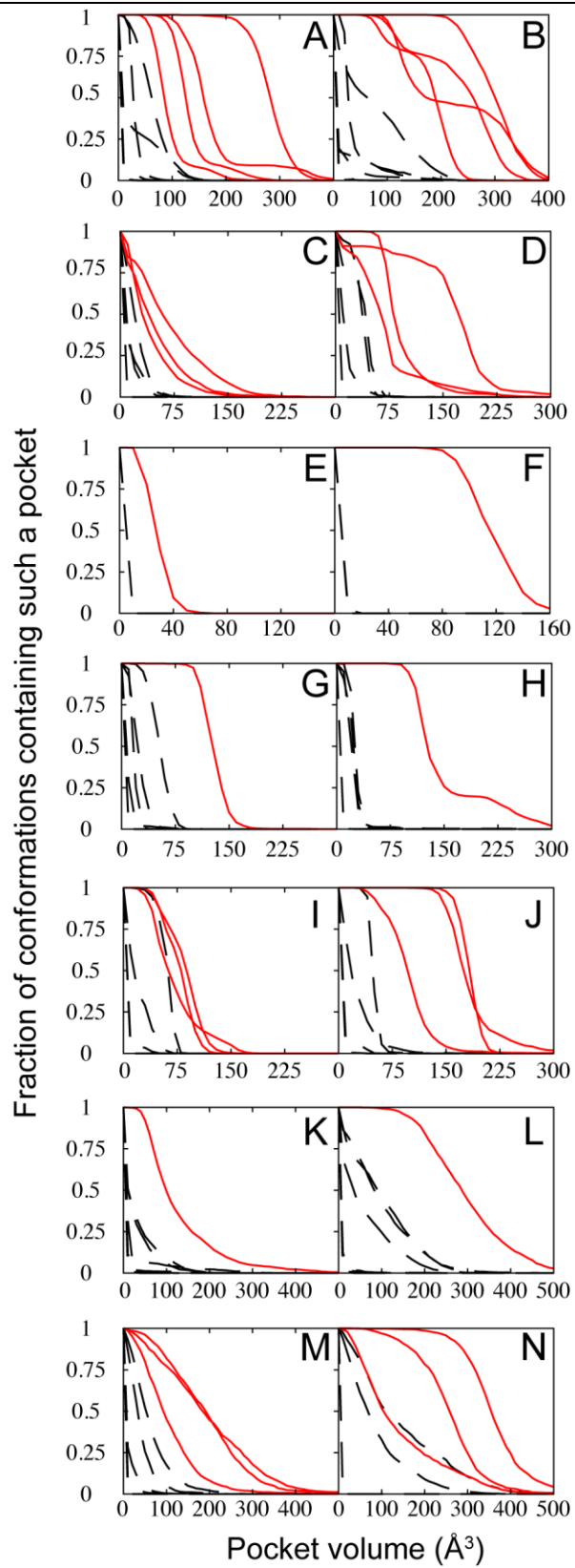
Figure S4: Surface pockets emerge only at druggable sites, upon inclusion of additional random surface sites (complements Figure 2). Pocket opening is not observed at random surface sites upon inclusion of additional sites that were previously excluded on the basis of Q-SiteFinder pocket volumes less than 100 Å^3 (*dashed black lines*). Symbols are as defined in Figure 2. (A) Bcl-X$_L$. (B) IL-2. (C) FKBP12. (D) HPV E2. (E) ZipA. (F) MDM2. (G) BIR3 domain of XIAP.

Figure axes — Columns: Unbiased, Biased. Rows A–G: BCL-xL, IL-2, FKBP12, HPV E2, ZipA, MDM2, XIAP. Y-axis: Fraction of conformations containing such a pocket. X-axis: Pocket volume (Å³).
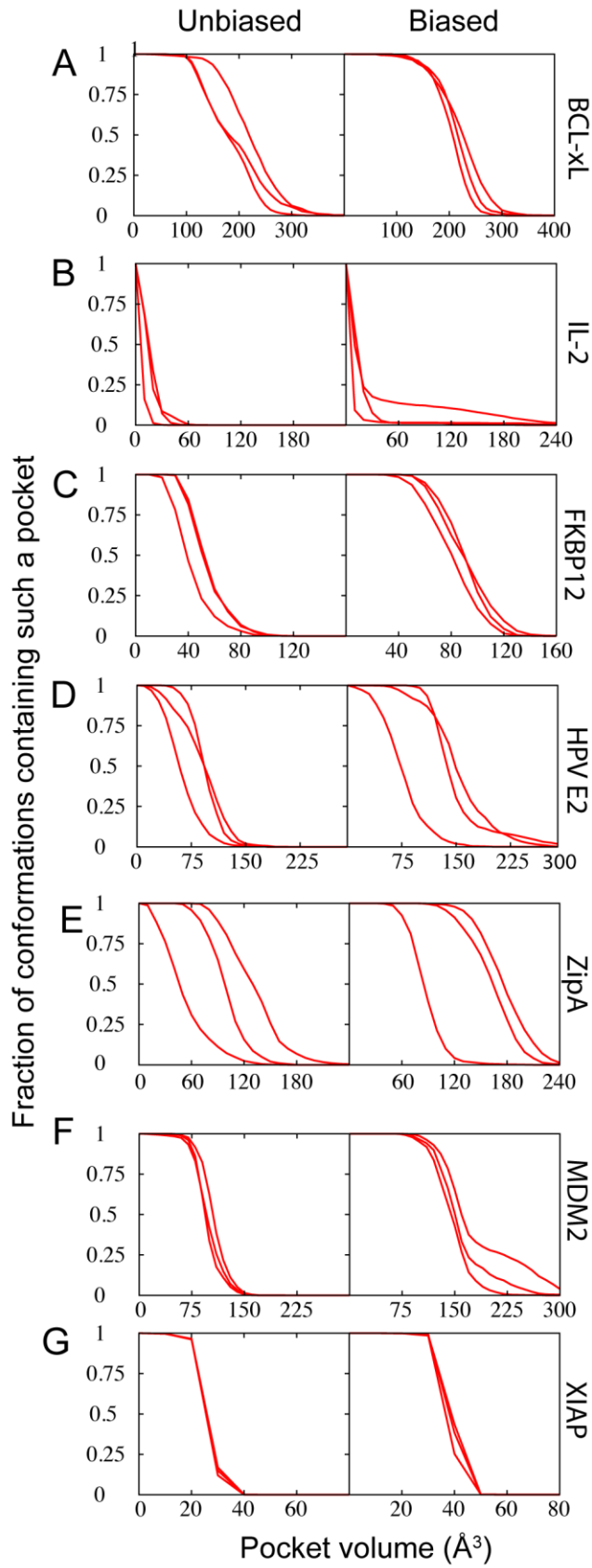
Figure S5: Surface pockets emerge at druggable sites, when starting from protein-bound conformations (complements Figure 2). Simulations were carried out only applying the biasing potential to the protein interaction site (*red*). With the exception of XIAP, results are in agreement with simulations started from the unbound protein structures (Figure 2). (A) Bcl-X$_L$. (B) IL-2. (C) FKBP12. (D) HPV E2. (E) ZipA. (F) MDM2. (G) BIR3 domain of XIAP.
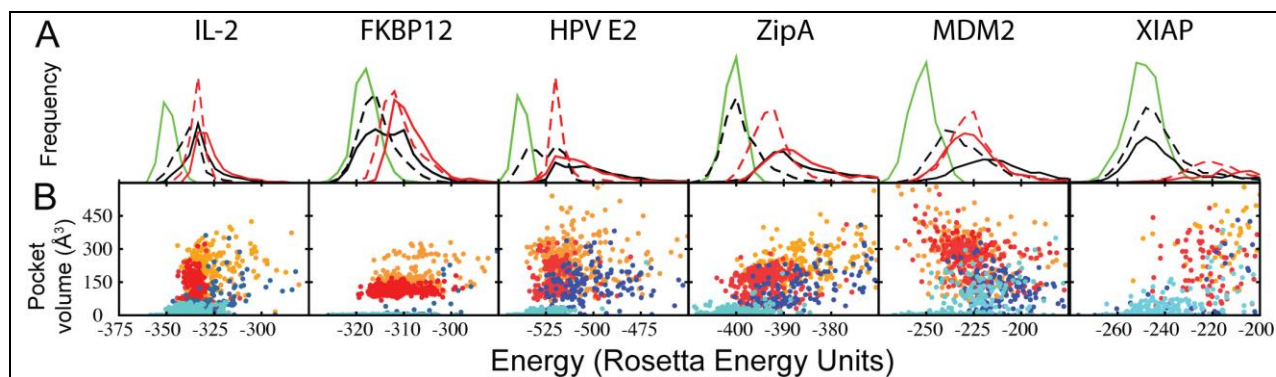
Figure S6: Energetic analysis of pocket opening for the other members of our test set (complements Figure 3). (A) Conformations generated using the biasing potential typically have a distribution of energies that overlaps those generated with the biasing potential, suggesting that these conformations represent low-energy states accessible to the unbound protein. (B) As with Bcl-xL, low-energy conformations containing large pockets are not observed for the other members of our test set unless the biasing potential is applied to the protein interaction site. Symbols are as defined in Figure 3.
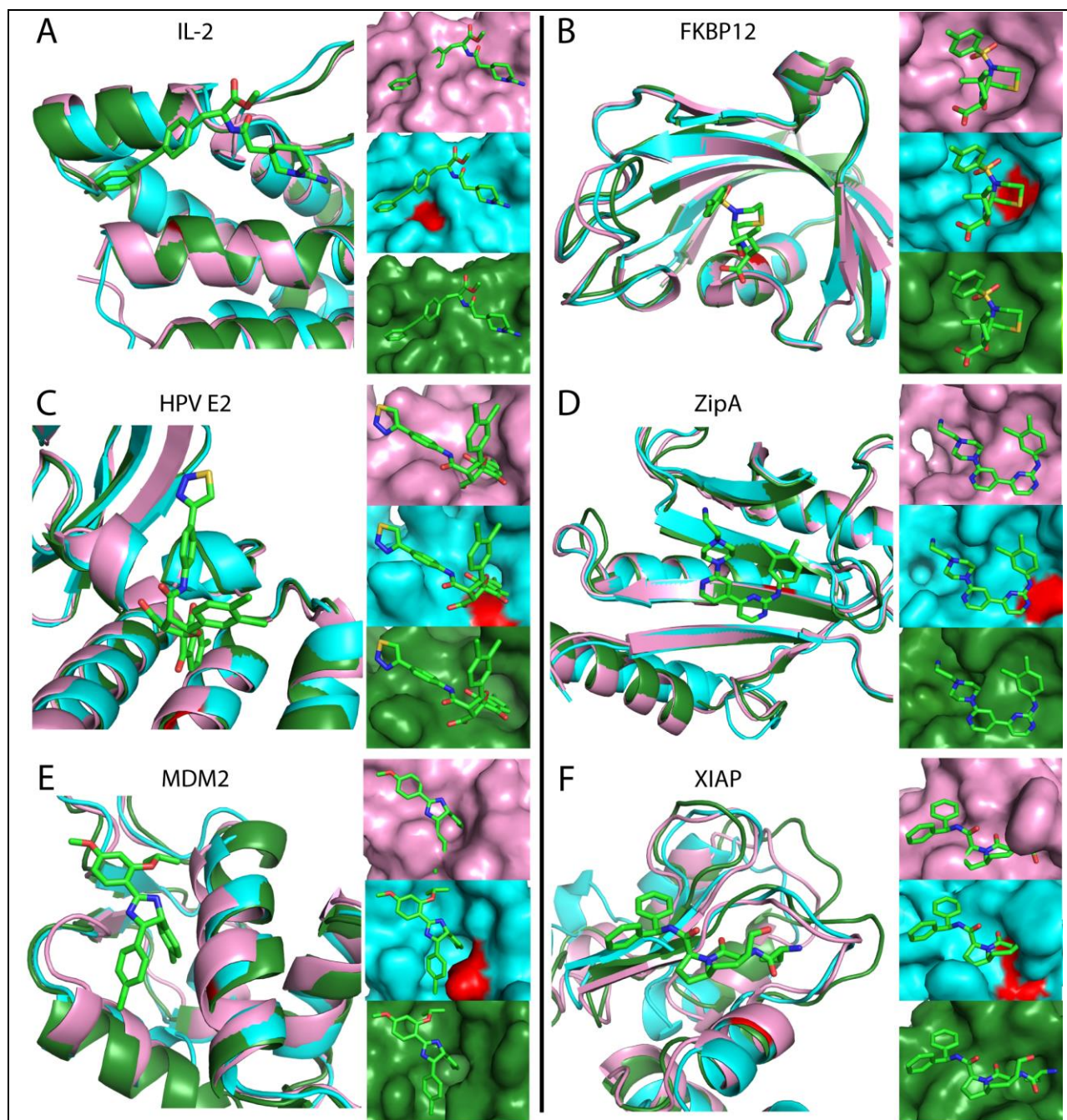
Figure S7: Representative conformations generated using the biasing potential (complements Figure 4).

An unbound crystal structure (*pink*), an inhibitor-bound crystal structure (*green, with inhibitor shown in sticks*), and a low-energy conformation generated from the unbound crystal structure using the biasing

potential (*cyan, with target residue in red*) are shown for each of the proteins comprising our test set (except Bcl-X$_L$, shown in Figure 4). (A) IL-2. (B) FKBP12. (C) HPV E2. (D) ZipA. (E) MDM2. (F) BIR3 domain of XIAP.
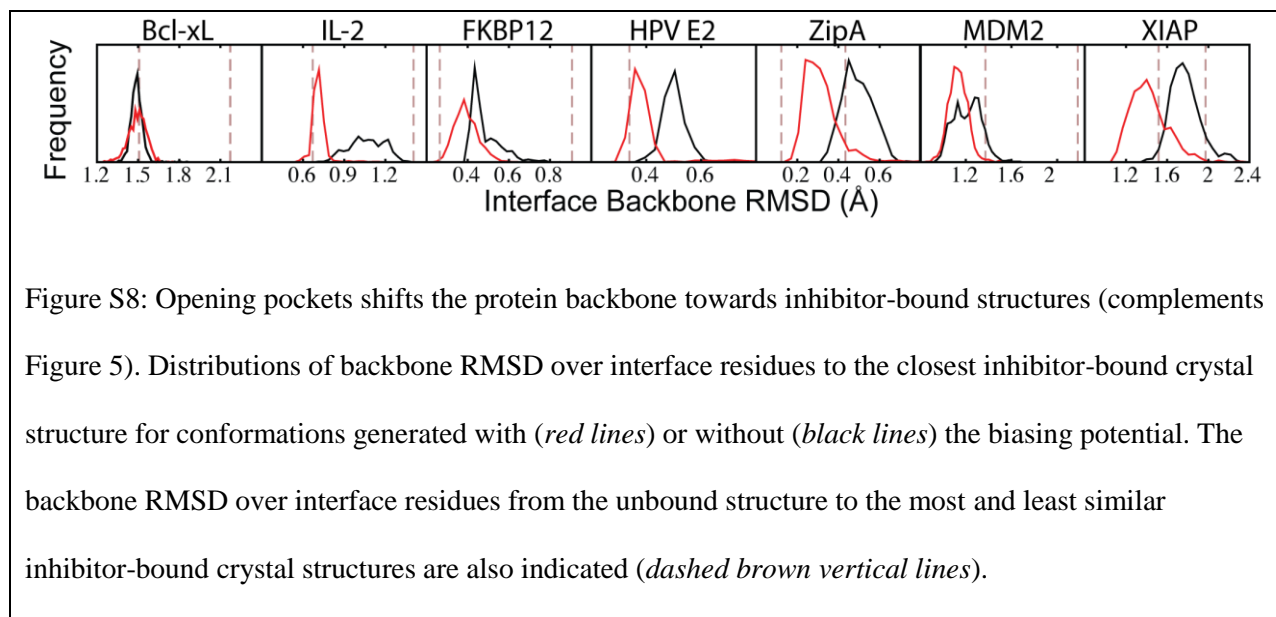
Figure S8: Opening pockets shifts the protein backbone towards inhibitor-bound structures (complements Figure 5). Distributions of backbone RMSD over interface residues to the closest inhibitor-bound crystal structure for conformations generated with (*red lines*) or without (*black lines*) the biasing potential. The backbone RMSD over interface residues from the unbound structure to the most and least similar inhibitor-bound crystal structures are also indicated (*dashed brown vertical lines*).
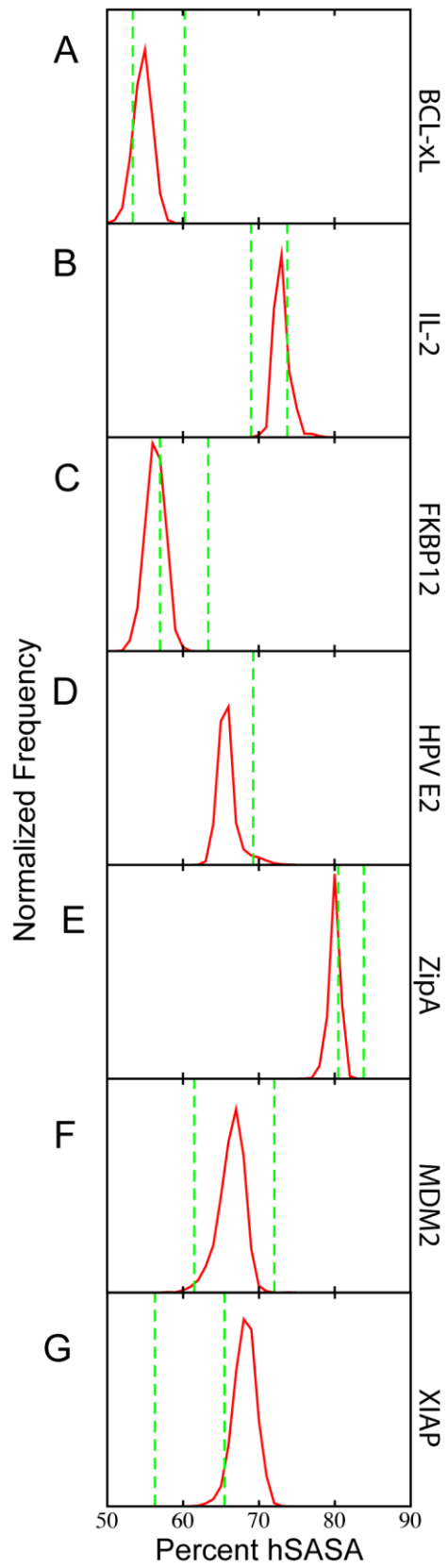
Figure S9: Hydrophobicity of pockets generated using the biasing potential are similar to the corresponding inhibitor-bound pockets (complements Figure 5). Distributions of percentage hydrophobic solvent accessible surface area (%hSASA) over interface atoms for conformations generated with the biasing potential (*red lines*). The lowest and highest %hSASAs from amongst all available inhibitor-bound crystal structures are indicated (*dashed green vertical lines*). (A) Bcl-X$_L$. (B) IL-2. (C) FKBP12. (D) HPV E2. (E) ZipA. (F) MDM2. (G) BIR3 domain of XIAP.

A.2 Supporting Information for Chapter 2

A2.1 Supplementary Methods

Rosetta software

Computational methods are implemented in the Rosetta software suite [18]. Calculations were carried out using git revision a27ab8d2408fb4a2 of the developer trunk source code.

A sample Rosetta command line used to identify target residues on chain A of an input.pdb using the Robetta ΔΔG calculations pasted into ddg.txt is as follows:

pocket_suggest_target_residues_by_ddg.linuxgccrelease -s input_pdb

–ddg_list ddg.txt –target_chain_list A

A sample Rosetta command line used to generate an exemplar is as follows:

make_exemplar.linuxgccrelease -s input_pdb -pocket_grid_size 12

-pocket_filter_by_exemplar –pocket_static_grid -central_relax_pdb_num 108

Dataset of Bcl-2 family complexes

Structures that met any of the following criterion were excluded from the set of known Bcl-2 family inhibitors: fragments below m.w. 250 Da, structures with multiple ligands in the binding site, structures that have the binding site defined by a (non-biologically relevant) homodimer, structures in which there was multiple occupancy of the inhibitor in the binding site, and fusion inhibitors that extended well out of the binding pocket. The complete list of complexes is presented as **Table S1**.

PDB structures used in calculations

Unbound structures used as a starting point for "pocket opening" simulations were from the following PDB IDs: 1R2D (Bcl-xL), 1G5M (Bcl-2), 1WSX (Mcl-1), 1MK3 (Bcl-w), 1F16 (Bax),

2BID (Bid), and 1OHU (Ced9). Peptide-bound structures used as a starting point for "pocket opening"

simulations were from the following PDB IDs: 1BXL (Bcl-xL), 2XA0 (Bcl-2), 2NL9 (Mcl-1), and 3MK8

(Mcl-1). In the case of NMR structures with multiple models, the first model was used.

Results for **Figure 8** were described for fifteen bromodomains comprising our test set: 2OSS

(first bromodomain of BRD4), 2OO1 (second bromodomain of BRD3), 3HMH (first bromodomain of

BRD3), 2OUO (second bromodomain of BRD4), 2RFJ (first bromodomain of BRDT), 3HMH (second

bromodomain of TAF1L), 2YYN (TIF1), 3DAI (ATAD2), 2GRC (SMARCA4), 3G0L (BAZ2B, 2F6N

(FALZ), 3GG3 (PCAF), 3IU5 (first bromodomain of PB1), 2RO1 (TRIM28), and 3G0J (fifth

bromodomain of PB1). Comparisons were to 3MXF (the inhibitor-bound structure of the first

bromodomain of BRD4). Data were collected for 5,000 independent simulations for each Bromodomain

at the homologous target residues for the BRD4-JQ1 bound structure (residues 94 and 146 in 3MXF).

Building ensembles of pocket-containing conformations

Ensembles of conformations were generated using the "relax" protocol in Rosetta with the pocket

biasing potential as follows:

```
relax.linuxgccrelease -s input_pdb -relax:fast -pocket_max_spacing 12

-pocket_zero_derivatives -pocket_psp false -pocket_sps -pocket_num_angles 2

-ex1 -ex1aro -ex2 -score:patch pocket.wts.patch -nstruct 1

-cst_fa_file constraints
```

The file pocket.wts.patch simply contains the text "pocket_constraint 1.0", while the file

constraints contains "Pocket 0.25 123:A,129:A", where the target residues are defined as residues 123 and

129 in chain A of the PDB file.

Output structures were then subjected to a local energy minimization in the absence of the pocket

biasing potential:

```
minimize.linuxgccrelease -s input_pdb
```

Conformations with very small pockets led to exemplars with a small number of atoms: exemplars with fewer than ten atoms were also removed from consideration. We also removed from consideration any conformations with energy more than 15 Rosetta Energy Units from those sampled in an analogous set of unbiased simulations.

Generating "exemplars" to represent pockets

Exemplar generation occurs in the "pocket" protocol within Rosetta, and begins by defining the "deep pocket" volume as described previously [14].

The ideal locations of hydrogen bond acceptors (per the Rosetta energy function) are then identified. This is accomplished by identifying all solvent exposed hydrogen bond donors, then calculating the location of an ideal hydrogen bond acceptor for each: an ideal acceptor lies 2.75 Å from the donor atom, on the vector that passes through the donor "base" atom and the hydrogen atom. Acceptor locations that touch the target pocket and do not clash with the protein are included in the exemplar.

Ideal hydrogen bond donors are then identified by iterating over all solvent exposed hydrogen bond acceptors: an ideal donor lies 2.75 Å from the acceptor atom, though its location also depends on the bond angles, Φ and Θ, of the local coordinate system defined by the acceptor atom (origin), its base atom (setting the direction of the X axis) and the hybridization of the accepting group [S1]. For each accepting group, potential donor locations are tested in search of one that touches the target pocket and does not clash with the protein; if found, this donor location is added to the exemplar.

Next, the deep pocket points are scanned to identify points whose 26 neighboring pocket grid points and three next-nearest neighbors in the positive X, Y, and Z directions are all deep pocket points not yet included in the exemplar. When such a point is found, that point is included in the exemplar as a carbon atom, and 26 neighboring grid points are excluded from further searching.

Exemplar atoms within 3.5 Å of one another are then clustered together; any clusters that are not in direct contact with the "target" residue used to define the deep pocket are removed.

Finally, the exemplar is printed to a file using PDB format.

118

Treatment of donors/acceptors in ROCS

The ROCS software is built to use "real" compounds as input, rather than exemplars. Accordingly, ROCS assigns hydrogen bond donors/acceptors upon reading in the template molecule. Oxygen and nitrogen atoms are identified by ROCS as either acceptors or donors (or both) based on their connectivity.

Since our exemplars lack this connectivity, it was necessary to define new atom types that are forced to serve as a pre-determined donor/acceptor regardless of context. To do this, we edited the beryllium and neon element types in ROCS such that these served as obligate hydrogen bond donors and acceptors, respectively. We then doubled the scale of the radius to allow for near optimal hydrogen bonding. Beryllium and neon were chosen because of their comparable size to nitrogen and oxygen, as well as their paucity in drug-like compounds.

When comparing exemplars, we used the modified chemical force field and included chemical complementarity only in the final scoring phase (i.e. only shape considerations were included in the optimization phase).

Comparison of chemical structures
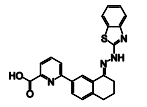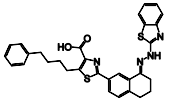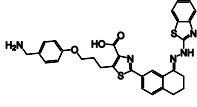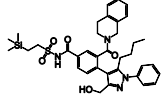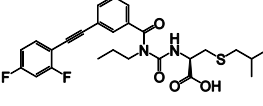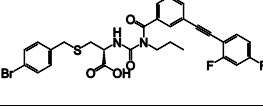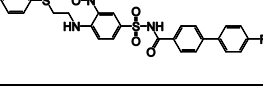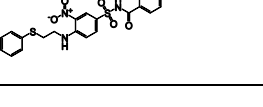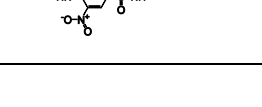
Chemical structures were performed by OpenBabel [S2,S3] using FP2 fingerprints. FP2 fingerprints index the small molecule by identifying linear segments of one to seven atoms that are then indexed in a 1024 bit vector [S2,S3]. Comparisons are scored using the Tanimoto coefficient (Jaccard index): the ratio of number of chemical moieties common to both molecules to the total number of chemical moieties used.

Numerical analysis

Spearman rank correlation coefficients and their significance (**Figure 2**) were also computed in the R statistical computing environment [S4]. Multidimensional scaling analysis was also carried out in

R [S4], using the cmdscale function. The reported statistical significance for the ROC plot (**Figure 7b**) was calculated using a one-tailed P-value with null hypothesis that the area under the curve is 0.5 [S5].

A2.2 Supplementary Tables

Complex	PDB ID	Protein	Inhibitor	Inhibitor common name
1	3zln	Bcl-xL		
2	3zlo	Bcl-xL		
3	3zlr	Bcl-xL		WEHI-539
4	4ehr	Bcl-xL		
5	4c52	Bcl-xL		
6	4c5d	Bcl-xL		
7	1ysi	Bcl-xL		
8 ª	1ysn	Bcl-xL		
9 ᵇ	2o2m	Bcl-xL		

10	2o2n	Bcl-xL		
11	2yxj	Bcl-xL		ABT-737
12	3inq	Bcl-xL		W1191542
13	3qkd	Bcl-xL		
14	3sp7	Bcl-xL		BM-903
15	4ieh	Bcl-2		
16	4lvt	Bcl-2		ABT-263 (navitoclax)
17	4man	Bcl-2		ABT-199
18	4lxd	Bcl-2		
19[b]	2o2f	Bcl-2		

20 [a]	1ysw	Bcl-2		
21	2o22	Bcl-2		
22	4aq3	Bcl-2		
23	2w3l	Bcl-2		
24	4hw2	Mcl-1		
25	4hw3	Mcl-1		
26	3wix	Mcl-1		
27	4oq5	Mcl-1		
28	4oq6	Mcl-1		

Table S1: Structures of complexes used in this study. At this time of writing, this table represents a comprehensive collection of all structures in the PDB containing a Bcl-2 family member in complex with a small-molecule inhibitor bound at the protein interaction site. Fragments (compounds with molecular weight less than 250 Da), molecules whose structure contains interactions with multiple chains that are not part of a biological unit, and molecules with multiple occupancies were excluded from this list. Numbering of compounds in this list corresponds to the order of rows in **Figure 2** and **Figure 7**, and the numbering of compounds in **Figure 4** and **Figure 5a**. Superscripted letters in the leftmost column denote cases in which the same compound has been solved in complex with different protein partners.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1.00	0.65	0.61	0.26	0.16	0.17	0.17	0.31	0.19	0.16	0.16	0.16	0.17	0.22	0.16	0.18	0.19	0.17	0.19	0.31	0.15	0.25	0.25	0.32	0.23	0.34	0.21	0.23
2	0.65	1.00	0.88	0.21	0.12	0.14	0.12	0.28	0.16	0.12	0.13	0.12	0.14	0.19	0.16	0.17	0.16	0.16	0.16	0.28	0.13	0.21	0.21	0.26	0.23	0.28	0.16	0.15
3	0.61	0.88	1.00	0.24	0.14	0.16	0.15	0.31	0.19	0.15	0.15	0.15	0.16	0.20	0.18	0.19	0.21	0.19	0.19	0.31	0.15	0.22	0.25	0.33	0.28	0.35	0.18	0.16
4	0.26	0.21	0.24	1.00	0.23	0.24	0.28	0.31	0.33	0.30	0.32	0.31	0.30	0.32	0.28	0.29	0.34	0.28	0.33	0.31	0.30	0.37	0.44	0.29	0.23	0.34	0.22	0.19
5	0.16	0.12	0.14	0.23	1.00	0.87	0.25	0.19	0.25	0.25	0.24	0.26	0.23	0.21	0.18	0.23	0.22	0.20	0.25	0.19	0.24	0.23	0.24	0.17	0.18	0.16	0.19	0.18
6	0.17	0.14	0.16	0.24	0.87	1.00	0.25	0.20	0.25	0.24	0.24	0.26	0.22	0.22	0.19	0.24	0.23	0.21	0.25	0.20	0.24	0.24	0.24	0.18	0.18	0.17	0.19	0.18
7	0.17	0.12	0.15	0.28	0.25	0.25	1.00	0.64	0.71	0.79	0.71	0.74	0.51	0.50	0.50	0.41	0.49	0.48	0.71	0.64	0.79	0.29	0.22	0.16	0.14	0.20	0.31	0.18
8	0.31	0.28	0.31	0.31	0.19	0.20	0.64	1.00	0.61	0.60	0.56	0.57	0.44	0.46	0.45	0.40	0.46	0.46	0.61	1.00	0.61	0.32	0.29	0.25	0.21	0.26	0.26	0.18
9	0.19	0.16	0.19	0.33	0.25	0.25	0.71	0.61	1.00	0.72	0.70	0.72	0.51	0.52	0.51	0.50	0.53	0.58	1.00	0.61	0.88	0.35	0.27	0.22	0.21	0.22	0.27	0.18
10	0.16	0.12	0.15	0.30	0.25	0.24	0.79	0.60	0.72	1.00	0.88	0.85	0.61	0.56	0.60	0.47	0.60	0.52	0.72	0.60	0.79	0.31	0.24	0.17	0.15	0.20	0.28	0.16
11	0.16	0.13	0.15	0.32	0.24	0.24	0.71	0.56	0.70	0.88	1.00	0.90	0.71	0.65	0.69	0.55	0.65	0.56	0.70	0.56	0.76	0.34	0.26	0.19	0.17	0.20	0.26	0.15
12	0.16	0.12	0.15	0.31	0.26	0.26	0.74	0.57	0.72	0.85	0.90	1.00	0.64	0.62	0.63	0.53	0.59	0.53	0.72	0.57	0.78	0.35	0.26	0.17	0.15	0.19	0.27	0.17
13	0.17	0.14	0.16	0.30	0.23	0.22	0.51	0.44	0.51	0.61	0.71	0.64	1.00	0.65	0.60	0.42	0.52	0.41	0.51	0.44	0.53	0.26	0.26	0.19	0.14	0.20	0.30	0.14
14	0.22	0.19	0.20	0.32	0.21	0.22	0.50	0.46	0.52	0.56	0.65	0.62	0.65	1.00	0.52	0.43	0.48	0.42	0.52	0.46	0.52	0.29	0.25	0.23	0.16	0.26	0.30	0.18
15	0.16	0.16	0.18	0.28	0.18	0.19	0.50	0.45	0.51	0.60	0.69	0.63	0.60	0.52	1.00	0.45	0.50	0.46	0.51	0.45	0.53	0.29	0.25	0.16	0.15	0.16	0.21	0.11
16	0.18	0.17	0.19	0.29	0.23	0.24	0.41	0.40	0.50	0.47	0.55	0.53	0.42	0.43	0.45	1.00	0.47	0.52	0.50	0.40	0.48	0.35	0.27	0.21	0.20	0.18	0.26	0.14
17	0.19	0.16	0.21	0.34	0.22	0.23	0.49	0.46	0.53	0.60	0.65	0.59	0.52	0.48	0.50	0.47	1.00	0.56	0.53	0.46	0.54	0.34	0.30	0.29	0.24	0.27	0.29	0.15
18	0.17	0.16	0.19	0.28	0.20	0.21	0.48	0.46	0.58	0.52	0.56	0.53	0.41	0.42	0.46	0.52	0.56	1.00	0.58	0.46	0.54	0.32	0.26	0.23	0.22	0.19	0.23	0.15
19	0.19	0.16	0.19	0.33	0.25	0.25	0.71	0.61	1.00	0.72	0.70	0.72	0.51	0.52	0.51	0.50	0.53	0.58	1.00	0.61	0.88	0.35	0.27	0.22	0.21	0.22	0.27	0.18
20	0.31	0.28	0.31	0.31	0.19	0.20	0.64	1.00	0.61	0.60	0.56	0.57	0.44	0.46	0.45	0.40	0.46	0.46	0.61	1.00	0.61	0.32	0.29	0.25	0.21	0.26	0.26	0.18
21	0.15	0.13	0.15	0.30	0.24	0.24	0.79	0.61	0.88	0.79	0.76	0.78	0.53	0.52	0.53	0.48	0.54	0.54	0.88	0.61	1.00	0.35	0.25	0.18	0.16	0.17	0.27	0.14
22	0.25	0.21	0.22	0.37	0.23	0.24	0.29	0.32	0.35	0.31	0.34	0.35	0.26	0.29	0.29	0.35	0.34	0.32	0.35	0.32	0.35	1.00	0.36	0.20	0.17	0.20	0.21	0.13
23	0.25	0.21	0.25	0.44	0.24	0.24	0.22	0.29	0.27	0.24	0.26	0.26	0.26	0.25	0.25	0.27	0.30	0.26	0.27	0.29	0.25	0.36	1.00	0.32	0.21	0.28	0.18	0.19
24	0.32	0.26	0.33	0.29	0.17	0.18	0.16	0.25	0.22	0.17	0.19	0.17	0.19	0.23	0.16	0.21	0.29	0.23	0.22	0.25	0.18	0.20	0.32	1.00	0.49	0.55	0.26	0.29
25	0.23	0.23	0.28	0.23	0.18	0.18	0.14	0.21	0.21	0.15	0.17	0.15	0.14	0.16	0.15	0.20	0.24	0.22	0.21	0.21	0.16	0.17	0.21	0.49	1.00	0.32	0.24	0.31
26	0.34	0.28	0.35	0.34	0.16	0.17	0.20	0.26	0.22	0.20	0.20	0.19	0.20	0.26	0.16	0.18	0.27	0.19	0.22	0.26	0.17	0.20	0.28	0.55	0.32	1.00	0.28	0.30
27	0.21	0.16	0.18	0.22	0.19	0.19	0.31	0.26	0.27	0.28	0.26	0.27	0.30	0.30	0.21	0.26	0.29	0.23	0.27	0.26	0.27	0.21	0.18	0.26	0.24	0.28	1.00	0.28
28	0.23	0.15	0.16	0.19	0.18	0.18	0.18	0.18	0.18	0.16	0.15	0.17	0.14	0.18	0.11	0.14	0.15	0.15	0.18	0.18	0.14	0.13	0.19	0.29	0.31	0.30	0.28	1.00

Table S2: 2D Tanimoto similarity of Bcl-2 family inhibitors. This table shows the raw data from which the heatmap in Figure 2a was created.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	2.00	1.45	0.55	0.37	0.52	0.48	0.42	0.42	0.49	0.45	0.45	0.41	0.44	0.38	0.41	0.33	0.39	0.38	0.52	0.44	0.40	0.50	0.50	0.71	0.71	0.72	0.55	0.52
2	1.45	2.00	1.53	0.36	0.46	0.52	0.47	0.40	0.42	0.39	0.38	0.32	0.39	0.39	0.35	0.35	0.32	0.42	0.46	0.46	0.45	0.47	0.41	0.56	0.60	0.52	0.50	0.68
3	0.55	1.53	2.00	0.31	0.46	0.51	0.48	0.42	0.36	0.42	0.37	0.36	0.36	0.45	0.36	0.40	0.39	0.44	0.47	0.53	0.45	0.40	0.38	0.53	0.55	0.45	0.49	0.45
4	0.37	0.36	0.31	2.00	0.35	0.45	0.39	0.33	0.56	0.40	0.46	0.27	0.45	0.23	0.43	0.43	0.36	0.48	0.54	0.39	0.39	0.75	0.49	0.51	0.36	0.35	0.39	0.41
5	0.52	0.46	0.46	0.35	2.00	0.84	0.47	0.42	0.44	0.43	0.40	0.36	0.35	0.33	0.36	0.33	0.37	0.33	0.44	0.41	0.48	0.46	0.46	0.50	0.56	0.46	0.42	0.49
6	0.48	0.52	0.51	0.45	0.84	2.00	0.47	0.45	0.45	0.52	0.41	0.42	0.34	0.38	0.36	0.29	0.32	0.42	0.41	0.55	0.42	0.44	0.42	0.52	0.59	0.49	0.50	0.42
7	0.42	0.47	0.48	0.39	0.47	0.47	2.00	1.23	1.41	0.58	1.10	0.35	0.97	0.73	0.90	1.03	0.44	0.39	1.30	0.54	1.35	0.51	0.47	0.60	0.51	0.41	0.49	0.42
8	0.42	0.40	0.42	0.33	0.42	0.45	1.23	2.00	1.19	0.91	0.83	0.86	0.72	0.62	0.74	0.43	0.62	0.70	0.85	0.89	1.12	0.50	0.40	0.40	0.41	0.54	0.36	0.50
9	0.49	0.42	0.36	0.56	0.44	0.45	1.41	1.19	2.00	1.25	1.15	1.09	1.05	0.82	1.00	1.05	0.77	1.05	1.20	1.13	1.25	0.48	0.35	0.41	0.48	0.54	0.44	0.35
10	0.45	0.39	0.42	0.40	0.43	0.52	0.58	0.91	1.25	2.00	1.11	1.02	1.05	0.88	0.95	1.08	0.85	1.03	1.33	1.10	1.19	0.51	0.34	0.37	0.43	0.45	0.36	0.34
11	0.45	0.38	0.37	0.46	0.40	0.41	1.10	0.83	1.15	1.11	2.00	1.34	1.67	1.03	1.52	1.43	0.86	1.39	1.12	0.89	1.03	0.45	0.40	0.38	0.36	0.45	0.40	0.40
12	0.41	0.32	0.36	0.27	0.36	0.42	0.35	0.86	1.09	1.02	1.34	2.00	1.16	0.91	1.11	1.09	0.86	1.09	0.98	0.79	1.02	0.47	0.34	0.37	0.36	0.42	0.41	0.40
13	0.44	0.39	0.36	0.45	0.35	0.34	0.97	0.72	1.05	1.05	1.67	1.16	2.00	1.02	1.50	1.36	0.96	1.31	1.02	0.81	0.93	0.47	0.43	0.36	0.37	0.39	0.37	0.29
14	0.38	0.39	0.45	0.23	0.33	0.38	0.73	0.62	0.82	0.88	1.03	0.91	1.02	2.00	1.03	0.91	0.77	0.87	0.74	0.74	0.71	0.45	0.29	0.31	0.35	0.39	0.33	0.28
15	0.41	0.35	0.36	0.43	0.36	0.36	0.90	0.74	1.00	0.95	1.52	1.11	1.50	1.03	2.00	1.19	0.99	1.27	0.96	0.83	0.73	0.48	0.38	0.36	0.35	0.41	0.47	0.27
16	0.33	0.35	0.40	0.43	0.33	0.29	1.03	0.43	1.05	1.08	1.43	1.09	1.36	0.91	1.19	2.00	0.82	1.23	1.00	0.83	0.92	0.42	0.30	0.33	0.32	0.41	0.43	0.29
17	0.39	0.32	0.39	0.36	0.37	0.32	0.44	0.62	0.77	0.85	0.86	0.86	0.96	0.77	0.99	0.82	2.00	0.82	0.83	0.77	0.44	0.43	0.32	0.36	0.33	0.43	0.37	0.33
18	0.38	0.42	0.44	0.48	0.33	0.42	0.39	0.70	1.05	1.03	1.39	1.09	1.31	0.87	1.27	1.23	0.82	2.00	1.07	0.91	0.93	0.46	0.34	0.37	0.34	0.46	0.35	0.44
19	0.52	0.46	0.47	0.54	0.44	0.41	1.30	0.85	1.20	1.33	1.12	0.98	1.02	0.74	0.96	1.00	0.83	1.07	2.00	1.06	1.32	0.49	0.43	0.44	0.41	0.50	0.43	0.37
20	0.44	0.46	0.53	0.39	0.41	0.55	0.54	0.89	1.13	1.10	0.89	0.79	0.81	0.74	0.83	0.83	0.77	0.91	1.06	2.00	0.98	0.43	0.29	0.40	0.48	0.48	0.43	0.40
21	0.40	0.45	0.45	0.39	0.48	0.42	1.35	1.12	1.25	1.19	1.03	1.02	0.93	0.71	0.73	0.92	0.44	0.93	1.32	0.98	2.00	0.48	0.46	0.53	0.57	0.53	0.52	0.42
22	0.50	0.47	0.40	0.75	0.46	0.44	0.51	0.50	0.48	0.51	0.45	0.47	0.47	0.45	0.48	0.42	0.43	0.46	0.49	0.43	0.48	2.00	0.50	0.49	0.46	0.39	0.45	0.46
23	0.50	0.41	0.38	0.49	0.46	0.42	0.47	0.40	0.35	0.34	0.40	0.34	0.43	0.29	0.38	0.30	0.32	0.34	0.43	0.29	0.46	0.50	2.00	0.54	0.56	0.49	0.60	0.44
24	0.71	0.56	0.53	0.51	0.50	0.52	0.60	0.40	0.41	0.37	0.38	0.37	0.36	0.31	0.36	0.33	0.36	0.37	0.44	0.40	0.53	0.49	0.54	2.00	1.60	1.12	0.67	0.67
25	0.71	0.60	0.55	0.36	0.56	0.59	0.51	0.41	0.48	0.43	0.36	0.36	0.37	0.35	0.35	0.32	0.33	0.34	0.41	0.48	0.57	0.46	0.56	1.60	2.00	1.16	0.68	0.93
26	0.72	0.52	0.45	0.35	0.46	0.49	0.41	0.54	0.54	0.45	0.45	0.42	0.39	0.39	0.41	0.41	0.43	0.46	0.50	0.48	0.53	0.39	0.49	1.12	1.16	2.00	0.60	0.51
27	0.55	0.50	0.49	0.39	0.42	0.50	0.49	0.36	0.44	0.36	0.40	0.41	0.37	0.33	0.47	0.43	0.37	0.35	0.43	0.43	0.52	0.45	0.60	0.67	0.68	0.60	2.00	0.53
28	0.52	0.68	0.45	0.41	0.49	0.42	0.42	0.50	0.35	0.34	0.40	0.40	0.29	0.28	0.27	0.29	0.33	0.44	0.37	0.40	0.42	0.46	0.44	0.67	0.93	0.51	0.53	2.00

Table S3: 3D Tanimoto (shape) similarity of Bcl-2 family inhibitors. This table shows the raw data from which the heatmap in Figure 2b was created.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	2.00	0.56	0.82	0.50	0.41	0.35	0.26	0.34	0.27	0.27	0.30	0.33	0.25	0.38	0.38	0.40	0.37	0.26	0.26	0.23	0.26	0.30	0.36	0.30	0.34	0.40	0.34	0.36
2	0.56	2.00	0.96	0.59	0.39	0.44	0.46	0.47	0.34	0.56	0.57	0.68	0.61	0.55	0.63	0.52	0.61	0.56	0.49	0.46	0.44	0.36	0.43	0.33	0.35	0.39	0.60	0.46
3	0.82	0.96	2.00	0.69	0.49	0.34	0.40	0.42	0.39	0.52	0.49	0.58	0.51	0.49	0.58	0.50	0.47	0.46	0.47	0.38	0.36	0.34	0.41	0.31	0.28	0.30	0.47	0.38
4	0.50	0.59	0.69	2.00	0.52	0.50	0.37	0.48	0.35	0.43	0.43	0.59	0.59	0.63	0.61	0.66	0.58	0.45	0.44	0.33	0.31	0.56	0.42	0.41	0.38	0.35	0.33	0.30
5	0.41	0.39	0.49	0.52	2.00	0.89	0.40	0.63	0.51	0.35	0.37	0.54	0.39	0.38	0.43	0.40	0.40	0.36	0.46	0.45	0.36	0.54	0.51	0.43	0.54	0.58	0.69	0.67
6	0.35	0.44	0.34	0.50	0.89	2.00	0.33	0.40	0.44	0.31	0.39	0.53	0.35	0.57	0.42	0.50	0.43	0.38	0.44	0.35	0.33	0.40	0.43	0.32	0.40	0.41	0.43	0.41
7	0.26	0.46	0.40	0.37	0.40	0.33	2.00	0.63	0.53	0.61	0.64	0.66	0.61	0.60	0.65	0.61	0.65	0.60	0.78	0.62	0.63	0.47	0.42	0.34	0.41	0.32	0.52	0.49
8	0.34	0.47	0.42	0.48	0.63	0.40	0.63	2.00	0.53	0.51	0.54	0.69	0.61	0.48	0.57	0.57	0.65	0.50	0.58	0.50	0.69	0.46	0.45	0.41	0.47	0.39	0.56	0.50
9	0.27	0.34	0.39	0.35	0.51	0.44	0.53	0.53	2.00	0.51	0.69	0.55	0.60	0.50	0.63	0.49	0.49	0.48	0.60	0.60	0.67	0.39	0.39	0.33	0.36	0.34	0.55	0.49
10	0.27	0.56	0.52	0.43	0.35	0.31	0.61	0.51	0.51	2.00	0.73	0.58	0.81	0.50	0.80	0.63	0.74	0.74	0.69	0.57	0.69	0.36	0.35	0.20	0.34	0.24	0.44	0.37
11	0.30	0.57	0.49	0.43	0.37	0.39	0.64	0.54	0.69	0.73	2.00	0.74	1.29	0.77	1.02	0.84	0.98	0.84	0.76	0.76	0.72	0.37	0.37	0.26	0.31	0.25	0.47	0.35
12	0.33	0.68	0.58	0.59	0.54	0.53	0.66	0.69	0.55	0.58	0.74	2.00	0.71	0.72	0.75	0.68	0.72	0.57	0.75	0.53	0.55	0.49	0.50	0.33	0.40	0.39	0.63	0.52
13	0.25	0.61	0.51	0.59	0.39	0.35	0.61	0.61	0.60	0.81	1.29	0.71	2.00	0.66	0.98	0.87	0.95	0.93	0.72	0.66	0.80	0.35	0.39	0.32	0.29	0.29	0.37	0.30
14	0.38	0.55	0.49	0.63	0.38	0.57	0.60	0.48	0.50	0.50	0.77	0.72	0.66	2.00	0.69	0.75	0.68	0.59	0.56	0.55	0.49	0.35	0.47	0.26	0.39	0.29	0.52	0.41
15	0.38	0.63	0.58	0.61	0.43	0.42	0.65	0.57	0.63	0.80	1.02	0.75	0.98	0.69	2.00	0.87	1.11	0.86	0.75	0.64	0.63	0.32	0.39	0.27	0.38	0.27	0.45	0.38
16	0.40	0.52	0.50	0.66	0.40	0.50	0.61	0.57	0.49	0.63	0.84	0.68	0.87	0.75	0.87	2.00	0.86	0.80	0.64	0.42	0.47	0.46	0.39	0.34	0.37	0.34	0.53	0.36
17	0.37	0.61	0.47	0.58	0.40	0.43	0.65	0.65	0.49	0.74	0.98	0.72	0.95	0.68	1.11	0.86	2.00	0.88	0.73	0.59	0.59	0.48	0.42	0.35	0.41	0.35	0.56	0.49
18	0.26	0.56	0.46	0.45	0.36	0.38	0.60	0.50	0.48	0.74	0.84	0.57	0.93	0.59	0.86	0.80	0.88	2.00	0.65	0.48	0.65	0.39	0.38	0.29	0.31	0.33	0.49	0.42
19	0.26	0.49	0.47	0.44	0.46	0.44	0.78	0.58	0.60	0.69	0.76	0.75	0.72	0.56	0.75	0.64	0.73	0.65	2.00	0.62	0.77	0.41	0.39	0.30	0.38	0.28	0.57	0.45
20	0.23	0.46	0.38	0.33	0.45	0.35	0.62	0.50	0.60	0.57	0.76	0.53	0.66	0.55	0.64	0.42	0.59	0.48	0.62	2.00	0.65	0.42	0.45	0.39	0.41	0.33	0.55	0.54
21	0.26	0.44	0.36	0.31	0.36	0.33	0.63	0.69	0.67	0.69	0.72	0.55	0.80	0.49	0.63	0.47	0.59	0.65	0.77	0.65	2.00	0.43	0.36	0.29	0.36	0.30	0.50	0.45
22	0.30	0.36	0.34	0.56	0.54	0.40	0.47	0.46	0.39	0.36	0.37	0.49	0.35	0.35	0.32	0.46	0.48	0.39	0.41	0.42	0.43	2.00	0.46	0.45	0.41	0.39	0.64	0.53
23	0.36	0.43	0.41	0.42	0.51	0.43	0.42	0.45	0.39	0.35	0.37	0.50	0.39	0.47	0.39	0.39	0.42	0.38	0.39	0.45	0.36	0.46	2.00	0.45	0.44	0.46	0.64	0.62
24	0.30	0.33	0.31	0.41	0.43	0.32	0.34	0.41	0.33	0.20	0.26	0.33	0.32	0.26	0.27	0.34	0.35	0.29	0.30	0.39	0.29	0.45	0.45	2.00	0.95	0.45	0.58	0.46
25	0.34	0.35	0.28	0.38	0.54	0.40	0.41	0.47	0.36	0.34	0.31	0.40	0.29	0.39	0.38	0.37	0.41	0.31	0.38	0.41	0.36	0.41	0.44	0.95	2.00	0.45	0.68	0.61
26	0.40	0.39	0.30	0.35	0.58	0.41	0.32	0.39	0.34	0.24	0.25	0.39	0.29	0.29	0.27	0.34	0.35	0.33	0.28	0.33	0.30	0.39	0.46	0.45	0.45	2.00	0.62	0.69
27	0.34	0.60	0.47	0.33	0.69	0.43	0.52	0.56	0.55	0.44	0.47	0.63	0.37	0.52	0.45	0.53	0.56	0.49	0.57	0.55	0.50	0.64	0.64	0.58	0.68	0.62	2.00	1.04
28	0.36	0.46	0.38	0.30	0.67	0.41	0.49	0.50	0.49	0.37	0.35	0.52	0.30	0.41	0.38	0.36	0.49	0.42	0.45	0.54	0.45	0.53	0.62	0.46	0.61	0.69	1.04	2.00

Table S4: Exemplar similarity of Bcl-2 family inhibitor bound structures. This table shows the raw data from which the heatmap in Figure 2c was created.

Complex	Bcl-xL	Bcl-2	Mcl-1	Bcl-w	Bax	Bid	Ced-9
1	0.58	0.55	0.49	0.45	0.55	0.48	0.51
2	0.75	0.74	0.52	0.51	0.58	0.55	0.59
3	0.65	0.64	0.52	0.41	0.51	0.55	0.57
4	0.67	0.72	0.51	0.42	0.53	0.52	0.60
5	0.88	0.68	0.75	0.56	0.66	0.71	0.79
6	0.85	0.70	0.51	0.45	0.49	0.56	0.54
7	0.67	0.64	0.54	0.61	0.71	0.67	0.51
8	0.65	0.59	0.54	0.49	0.58	0.56	0.63
9	0.61	0.60	0.47	0.50	0.63	0.58	0.54
10	0.67	0.69	0.43	0.55	0.58	0.57	0.46
11	0.71	0.72	0.44	0.57	0.62	0.61	0.45
12	0.81	0.73	0.53	0.54	0.61	0.62	0.66
13	0.69	0.72	0.43	0.56	0.66	0.59	0.44
14	0.70	0.60	0.48	0.55	0.61	0.61	0.52
15	0.74	0.74	0.47	0.57	0.68	0.68	0.48
16	0.70	0.65	0.57	0.58	0.62	0.59	0.54
17	0.87	0.73	0.54	0.59	0.71	0.70	0.51
18	0.78	0.74	0.57	0.59	0.70	0.65	0.50
19	0.76	0.78	0.52	0.66	0.74	0.71	0.56
20	0.58	0.63	0.46	0.54	0.63	0.63	0.54
21	0.59	0.59	0.49	0.51	0.57	0.57	0.49
22	0.61	0.53	0.57	0.42	0.59	0.59	0.56
23	0.67	0.77	0.66	0.64	0.64	0.64	0.77
24	0.47	0.63	0.62	0.58	0.59	0.61	0.64
25	0.48	0.63	0.78	0.61	0.69	0.66	0.61
26	0.50	0.71	0.83	0.61	0.79	0.75	0.68
27	0.74	0.90	0.98	0.91	1.11	1.04	0.93
28	0.61	0.76	0.84	0.71	1.12	0.87	0.81

Table S5: Exemplar similarity of top (closest) pocket optimized structures to inhibitor-bound structures.

This table shows the raw data from which Figure 7 and Figure S8 were created.

Complex	Bcl-xL	Bcl-2	Mcl-1	Bcl-w	Bax	Bid	Ced-9
1	0.65	0.65	0.69	0.51	0.67	0.65	0.60
2	0.68	0.52	0.50	0.47	0.56	0.64	0.54
3	0.60	0.54	0.47	0.47	0.54	0.62	0.53
4	0.52	0.46	0.46	0.48	0.49	0.52	0.40
5	0.65	0.63	0.51	0.52	0.56	0.61	0.63
6	0.60	0.52	0.50	0.51	0.53	0.57	0.51
7	0.64	0.53	0.52	0.50	0.54	0.57	0.49
8	0.54	0.49	0.43	0.48	0.51	0.50	0.45
9	0.59	0.56	0.45	0.51	0.53	0.54	0.45
10	0.53	0.52	0.43	0.51	0.50	0.56	0.40
11	0.52	0.45	0.42	0.47	0.44	0.50	0.38
12	0.52	0.45	0.42	0.44	0.46	0.49	0.40
13	0.51	0.46	0.41	0.49	0.45	0.48	0.37
14	0.49	0.46	0.41	0.46	0.43	0.46	0.32
15	0.52	0.44	0.42	0.49	0.45	0.49	0.37
16	0.48	0.43	0.43	0.48	0.39	0.46	0.37
17	0.52	0.50	0.44	0.53	0.44	0.48	0.39
18	0.52	0.51	0.44	0.47	0.45	0.47	0.41
19	0.61	0.58	0.47	0.53	0.54	0.55	0.45
20	0.53	0.49	0.41	0.50	0.47	0.51	0.42
21	0.64	0.61	0.50	0.55	0.59	0.58	0.52
22	0.58	0.52	0.45	0.54	0.52	0.57	0.48
23	0.46	0.52	0.46	0.46	0.59	0.60	0.43
24	0.60	0.65	0.69	0.65	0.66	0.67	0.55
25	0.57	0.63	0.68	0.66	0.66	0.66	0.58
26	0.54	0.58	0.62	0.58	0.59	0.58	0.43
27	0.52	0.57	0.57	0.50	0.62	0.57	0.40
28	0.62	0.87	0.82	0.78	0.83	0.85	0.79

Table S6: Exemplar similarity of top (closest) pocket optimized structures to native inhibitor conformer.

This table shows the raw data from which Figure S9 was created.

Compound	Bcl-xL	Bcl-2	Mcl-1	Bcl-w	Bax	Bid	Ced-9
1	*A*	*A*	*A*	*A*			
2	*A*						
3	*A*	*A*	*A*	*A*			
4	*B*	*B*					
5	*C*	*C*	*C*	*C*			
6	*C*	*C*	*C*	*C*			
7	*D*	*D*					
8							
9							
10							
11	*D*	*D*	*D*	*D*			
12	*E*	*E*	*E*	*E*			
13	*F*	*F*	*F*	*F*	*S*		
14	*G*	*G*	*G*				
15		*H*					
16	*I*	*I*	*I*	*I*			
17	*I*	*I*	*I*	*I*			
18	*I*	*I*					
19							
20							
21	*J*	*J*					
22	*K*	*K*					
23	*L*	*L*					
24	*M*	*M*	*M*				
25	*M*	*M*	*M*				
26	*N*		*N*				
27							
28							

Table S7: Citations for experimental data for ligand selectivity. This table points to the sources of experimental data from which **Figures 7, S8, and S9** were created. Citations are as follows:

A: Lessene G, Czabotar PE, Sleebs BE, Zobel K, Lowes KN, et al. (2013) Structure-guided design of a selective BCL-X(L) inhibitor. Nat Chem Biol 9: 390-397.

B: Schroeder GM, Wei D, Banfi P, Cai ZW, Lippy J, et al. (2012) Pyrazole and pyrimidine

phenylacylsulfonamides as dual Bcl-2/Bcl-xL antagonists. Bioorg Med Chem Lett 22: 3951-3956.

C: Brady RM, Vom A, Roy MJ, Toovey N, Smith BJ, et al. (2014) De-novo designed library of

benzoylureas as inhibitors of BCL-XL: synthesis, structural and biochemical characterization.

J Med Chem 57: 1323-1343.

D: Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, et al. (2005) An inhibitor of

Bcl-2 family proteins induces regression of solid tumours. Nature 435: 677-681.

E: Lee EF, Czabotar PE, Yang H, Sleebs BE, Lessene G, et al. (2009) Conformational changes in Bcl-2

pro-survival proteins determine their capacity to bind ligands. J Biol Chem 284: 30508-30517.

F: Sleebs BE, Czabotar PE, Fairbrother WJ, Fairlie WD, Flygare JA, et al. (2011) Quinazoline

sulfonamides as dual binders of the proteins B-cell lymphoma 2 and B-cell lymphoma extra long with

potent proapoptotic cell-based activity. J Med Chem 54: 1914-1926.

G: Zhou H, Aguilar A, Chen J, Bai L, Liu L, et al. (2012) Structure-based design of potent Bcl-2/Bcl-xL

inhibitors with strong in vivo antitumor activity. J Med Chem 55: 6149-6161.

H: Touré BB, Miller-Moslin K, Yusuff N, Perez L, Doré M, et al. (2013) The Role of the Acidity of N-

Heteroaryl Sulfonamides as Inhibitors of Bcl-2 Family Protein–Protein Interactions. ACS Med Chem Lett

4: 186-190.

I: Petros AM, Dinges J, Augeri DJ, Baumeister SA, Betebenner DA, et al. (2006) Discovery of a potent

inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. J Med Chem 49: 656-663.

J: Bruncko M, Oost TK, Belli BA, Ding H, Joseph MK, et al. (2007) Studies leading to potent, dual

inhibitors of Bcl-2 and Bcl-xL. J Med Chem 50: 641-662.

K: Perez HL, Banfi P, Bertrand J, Cai ZW, Grebinski JW, et al. (2012) Identification of a

phenylacylsulfonamide series of dual Bcl-2/Bcl-xL antagonists. Bioorg Med Chem Lett 22: 3946-3950.

L: Porter J, Payne A, de Candole B, Ford D, Hutchinson B, et al. (2009) Tetrahydroisoquinoline amide

substituted phenyl pyrazoles as selective Bcl-2 inhibitors. Bioorg Med Chem Lett 19: 230-233.

M: Friberg A, Vigil D, Zhao B, Daniels RN, Burke JP, et al. (2013) Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. J Med Chem 56: 15-30.

N: Tanaka Y, Aikawa K, Nishida G, Homma M, Sogabe S, et al. (2013) Discovery of potent Mcl-1/Bcl-xL dual inhibitors by using a hybridization strategy based on structural analysis of target proteins. J Med Chem 56: 9635-9645.

Active	
BRD4(1)	0.83
BRD3(2)	0.84
BRD3(1)	0.95
BRD4(2)	0.70
BRDT(1)	0.65
Inactive	
TAF1L(2)	0.57
TIF1	0.64
ATAD2	0.56
SMARCA4	0.52
BAZ2B	0.55
FALZ	0.50
PCAF	0.55
PB1(1)	0.52
TRIM28	0.48
PB1(5)	0.48

Table S8: Exemplar similarity of top (closest) pocket optimized structures to (+)-JQ1-BRD4(1) bound structure. This table shows the raw data from which the heatmap in **Figure 8** was created.

A2.3 Supplementary Figures
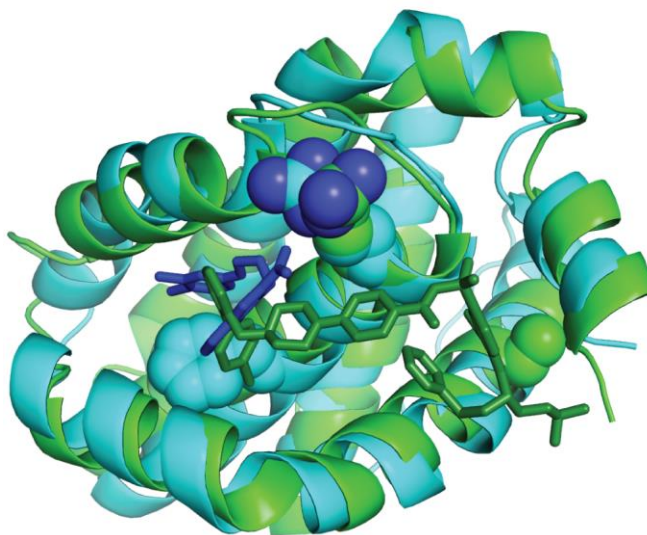
Figure S1: Effect of orientation on exemplar comparison. Bcl-xL bound to ABT-737 (PDB ID 2yxj), a compound similar to ABT-737 (PDB ID 3qkd), an unrelated Bcl-xL inhibitor (PDB ID 3zln), and Mcl-1 bound to an Mcl-1 inhibitor (PDB ID 4hw2) were rotated for 100 random orientations, and an all-vs-all comparison of the ensemble of ABT-737 bound exemplars to those generated for each orientation was performed. Histograms of the 3D Tanimoto scores of the ABT-737 exemplars compared to themselves (*black*), a closely-related compound (*green*), an unrelated compound (*orange*), and an Mcl-1 pocket (*red*) demonstrate overlap between similar and dissimilar pockets.
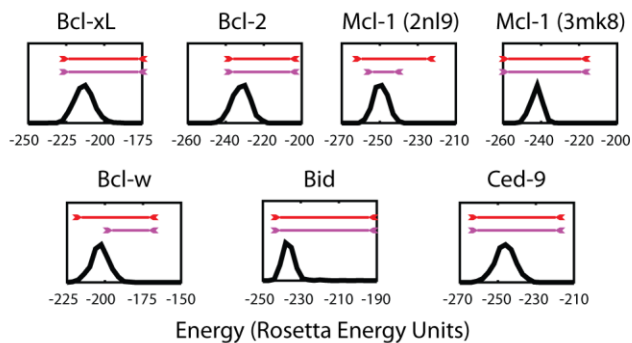
Figure S2: Bcl-2 family members recognize different inhibitors using distinct surface pockets. In all cases color gradient indicates similarity between experimentally-derived structures, expressed as Z-scores (*green* are most similar, *red* are most dissimilar). Numbering in all cases corresponds to complexes in Table S1. (A) Chemical similarity of the inhibitors. (B) Three-dimensional similarity of the inhibitors' active conformation. (C) Similarity of protein surface pockets, quantified through exemplar similarity. There is a statistically significant correlation between each pair of panels, as described in the text. The underlying raw data used to generate these plots are included in Tables S2, S3, and S4.
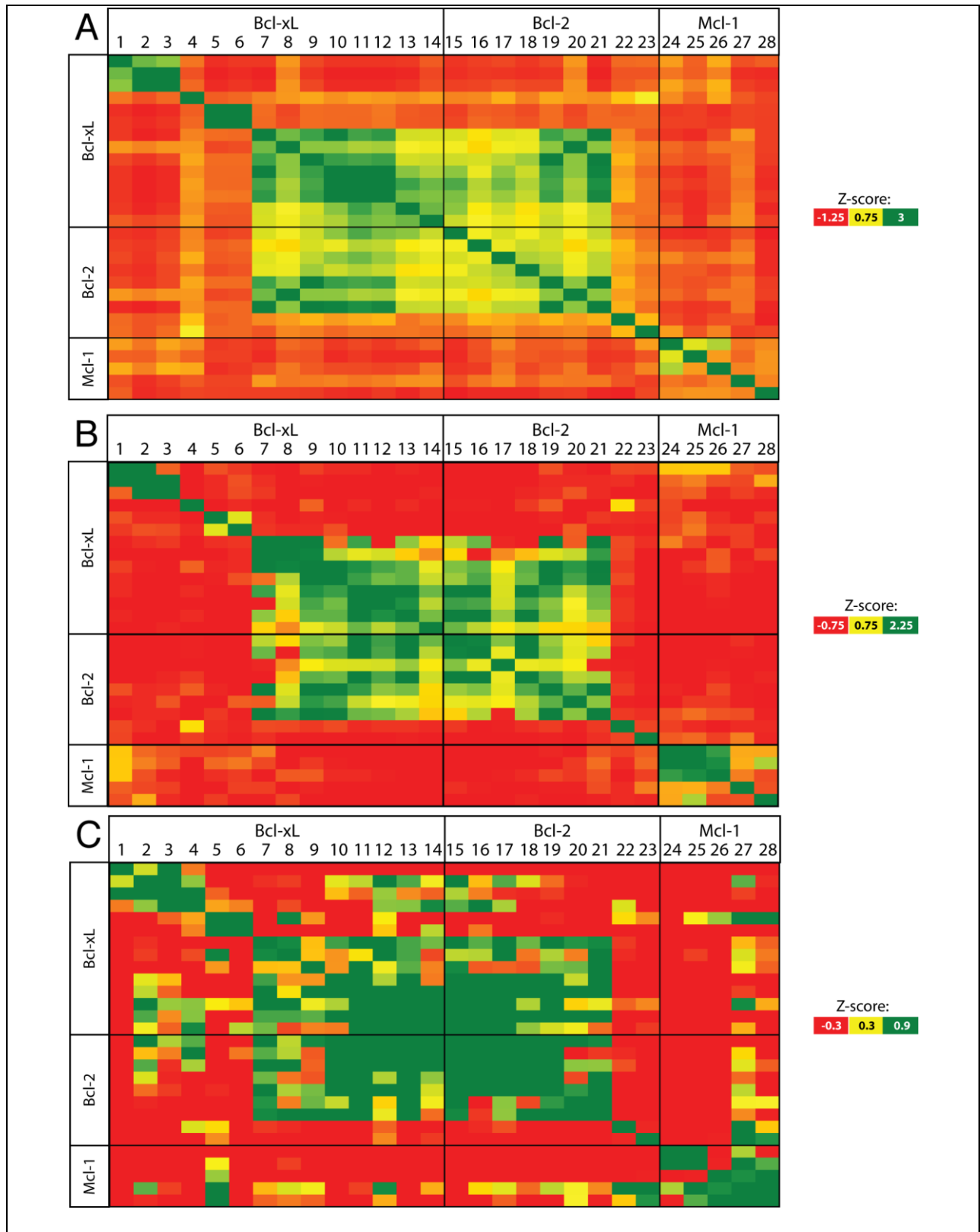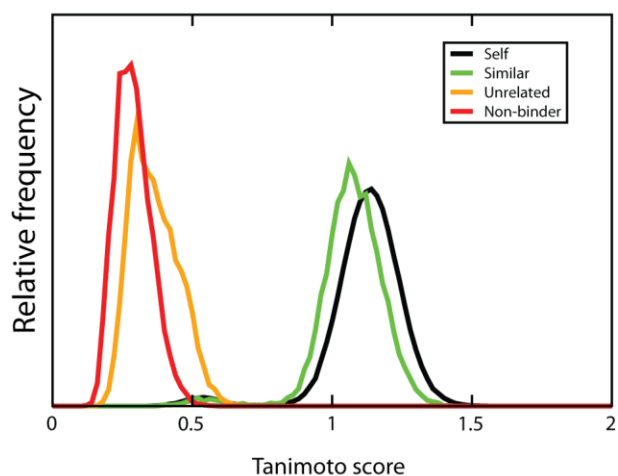
Figure S4: Location of target residues. A superposition of representative inhibitor-bound structures of Bcl-xL (*light green cartoon, inhibitor in dark green sticks*) and Mcl-1 (*cyan cartoon, inhibitor in dark blue sticks*) is shown. The target residues for pocket opening generated from the corresponding peptide-bound structures differed for Bcl-xL (Ala93 and Arg143) and for Mcl-1 (Arg263 and Phe270) (*sidechains shown using spacefill*), reflecting different energetic contributions to the protein-protein interaction from each protein surface.
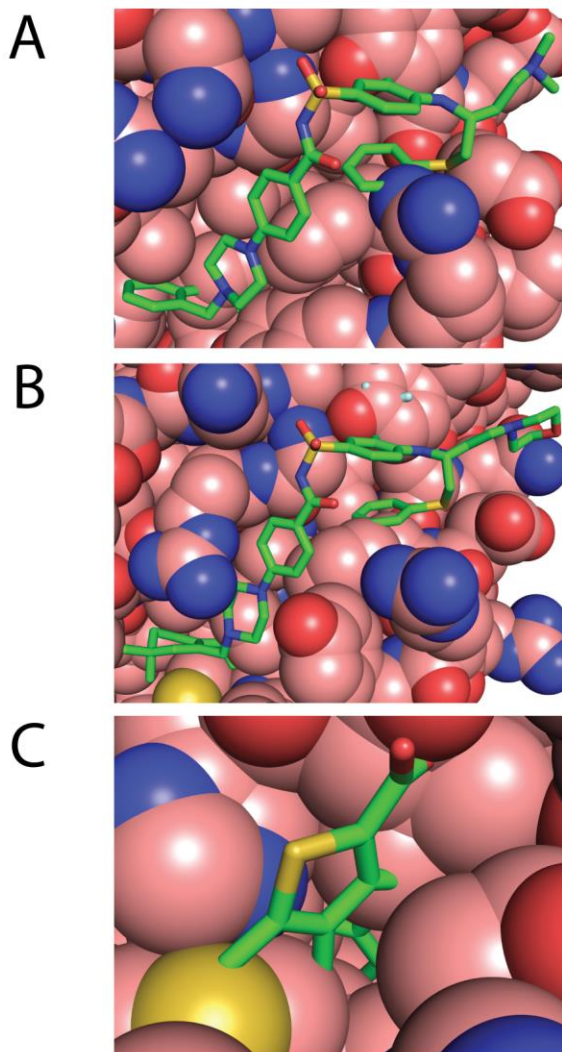
Figure S5: Energetic analysis of conformations generated starting from a peptide bound structure.
For each peptide-bound complex used in this analysis, the histograms of energies of conformations
generated without the use of a biasing potential (*black line*) overlap with the range of energies of the

conformations used in all subsequent analyses generated by using a biasing potential at the Bcl-xL protein

interaction site (*magenta line*) and the Mcl-1 protein interaction site (*red line*); this suggests that many of

these conformations are energetically accessible to these proteins under physiological conditions. All

energies shown here were evaluated in the absence of the biasing potential, for fair comparison.
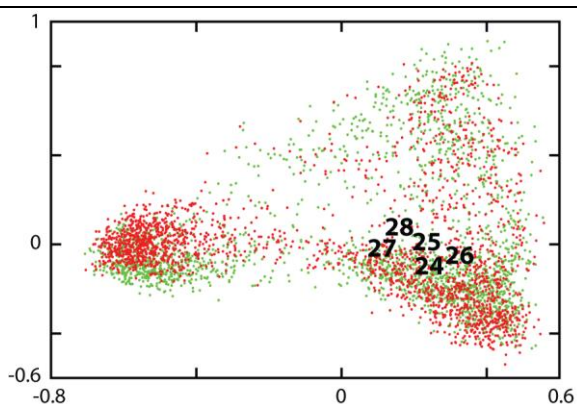
Figure S6: Comparison of pocket space for human and mouse Mcl-1. Constructs of Mcl-1 were made from peptide-bound (2nl9 and 3mk8) and unbound (1wsx) structures of Mcl-1 with human (*red*) and mouse (*green*) sequences. Individual conformations are represented as points on a two-dimensional projection that best reflects the pairwise distances between their exemplars.
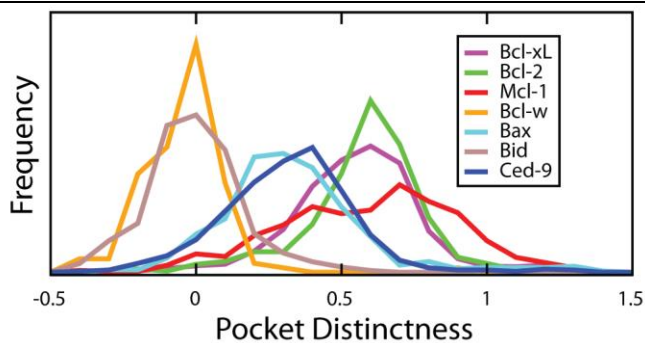

Figure S7: "Distinctness" of pockets in ensembles generated by simulation of Bcl-2 family members. The "distinctness" of a pocket is defined as the exemplar distance of the difference in exemplar distances of the closest conformation from a different family member, and the closest conformation from the ensemble of this family member.
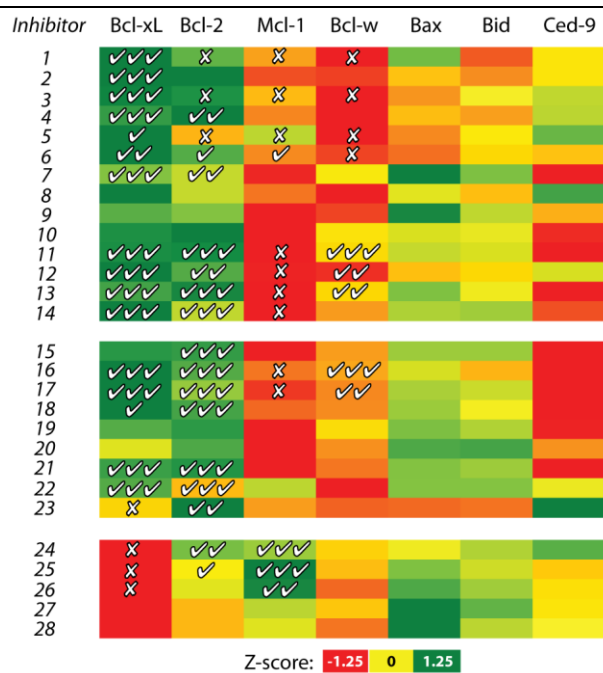
Figure S8: Pocket shape similarity explains ligand selectivity across the Bcl-2 family. For each inhibitor-bound pocket, the exemplar distance to the most similar pocket is indicated by color gradient expressed as Z-scores (*green* are most similar, *red* are most dissimilar; the range of colors for each row is normalized across that row). Markings inside the cells denote experimental measurements of binding for this protein-ligand pair (Kd or Ki where available, otherwise IC50): ✓✓✓ indicates < 0.1 μM, ✓✓ indicates 0.1-1 μM, ✓ indicates binding weaker than 1 μM, and ✗ indicates that binding was not detectable/quantifiable. Cells that do not include markings correspond to protein-ligand pairs for which binding data has not been reported. Numbering corresponds to complexes as in Table S1. The underlying raw data are included in Table S5 (exemplar distances) and Table S7 (citations to binding data).

Figure S9: Pocket shape versus ligand shape similarity explains ligand selectivity across the Bcl-2 family**.** As opposed to Figure S8, in which we compared conformations from simulation to exemplars from inhibitor-bound pockets, here we compare conformations from simulation directly to the structures of the inhibitors. As previously, the exemplar distance to the most similar pocket is indicated by color gradient expressed as Z-scores (*green* are most similar, *red* are most dissimilar; the range of colors for each row is normalized across that row). Markings inside the cells denote experimental measurements of binding for this protein-ligand pair (Kd or Ki where available, otherwise IC50): ✔✔✔ indicates < 0.1 μM, ✔✔ indicates 0.1-1 μM,, ✔ indicates binding weaker than 1 μM, and ✗ indicates that binding was not detectable/quantifiable. Cells that do not include markings correspond to protein-ligand pairs for which binding data has not been reported. Numbering corresponds to complexes as in Table S1. The underlying raw data are included in Table S6 (exemplar distances) and Table S7 (citations to binding data)**.**
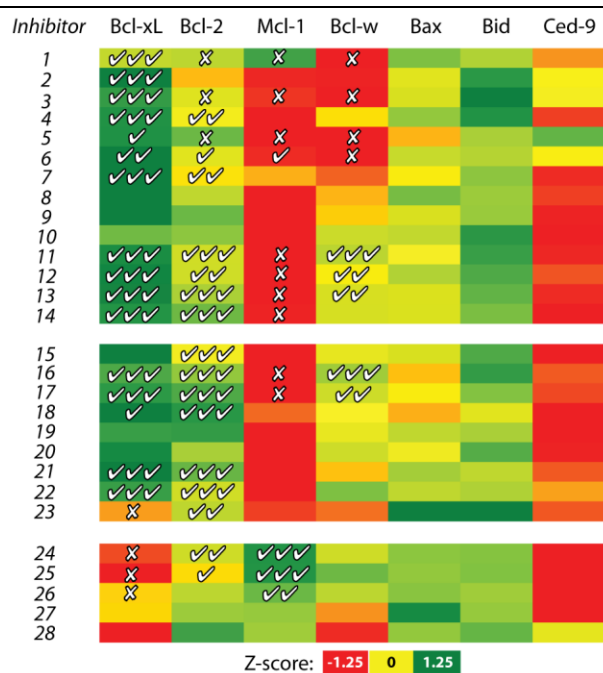
A3 Supporting Information for Chapter 3

A3.1 Supplementary Methods

Rosetta command-lines

All Rosetta calculations were carried out using the developer trunk source code, git revision fe2036f5f991f7c48221ee8916e38da6b28fc949.

Ensembles of conformations were generated using the "relax" protocol in Rosetta, with the pocket-biasing potential, as follows:

```
relax.linuxgccrelease -s input_pdb -relax:fast -nstruct 1 -pocket_max_spacing 12

       -pocket_zero_derivatives -pocket_psp false -pocket_sps -pocket_num_angles 2

       -ex1 -ex1aro -ex2 -score:patch pocket.wts.patch -cst_fa_file constraints
```

The file "pocket.wts.patch" simply contains the text "pocket_constraint 1.0". The file "constraints" contains the text "Pocket 0.25 123:A,129:A" (to indicate target residues 123 and 129 of chain A in the input PDB). It takes about 5 minutes on a single core to generate each conformation; jobs are trivially parallel.

Output structures were then subjected to a local energy minimization in the absence of the pocket biasing term:

```
minimize.linuxgccrelease -s input_pdb
```

In order to limit the conformations to physiologically relevant pocket containing conformations, we only analyzed pockets from conformations within 15 Rosetta energy units from the highest energy output structure in the unbiased ensemble.

The Rosetta command to generate an exemplar is as follows:

```
make_exemplar.linuxgccrelease -s input_pdb -pocket_grid_size 12

        -pocket_filter_by_exemplar --pocket_static_grid -central_relax_pdb_num 108
```

Scoring in ROCS / FastROCS

Because the ROCS software is designed to read "real" compounds as input (rather than exemplars), ROCS automatically assigns hydrogen bond donors/acceptors based on the connectivity within the template molecule. Since our exemplars lack connectivity between atoms, we therefore needed to explicitly define hydrogen bond donors/acceptors, regardless of context.

To do so, we edited the "color force field" used by ROCS, re-defining the beryllium and neon element types to map to hydrogen bond donors and acceptors, respectively. We also doubled the scale of the radius to accommodate near optimal hydrogen bonding. Beryllium and neon were selected due to their comparable van der Waals radii with nitrogen and oxygen, as well as their infrequent occurrence in drug-like compounds.

When comparing exemplars to small molecules, we used this modified chemical force field, and considered only shape in the optimization (overlay) phase. We then included the complete force field (i.e. including chemical complementarity) in the final scoring phase.

A patch file to create the exemplar color force field (exemplar.cff) from ImplicitMillsDean.cff (included with ROCS) is pasted below. We recommended copying ImplicitMillsDean.cff into a separate

directory, renaming it to exemplar.cff, and then applying this patch (to avoid overwriting the original file).

--- ImplicitMillsDean.cff   2015-09-13 19:53:25.324724972 -0500

+++ exemplar.cff 2015-09-13 19:54:12.418808086 -0500

@@ -51,6 +51,7 @@

 #

 # acceptors

 #

+DEFINE AcceptorAt [Ne]

 DEFINE ACamine [N;!$(N*=[!#6]);!$(N~[!#6;!#1]);!$(Na);!$(N#*);!$(N=*)]

 DEFINE ACphosphate [O;$hd1;$(O~P(~O)~O)]

 DEFINE ACcarboxylate [O;$hd1;$(O[C;!$(*N)]=O),$(O=[C;!$(*N)][O;$hd1])]

@@ -79,7 +80,7 @@

 DEFINE AChet5O [o;X2;$(o1cccc1),$(o1ccccc1);!$(*[#6]=O)]

 DEFINE ACsulphone [O;$(O=[S;$(S(~O)(~O)([#6,#7])[#6])])]

 # strong acceptors

-DEFINE strongAcceptor [$ACphosphate,$ACcarboxylate,$ACwater,$AChet6N,$ACphosphinyl]

+DEFINE strongAcceptor [$AcceptorAt,$ACphosphate,$ACcarboxylate,$ACwater,$AChet6N,$ACphosphinyl]

 # moderate acceptors

 #DEFINE moderateAcceptor [$ACsulphoxide,$ACprimaryAmine,$AChet5N,$ACthiocarbonyl,$AChydroxyl,$ACsulphate,$ACtertiaryAmine,$ACamide,$ACcarbamate,$ACurea]

DEFINE moderateAcceptor

[$ACsulphoxide,$AChet5N,$ACthiocarbonyl,$AChydroxyl,$ACsulphate,$ACamide,$ACcarbamate,$ACurea]

@@ -89,6 +90,7 @@

 #

 # donors

 #

+DEFINE DonorAt [Be]

 DEFINE Damine [N;!$(N*=[!#6]);!$(N~[!#6;!#1]);!$(Na);!$(N#*);!$(N=*)]

 DEFINE Dhet5NH [nH;$(n1aaaa1),$(n1aaaaa1)]

 DEFINE DNpH [NH,H2,H3;+]

@@ -107,7 +109,7 @@

 DEFINE DsecondaryAmine [$Damine;$hd2]

 # strong donors

 #DEFINE strongDonor [$Dhet5NH,$DNpH,$DacidOH,$Dhydroxyl]

-DEFINE strongDonor [$Dhet5NH,$DNpH,$Dhydroxyl]

+DEFINE strongDonor [$DonorAt,$Dhet5NH,$DNpH,$Dhydroxyl]

 # moderate donors

 DEFINE moderateDonor [$Dwater,$DprimaryAmide,$DanilineNH,$DamidineNH,$DsecondaryAmide,$DanilineNH2]

 # weak donors

@@ -184,12 +186,12 @@

 #######################################

 #   Interaction Definitions   #

148

###

-INTERACTION donor donor attractive gaussian weight=1.0 radius=1.0

-INTERACTION acceptor acceptor attractive gaussian weight=1.0 radius=1.0

+INTERACTION donor donor attractive gaussian weight=1.0 radius=2.0

+INTERACTION acceptor acceptor attractive gaussian weight=1.0 radius=2.0

 INTERACTION rings rings attractive gaussian weight=1.0 radius=1.0

 INTERACTION cation cation attractive gaussian weight=1.0 radius=1.0

 INTERACTION anion anion attractive gaussian weight=1.0 radius=1.0

-INTERACTION hydrophobe hydrophobe attractive gaussian weight=1.0 radius=1.0

+INTERACTION hydrophobe hydrophobe attractive gaussian weight=1.0 radius=2.0

###

Unused Types

###

Rapid overlays using FastROCS

FastROCS is composed of two Python scripts. One script, ShapeDatabaseServer.py, reads in the database, performs the overlay, and outputs the top overlays. The other script, ShapeDatabaseClient.py, reads in the query compounds (in our case, the exemplars), and sends them to the ShapeDatabaseServer. Because the command line interface is very limited, we modified the ShapeDatabaseServer.py script in order to use the modified chemical force field described above, and to return more than just the highest scoring conformer when returning the hits.

A patch file to create the ShapeDatabaseExemplarServer.py from the ShapeDatabaseServer.py that is included with FastROCS is pasted below. The location to exemplar.cff needs to be changed to

match the location of the patched exemplar.cff file, as created in the section above. We recommended copying ShapeDatabaseServer.py to ShapeDatabaseExemplarServer.py, and then applying this patch (to avoid overwriting the original file).

```
--- ShapeDatabaseServer.py        2015-09-13 19:37:12.160003637 -0500

+++ ShapeDatabaseExemplarServer.py        2015-09-13 19:43:54.344715620 -0500

@@ -87,6 +87,10 @@

    self.options = OEShapeDatabaseOptions()

    self.options.SetTracer(self.tracer)

    self.options.SetLimit(nhits)

+      self.cff = OEColorForceField()

+      self.cff.Init("/home/labs/karanicolaslab/exemplar.cff")

+      self.options.SetColorForceField(self.cff)

+      self.options.SetMaxConfs(10)

    self.lock = Lock()


  def run(self):
```

The commands to run FastROCS are as follows:

```
ShapeDatabaseExemplarServer.py /path/to/db/dbfile1.oeb.gz

ShapeDatabaseClient.py localhost:8080 /path/to/exemplar.pdb overlays.sdf 2000
```

Such a command would load one dbfile1 from a database, and the Client would read in the send the exemplar, specify the output file to be named overlays.sdf, and request the top 2000 hits. In order to limit the time spent writing data for the lower ranking compounds to disk, FastROCS does not output a score file, and the scores included in the output SDF file do not include the Tversky scores.

In order to re-rank by Tversky scores, we performed a secondary call to standard ROCS with the "-dbase" flag, which quickly recalculates the scores:

```
rocs -dbase overlays.sdf -query /path/to/exemplar.pdb –report one –scoreonly

      -nostructs -stats all -chemff /path/to/exemplar.cff -prefix rescore
```

PDB structures used in calculations

Unbound structures (including those used as starting points for "pocket optimization" simulations) in **Figures 3-6** were from PDB IDs: 1R2D (Bcl-xL) and 1F9X (XIAP). Small-molecule bound structures used in **Figures 2**, **3**, and **5** were from PDB IDs 2YXJ (Bcl-xL) and 2JK7 (XIAP).

Unbound structures used for "pocket optimization" simulations in **Table 1** and **Table S3** were from the following PDB IDs: 1M47 (IL-2), 1R6K (HPV E2), 1F9X (XIAP), 1R2D (Bcl-xL), 2OSS (BRD4), 1GHU (Grb2-SH2), 2GNQ (WDR5), 3ZRF (VHL), 1BIS (HIV Integrase), 1Z1M (Mdm2), and 3U84 (Menin). Small-molecule bound structures used in **Table S2** were from the following PDB IDs: 1PW6 (IL-2), 1R6N (HPV E2), 1TFT (XIAP), 1YSI (Bcl-xL), 2YEL (BRD4), 3IN7 (Grb2-SH2), (WDR5). 3ZRC (VHL), 4E1N (HIV Integrase), 4ERF (Mdm2), and 4GQ4 (Menin).

In the case of NMR structures, the first model was used. HET atoms were removed from all PDB files, and only one biological unit was retained.

Generating Exemplars for 1R2D and 1F9X

Because the unbound structures contain no small molecule binding pocket, extra flags were used to generate a suitable exemplar for screening. The flag "-pocket_surface_dist" indicates how much of the top layer of pocket residues to remove when finding the "deep volume" (the default value is 2.5 Å).

1R2D: -pocket_surface_dist 1.0

1F9X: -pocket_surface_dist 1.5

Generating Exemplars for 1TFT, 1Y2F, 2JK7, 3IN7, 3VKX, 3ZRC, AND 4GQ4

Because the small molecule binding pocket in these structures were small, extra flags were used to generate a suitable exemplar for screening. The flag "-pocket_surface_dist" indicates how much of the top layer of pocket residues to remove when finding the "deep volume" (the default value is 2.5 Å). The flag "-pocket_grid_size" indicates how far (in Å) in the X,Y,Z dimensions to extend the grid from the target residue(s) (the default value is 12 Å). Lower values, together with lower values of pocket_surface_dist, allow for better descriptions of small, shallow pockets.

1TFT: -pocket_surface_dist 0.5 -pocket_grid_size 10

1Y2F: -pocket_surface_dist 1.0 -pocket_grid_size 6

2JK7: -pocket_surface_dist 0.5 -pocket_grid_size 10

3IN7: -pocket_surface_dist 1.5 -pocket_grid_size 6

3VKX: -pocket_surface_dist 1.0 -pocket_grid_size 6

3ZRC: -pocket_surface_dist 1.0 -pocket_grid_size 6

4GQ4: -pocket_surface_dist 1.5 -pocket_grid_size 6

A3.2 Supplementary Tables

Protein target	Unique inhibitors	Unbound PDB ID	Inhibitor-bound PDB ID	Protein-bound PDB ID	Target Residue(s) (Resi:Chain)
HIV-1 integrase	87	1BIS	4E1N	3WNE	169:A,124:B
Mdm2	34	1Z1M	4ERF	1YCR	54:A,62:A
Bcl-xL	25	1R2D	1YSI	1BXL	93:A,143:A
BRD4	20	2OSS	2YEL	2MJV	385:B,439:B
XIAP-BIR3	12	1F9X	1TFT	1G73	306:D,308:D
Grb-SH2	9	3IN7	3WA4	1GHU	67:A,109:A
Interleukin-2	6	1M47	1PW6	1Z92	42:A
HPV E2	4	1R6K	1R6N	1TUE	40:A
Menin	3	3U84	4GQ4	4I80	319:A
WDR5	3	2GNQ	3UR4	4CY2	88:A,133:A
VHL	1	3ZRF	3ZRC	4AJY	98:V,112:V

Table S1: Protein targets used in the large-scale screening benchmark (Table 1, Tables S2 and S3).

We note that the active compounds for Bcl-xL and XIAP differ slightly here relative to our previous benchmark (used in Figures 3-6), because we updated these at the same time as collecting inhibitors for each of the other target proteins in our set. The target residues' chain and residue numbers refer to the numbering scheme in the protein complex, and were then adjusted accordingly for the inhibitor-bound and unbound protein structures. These target residues were used for "pocket optimization" simulations, and also for identifying pockets to generate exemplars.

Protein name	Best rank of known active	Number of known actives in top 100	Fraction of known actives in top 1%
HIV-1 integrase (87)	1	1	0.01
Mdm2 (34)	1	3	0.09
Bcl-xL (25)	1	7	0.28
BRD4 (20)	NR	0	0.00
XIAP-BIR3 (12)	NR	0	0.00
Grb2-SH2 (9)	NR	0	0.00
Interleukin-2 (6)	NR	0	0.00
HPV E2 (4)	NR	0	0.00
Menin (3)	NR	0	0.00
WDR5 (3)	NR	0	0.00
VHL (1)	NR	0	0.00

Table S2: Virtual screening for a variety of protein-protein interaction targets, using an inhibitor-bound protein conformation. This screen is analogous to that presented in Table 1. NR indicates that no active compounds were in the top 0.75% from the initial FastROCS ranking, and therefore no active compounds were included when re-ranking with ROCS.

Protein name	Best rank of known active	Number of known actives in top 100	Fraction of known actives in top 1%
HIV-1 integrase (87)	1	2	0.03
Mdm2 (34)	NR	0	0.00
Bcl-xL (25)	NR	0	0.00
BRD4 (20)	NR	0	0.00
XIAP-BIR3 (12)	NR	0	0.00
Grb2-SH2 (9)	4420	0	0.11
Interleukin-2 (6)	NR	0	0.00
HPV E2 (4)	6534	0	0.25
Menin (3)	132	0	0.33
WDR5 (3)	NR	0	0.00
VHL (1)	NR	0	0.00

Table S3: Virtual screening for a variety of protein-protein interaction targets, using the single top-scoring pocket-optimized conformation. NR indicates that no active compounds were in the top 0.75% from the initial FastROCS ranking, and therefore no active compounds were included when re-ranking with ROCS.

REFERENCES

1.   Sun H*, et al.* (2008) Structure-based design, synthesis, evaluation, and crystallographic studies of conformationally constrained Smac mimetics as inhibitors of the X-linked inhibitor of apoptosis protein (XIAP). *J. Med. Chem.* 51(22):7169-7180.

2.   Bruncko M*, et al.* (2007) Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J. Med. Chem.* 50(4):641-662.

3.   Rush TS, 3rd, Grant JA, Mosyak L, & Nicholls A (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *Journal of medicinal chemistry* 48(5):1489-1495.

4.   Wang Y*, et al.* (2004) Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor. *J. Biol. Chem.* 279(8):6976-6985.

5.   Thanos CD, Randal M, & Wells JA (2003) Potent small-molecule binding to a dynamic hot spot on IL-2. *J. Am. Chem. Soc.* 125(50):15280-15281.

6.   Becker JW*, et al.* (1999) 32-Indolyl ether derivatives of ascomycin: three-dimensional structures of complexes with FK506-binding protein. *J. Med. Chem.* 42(15):2798-2804.

7.   Vassilev LT*, et al.* (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303(5659):844-848.

8.   Wells JA & McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450(7172):1001-1009.

9.   Alsmadi O, Herz R, Murphy E, Pinter A, & Tilley SA (1997) A novel antibody-dependent cellular cytotoxicity epitope in gp120 is identified by two monoclonal antibodies isolated from a long-term survivor of human immunodeficiency virus type 1 infection. *J Virol* 71(2):925-933.

10.   Domling A (2008) Small molecular weight protein-protein interaction antagonists: an insurmountable challenge? *Curr Opin Chem Biol* 12(3):281-291.

11.   Rechfeld F, Gruber P, Hofmann J, & Kirchmair J (2011) Modulators of protein-protein

interactions: novel approaches in targeting protein kinases and other pharmaceutically relevant

biomolecules. *Current topics in medicinal chemistry* 11(11):1305-1319.

12.   Zinzalla G & Thurston DE (2009) Targeting protein-protein interactions for therapeutic

intervention: a challenge for the future. *Future medicinal chemistry* 1(1):65-93.

13.   Kenny CH*, et al.* (2003) Development of a fluorescence polarization assay to screen for inhibitors

of the FtsZ/ZipA interaction. *Anal Biochem* 323(2):224-233.

14.   Hajduk PJ, Huth JR, & Fesik SW (2005) Druggability indices for protein targets derived from

NMR-based screening data. *Journal of medicinal chemistry* 48(7):2518-2525.

15.   Mattos C*, et al.* (2006) Multiple solvent crystal structures: probing binding sites, plasticity and

hydration. *J Mol Biol* 357(5):1471-1482.

16.   Hu Y & Bajorath J (2013) Compound promiscuity: what can we learn from current data? *Drug

Discov. Today* 18(13-14):644-650.

17.   Peters JU (2013) Polypharmacology - foe or friend? *J. Med. Chem.* 56(22):8955-8971.

18.   Roberts AW*, et al.* (2012) Substantial susceptibility of chronic lymphocytic leukemia to BCL2

inhibition: results of a phase I study of navitoclax in patients with relapsed or refractory disease.

J Clin Oncol 30(5):488-496.

19.   Wilson WH*, et al.* (2010) Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid

malignancies: a phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics,

and antitumour activity. *Lancet Oncol* 11(12):1149-1159.

20.   Liu L*, et al.* (2012) Structure-based design of novel class II c-Met inhibitors: 2. SAR and kinase

selectivity profiles of the pyrazolone series. *J. Med. Chem.* 55(5):1868-1897.

21.   Filippakopoulos P*, et al.* (2010) Selective inhibition of BET bromodomains. *Nature*

468(7327):1067-1073.

22.   Friberg A*, et al.* (2013) Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J. Med. Chem.* 56(1):15-30.

23.   Kurumbail RG*, et al.* (1996) Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* 384(6610):644-648.

24.   Wermuth CG, Ganellin CR, Lindberg P, & Mitscher LA (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry* 70(5):1129-1143.

25.   Bohacek R, Boosalis MS, McMartin C, Faller DV, & Perrine SP (2006) Identification of novel small-molecule inducers of fetal hemoglobin using pharmacophore and 'PSEUDO' receptor models. *Chemical biology & drug design* 67(5):318-328.

26.   Cushman M, Insaf S, Ruell JA, Schaeffer CA, & Rice WG (1998) Synthesis of a cosalane analog with an extended polyanionic pharmacophore conferring enhanced potency as an anti-HIV agent. *Bioorg. Med. Chem. Lett.* 8(7):833-836.

27.   Wender PA, Koehler KF, Sharkey NA, Dell'Aquila ML, & Blumberg PM (1986) Analysis of the phorbol ester pharmacophore on protein kinase C as a guide to the rational design of new classes of analogs. *Proceedings of the National Academy of Sciences of the United States of America* 83(12):4214-4218.

28.   Garland R M, C. David B, Heinz E B, Richard A D, & Deborah A D (1979) The Conformational Parameter in Drug Design: The Active Analog Approach. *Computer-Assisted Drug Design,* ACS Symposium Series, (AMERICAN CHEMICAL SOCIETY), Vol 112, pp 205-226.

29.   Wolber G & Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of chemical information and modeling* 45(1):160-169.

30.   Koes DR & Camacho CJ (2012) PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic acids research* 40(Web Server issue):W387-392.

31.   Koes D, *et al.* (2012) Enabling large-scale design, synthesis and validation of small molecule

        protein-protein antagonists. *PloS one* 7(3):e32839.

32.   Metz A, Schanda J, Grez M, Wichmann C, & Gohlke H (2013) From determinants of RUNX1/ETO

        tetramerization to small-molecule protein-protein interaction inhibitors targeting acute myeloid

        leukemia. *Journal of chemical information and modeling* 53(9):2197-2202.

33.   De Luca L, *et al.* (2009) Pharmacophore-based discovery of small-molecule inhibitors of protein-

        protein interactions between HIV-1 integrase and cellular cofactor LEDGF/p75. *ChemMedChem*

        4(8):1311-1316.

34.   Christ F, *et al.* (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase

        interaction and HIV replication. *Nat. Chem. Biol.* 6(6):442-448.

35.   Mustata G, *et al.* (2011) Development of small-molecule PUMA inhibitors for mitigating

        radiation-induced cell death. *Curr. Top. Med. Chem.* 11(3):281-290.

36.   Corradi V, *et al.* (2011) Computational techniques are valuable tools for the discovery of protein-

        protein interaction inhibitors: the 14-3-3sigma case. *Bioorg. Med. Chem. Lett.* 21(22):6867-6871.

37.   Voet A, *et al.* (2011) Structure based discovery of small molecule suppressors targeting bacterial

        lysozyme inhibitors. *Biochem. Biophys. Res. Commun.* 405(4):527-532.

38.   Metz A, *et al.* (2012) Hot spots and transient pockets: predicting the determinants of small-

        molecule binding to a protein-protein interface. *J Chem Inf Model* 52(1):120-133.

39.   Seco J, Luque FJ, & Barril X (2009) Binding site detection and druggability index from first

        principles. *Journal of medicinal chemistry* 52(8):2363-2371.

40.   Eyrisch S & Helms V (2007) Transient pockets on protein surfaces involved in protein-protein

        interaction. *J. Med. Chem.* 50(15):3457-3464.

41.   Laurie AT & Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of

        protein-ligand binding sites. *Bioinformatics* 21:1908-1916.

42.   Foster TJ, MacKerell AD, Jr., & Guvench O (2012) Balancing target flexibility and target

        denaturation in computational fragment-based inhibitor discovery. *Journal of computational*

        chemistry 33(23):1880-1891.

43.   Clodfelter KH, Waxman DJ, & Vajda S (2006) Computational solvent mapping reveals the

        importance of local conformational changes for broad substrate specificity in mammalian

        cytochromes P450. *Biochemistry* 45(31):9393-9407.

44.   Landon MR*, et al.* (2008) Novel druggable hot spots in avian influenza neuraminidase H5N1

        revealed by computational solvent mapping of a reduced and representative receptor

        ensemble. *Chem Biol Drug Des* 71(2):106-116.

45.   Landon MR, Lancia DR, Jr., Yu J, Thiel SC, & Vajda S (2007) Identification of hot spots within

        druggable binding regions by computational solvent mapping of proteins. *Journal of medicinal*

        chemistry 50(6):1231-1240.

46.   Prasad JC, Goldstone JV, Camacho CJ, Vajda S, & Stegeman JJ (2007) Ensemble modeling of

        substrate binding to cytochromes P450: analysis of catalytic differences between CYP1A

        orthologs. *Biochemistry* 46(10):2640-2654.

47.   Eyrisch S & Helms V (2009) What induces pocket openings on protein surface patches involved

        in protein-protein interactions? *J Comput Aided Mol Des* 23(2):73-86.

48.   Bakan A, Nevins N, Lakdawala AS, & Bahar I (2012) Druggability Assessment of Allosteric

        Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J Chem Theory Comput*

        8(7):2435-2447.

49.   Kozakov D*, et al.* (2011) Structural conservation of druggable hot spots in protein-protein

        interfaces. *Proceedings of the National Academy of Sciences of the United States of America*

        108(33):13528-13533.

50.   Huang B & Schroeder M (2006) LIGSITEcsc: predicting ligand binding sites using the Connolly

        surface and degree of conservation. *BMC Struct Biol* 6:19.

51.   Brady GP, Jr. & Stouten PF (2000) Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 14(4):383-401.

52.   Hendlich M, Rippmann F, & Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* 15(6):359-363, 389.

53.   Liang J, Edelsbrunner H, & Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7(9):1884-1897.

54.   Peters KP, Fauck J, & Frommel C (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J Mol Biol* 256(1):201-213.

55.   Weisel M, Proschak E, & Schneider G (2007) PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chemistry Central journal* 1:7.

56.   Leaver-Fay A*, et al.* (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487:545-574.

57.   Velculescu VE*, et al.* (1999) Analysis of human transcriptomes. *Nat Genet* 23(4):387-388.

58.   Paik S*, et al.* (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27):2817-2826.

59.   Rodel F*, et al.* (2008) Survivin antisense oligonucleotides effectively radiosensitize colorectal cancer cells in both tissue culture and murine xenograft models. *Int J Radiat Oncol Biol Phys* 71(1):247-255.

60.   Lewis KD*, et al.* (2011) A multi-center phase II evaluation of the small molecule survivin suppressor YM155 in patients with unresectable stage III or IV melanoma. *Investigational new drugs* 29(1):161-166.

61.   Kita A*, et al.* (2011) Antitumor effects of YM155, a novel survivin suppressant, against human aggressive non-Hodgkin lymphoma. *Leuk Res* 35(6):787-792.

62.   Altieri DC (2008) Survivin, cancer networks and pathway-directed drug discovery. *Nat Rev Cancer* 8(1):61-70.

63.   Sun H*, et al.* (2010) Nonpeptidic and potent small-molecule inhibitors of cIAP-1/2 and XIAP proteins. *J Med Chem* 53(17):6361-6367.

64.   Dean EJ*, et al.* (2010) A small molecule inhibitor of XIAP induces apoptosis and synergises with vinorelbine and cisplatin in NSCLC. *Br J Cancer* 102(1):97-103.

65.   Wang Z*, et al.* (2004) Cellular, biochemical, and genetic analysis of mechanism of small molecule IAP inhibitors. *J Biol Chem* 279(46):48168-48176.

66.   Schimmer AD*, et al.* (2004) Small-molecule antagonists of apoptosis suppressor XIAP exhibit broad antitumor activity. *Cancer cell* 5(1):25-35.

67.   Wendt MD*, et al.* (2007) Discovery of a novel small molecule binding site of human survivin. *Bioorg Med Chem Lett* 17(11):3122-3129.

68.   Cheng AC*, et al.* (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* 25(1):71-75.

69.   Quintus F, Sperandio O, Grynberg J, Petitjean M, & Tuffery P (2009) Ligand scaffold hopping combining 3D maximal substructure search and molecular similarity. *BMC Bioinformatics* 10:245.

70.   Shuker SB, Hajduk PJ, Meadows RP, & Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274(5292):1531-1534.

71.   Qian B*, et al.* (2007) High-resolution structure prediction and the crystallographic phase problem. *Nature* 450(7167):259-264.

72.   Reddy EP & Aggarwal AK (2012) The ins and outs of bcr-abl inhibition. *Genes Cancer* 3(5-6):447-454.

73.   Socinski MA (2011) Multitargeted receptor tyrosine kinase inhibition: an antiangiogenic strategy in non-small cell lung cancer. *Cancer Treat Rev* 37(8):611-617.

74.   Miyamoto S, Miyake N, Jarskog LF, Fleischhacker WW, & Lieberman JA (2012) Pharmacological treatment of schizophrenia: a critical review of the pharmacology and clinical effects of current and future therapeutic agents. *Mol. Psychiatry* 17(12):1206-1227.

75.   Arkin MR*, et al.* (2003) Binding of small molecules to an adaptive protein-protein interface. *Proceedings of the National Academy of Sciences of the United States of America* 100(4):1603-1608.

76.   Johnson DK & Karanicolas J (2013) Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Comput Biol* 9(3):e1002951.

77.   Boehr DD, Nussinov R, & Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5(11):789-796.

78.   Moldoveanu T, Follis AV, Kriwacki RW, & Green DR (2014) Many players in BCL-2 family affairs. *Trends Biochem. Sci.* 39(3):101-111.

79.   Adams JM & Cory S (2007) The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* 26(9):1324-1337.

80.   Levitt DG & Banaszak LJ (1992) POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J Mol Graph* 10(4):229-234.

81.   Li B*, et al.* (2008) Characterization of local geometry of protein surfaces with the visibility criterion. *Proteins* 71(2):670-683.

82.   Le Guilloux V, Schmidtke P, & Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 10:168.

83.   Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 13(5):323-330, 307-328.

84.   Durrant JD, de Oliveira CA, & McCammon JA (2011) POVME: an algorithm for measuring binding-pocket volumes. *J Mol Graph Model* 29(5):773-776.

85. Kortemme T & Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* 99(22):14116-14121.

86. Kortemme T, Kim DE, & Baker D (2004) Computational alanine scanning of protein-protein interfaces. *Science's STKE : signal transduction knowledge environment* 2004(219):pl2.

87. Barnum D, Greene J, Smellie A, & Sprague P (1996) Identification of common functional configurations among molecules. *Journal of chemical information and computer sciences* 36(3):563-571.

88. Chen E*, et al.* (2014) Computation-guided discovery of influenza endonuclease inhibitors. *ACS Med Chem Lett* 5(1):61-64.

89. Fry DC (2012) Small-molecule inhibitors of protein-protein interactions: how to mimic a protein partner. *Curr. Pharm. Des.* 18(30):4679-4684.

90. Voet A, Banwell EF, Sahu KK, Heddle JG, & Zhang KY (2013) Protein interface pharmacophore mapping tools for small molecule protein: protein interaction inhibitor discovery. *Curr. Top. Med. Chem.* 13(9):989-1001.

91. Tintori C, Corradi V, Magnani M, Manetti F, & Botta M (2008) Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *Journal of chemical information and modeling* 48(11):2166-2179.

92. Barillari C, Marcou G, & Rognan D (2008) Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *Journal of chemical information and modeling* 48(7):1396-1410.

93. Hu B & Lill MA (2013) Exploring the potential of protein-based pharmacophore models in ligand pose prediction and ranking. *Journal of chemical information and modeling* 53(5):1179-1190.

94. Masek BB, Merchant A, & Matthew JB (1993) Molecular shape comparison of angiotensin II receptor antagonists. *J. Med. Chem.* 36(9):1230-1238.

95. Grant JA, Gallardo MA, & Pickup B (1996) A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J Comp Chem* 17(14):1653-1666.

96. OpenEye Scientific Software; ROCS), 3.2.0.3.

97. Abulwerdi F*, et al.* (2014) A novel small-molecule inhibitor of mcl-1 blocks pancreatic cancer growth in vitro and in vivo. *Mol. Cancer Ther.* 13(3):565-575.

98. Mencher SK & Wang LG (2005) Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). *BMC Clin Pharmacol* 5:3.

99. Arnold AA*, et al.* (2008) Preclinical studies of Apogossypolone: a new nonpeptidic pan small-molecule inhibitor of Bcl-2, Bcl-XL and Mcl-1 proteins in Follicular Small Cleaved Cell Lymphoma model. *Mol Cancer* 7:20.

100. Bernardo PH*, et al.* (2010) Structural insights into the design of small molecule inhibitors that selectively antagonize Mcl-1. *J. Med. Chem.* 53(5):2314-2318.

101. Chen J*, et al.* (2012) Structure-based discovery of BM-957 as a potent small-molecule inhibitor of Bcl-2 and Bcl-xL capable of achieving complete tumor regression. *J. Med. Chem.* 55(19):8502-8514.

102. Degterev A*, et al.* (2001) Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat. Cell Biol.* 3(2):173-182.

103. Lin J*, et al.* (2009) A novel Bcl-XL inhibitor Z36 that induces autophagic cell death in Hela cells. *Autophagy* 5(3):314-320.

104. Nguyen M*, et al.* (2007) Small molecule obatoclax (GX15-070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* 104(49):19512-19517.

105. Niino S, Nakamura Y, Hirabayashi Y, Nagano-Ito M, & Ichikawa S (2013) A small molecule inhibitor of Bcl-2, HA14-1, also inhibits ceramide glucosyltransferase. *Biochem. Biophys. Res. Commun.* 433(2):170-174.

106. Petros AM*, et al.* (2014) Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein. *Bioorg. Med. Chem. Lett.* 24(6):1484-1488.

107. Petros AM*, et al.* (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J. Med. Chem.* 49(2):656-663.

108. Petros AM*, et al.* (2010) Discovery of a potent and selective Bcl-2 inhibitor using SAR by NMR. *Bioorg. Med. Chem. Lett.* 20(22):6587-6591.

109. Shoemaker AR*, et al.* (2006) A small-molecule inhibitor of Bcl-XL potentiates the activity of cytotoxic drugs in vitro and in vivo. *Cancer Res.* 66(17):8731-8739.

110. Shore GC & Viallet J (2005) Modulating the bcl-2 family of apoptosis suppressors for potential therapeutic benefit in cancer. *Hematology Am Soc Hematol Educ Program*:226-230.

111. Wang G*, et al.* (2006) Structure-based design of potent small-molecule inhibitors of anti-apoptotic Bcl-2 proteins. *J. Med. Chem.* 49(21):6139-6142.

112. Wang JL*, et al.* (2000) Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7124-7129.

113. Wang Z*, et al.* (2008) TW-37, a small-molecule inhibitor of Bcl-2, inhibits cell growth and invasion in pancreatic cancer. *Int. J. Cancer* 123(4):958-966.

114. Wei J*, et al.* (2009) Apogossypol derivatives as pan-active inhibitors of antiapoptotic B-cell lymphoma/leukemia-2 (Bcl-2) family proteins. *J. Med. Chem.* 52(14):4511-4523.

115. Wei J*, et al.* (2010) BI-97C1, an optically pure Apogossypol derivative as pan-active inhibitor of antiapoptotic B-cell lymphoma/leukemia-2 (Bcl-2) family proteins. *J. Med. Chem.* 53(10):4166-4176.

116.   Lessene G*, et al.* (2013) Structure-guided design of a selective BCL-X(L) inhibitor. *Nat. Chem. Biol.* 9(6):390-397.

117.   Schneider G, Neidhart W, Giller T, & Schmid G (1999) "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew Chem Int Ed Engl* 38(19):2894-2896.

118.   Eyrisch S, Medina-Franco JL, & Helms V (2012) Transient pockets on XIAP-BIR2: toward the characterization of putative binding sites of small-molecule XIAP inhibitors. *J. Mol. Model.* 18(5):2031-2042.

119.   Hodis E*, et al.* (2008) Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome biology* 9(8):R121.

120.   Prilusky J*, et al.* (2011) Proteopedia: a status report on the collaborative, 3D web-encyclopedia of proteins and other biomolecules. *Journal of structural biology* 175(2):244-252.

121.   Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 1):12-21.

122.   Lee EF*, et al.* (2009) Conformational changes in Bcl-2 pro-survival proteins determine their capacity to bind ligands. *J. Biol. Chem.* 284(44):30508-30517.

123.   Oltersdorf T*, et al.* (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435(7042):677-681.

124.   Perez HL*, et al.* (2012) Identification of a phenylacylsulfonamide series of dual Bcl-2/Bcl-xL antagonists. *Bioorg. Med. Chem. Lett.* 22(12):3946-3950.

125.   Porter J*, et al.* (2009) Tetrahydroisoquinoline amide substituted phenyl pyrazoles as selective Bcl-2 inhibitors. *Bioorg. Med. Chem. Lett.* 19(1):230-233.

126.   Schroeder GM*, et al.* (2012) Pyrazole and pyrimidine phenylacylsulfonamides as dual Bcl-2/Bcl-xL antagonists. *Bioorg. Med. Chem. Lett.* 22(12):3951-3956.

127.   Sleebs BE*, et al.* (2011) Quinazoline sulfonamides as dual binders of the proteins B-cell lymphoma 2 and B-cell lymphoma extra long with potent proapoptotic cell-based activity. *J. Med. Chem.* 54(6):1914-1926.

128.   Tanaka Y*, et al.* (2013) Discovery of potent Mcl-1/Bcl-xL dual inhibitors by using a hybridization strategy based on structural analysis of target proteins. *J. Med. Chem.* 56(23):9635-9645.

129.   Touré BB*, et al.* (2013) The Role of the Acidity of N-Heteroaryl Sulfonamides as Inhibitors of Bcl-2 Family Protein–Protein Interactions. *ACS Med. Chem. Lett.* 4(2):186-190.

130.   Zhou H*, et al.* (2012) Structure-based design of potent Bcl-2/Bcl-xL inhibitors with strong in vivo antitumor activity. *J. Med. Chem.* 55(13):6149-6161.

131.   Ehrlich P (1909) *Beiträge zur experimentellen Pathologie und Chemotherapie* (Akademische Verlagsgesellschaft, Leipzig,) pp vii p., 2 l., 3 -247 p.

132.   Guner OF & Bowen JP (2014) Setting the record straight: the origin of the pharmacophore concept. *Journal of chemical information and modeling* 54(5):1269-1283.

133.   Kier LB (1967) Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. *Mol. Pharmacol.* 3(5):487-494.

134.   Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: Dassault Systèmes, 2015.).

135.   Ebalunode JO, Ouyang Z, Liang J, & Zheng W (2008) Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *Journal of chemical information and modeling* 48(4):889-901.

136.   Lee HS, Lee CS, Kim JS, Kim DH, & Choe H (2009) Improving virtual screening performance against conformational variations of receptors by shape matching with ligand binding pocket. *Journal of chemical information and modeling* 49(11):2419-2428.

137.   Richmond NJ*, et al.* (2006) GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. *J Comput Aided Mol Des* 20(9):567-587.

138.   Zhu LL, Hou TJ, Chen LR, & Xu XJ (2001) 3D QSAR analyses of novel tyrosine kinase inhibitors based on pharmacophore alignment. *Journal of chemical information and computer sciences* 41(4):1032-1040.

139.   Gao Q, Yang L, & Zhu Y (2010) Pharmacophore based drug design approach as a practical process in drug discovery. *Current computer-aided drug design* 6(1):37-49.

140.   Palomer A*, et al.* (2002) Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. *J. Med. Chem.* 45(7):1402-1411.

141.   Kraemer O, Hazemann I, Podjarny AD, & Klebe G (2004) Virtual screening for inhibitors of human aldose reductase. *Proteins* 55(4):814-823.

142.   Chen JZ, Wang J, & Xie XQ (2007) GPCR structure-based virtual screening approach for CB2 antagonist search. *Journal of chemical information and modeling* 47(4):1626-1637.

143.   Evers A & Klabunde T (2005) Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* 48(4):1088-1097.

144.   Marriott DP, Dougall IG, Meghani P, Liu YJ, & Flower DR (1999) Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J. Med. Chem.* 42(17):3210-3216.

145.   Ekins S*, et al.* (2005) In vitro and pharmacophore-based discovery of novel hPEPT1 inhibitors. *Pharmaceutical research* 22(4):512-517.

146.   Enyedy IJ, Sakamuri S, Zaman WA, Johnson KM, & Wang S (2003) Pharmacophore-based discovery of substituted pyridines as novel dopamine transporter inhibitors. *Bioorg. Med. Chem. Lett.* 13(3):513-517.

147.   Chang C, Ekins S, Bahadduri P, & Swaan PW (2006) Pharmacophore-based discovery of ligands for drug transporters. *Adv Drug Deliv Rev* 58(12-13):1431-1450.

148.   Fauman EB, Rai BK, & Huang ES (2011) Structure-based druggability assessment--identifying

        suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* 15(4):463-468.

149.   Rush TS, 3rd, Grant JA, Mosyak L, & Nicholls A (2005) A shape-based 3-D scaffold hopping

        method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48(5):1489-

        1495.

150.   Surade S & Blundell TL (2012) Structural biology and drug discovery of difficult targets: the limits

        of ligandability. *Chem. Biol.* 19(1):42-50.

151.   Kuenemann MA*, et al.* (2015) In silico design of low molecular weight protein-protein

        interaction inhibitors: Overall concept and recent advances. *Progress in biophysics and

        molecular biology* 119(1):20-32.

152.   Claussen H, Buning C, Rarey M, & Lengauer T (2001) FlexE: efficient molecular docking

        considering protein structure variations. *J. Mol. Biol.* 308(2):377-395.

153.   Cavasotto CN & Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to

        protein kinases. *J. Mol. Biol.* 337(1):209-225.

154.   Osguthorpe DJ, Sherman W, & Hagler AT (2012) Exploring protein flexibility: incorporating

        structural ensembles from crystal structures and simulation into virtual screening protocols. *The

        journal of physical chemistry. B* 116(23):6952-6959.

155.   Huang SY & Zou X (2007) Ensemble docking of multiple protein structures: considering protein

        structural variations in molecular docking. *Proteins* 66(2):399-421.

156.   Cosconati S*, et al.* (2012) Protein flexibility in virtual screening: the BACE-1 case study. *Journal of

        chemical information and modeling* 52(10):2697-2704.

157.   Cavasotto CN, Kovacs JA, & Abagyan RA (2005) Representing receptor flexibility in ligand docking

        through relevant normal modes. *Journal of the American Chemical Society* 127(26):9632-9640.

158. Dixit A & Verkhivker GM (2012) Integrating ligand-based and protein-centric virtual screening of kinase inhibitors using ensembles of multiple protein kinase genes and conformations. *Journal of chemical information and modeling* 52(10):2501-2515.

159. Cheng LS*, et al.* (2008) Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* 51(13):3878-3894.

160. Ellingson SR, Miao Y, Baudry J, & Smith JC (2015) Multi-conformer ensemble docking to difficult protein targets. *The journal of physical chemistry. B* 119(3):1026-1034.

161. Johnson DK & Karanicolas J (2015) Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations. *PLoS Comput Biol* 11(2):e1004081.

162. Gonder-Frederick LA*, et al.* (2013) Examining the Behaviour subscale of the Hypoglycaemia Fear Survey: an international study. *Diabet Med* 30(5):603-609.

163. Hawkins PC, Skillman AG, & Nicholls A (2007) Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50(1):74-82.

164. Brenke R*, et al.* (2012) Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics* 28(20):2608-2614.

165. Lyskov S*, et al.* (2013) Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PloS one* 8(5):e63906.

166. Das R*, et al.* (2007) Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins* 69 Suppl 8:118-128.

167. Gonder-Frederick L, Rice P, Warren D, Vajda K, & Shepard J (2013) Diabetic alert dogs: a preliminary survey of current users. *Diabetes Care* 36(4):e47.

168. Vajda A*, et al.* (2013) Gene expression analysis in prostate cancer: the importance of the endogenous control. *Prostate* 73(4):382-390.

169. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, & Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank

and Cambridge Structural Database. *Journal of chemical information and modeling* 50(4):572-584.

170.   Higueruelo AP, Jubb H, & Blundell TL (2013) TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database (Oxford)* 2013:bat039.

171.   Higueruelo AP*, et al.* (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chemical biology & drug design* 74(5):457-467.

172.   Bento AP*, et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* 42(Database issue):D1083-1090.

173.   Kalliokoski T, Kramer C, Vulpetti A, & Gedeck P (2013) Comparability of mixed IC(5)(0) data - a statistical analysis. *PloS one* 8(4):e61007.

174.   Horvath D, Marcou G, & Varnek A (2013) Do not hesitate to use Tversky-and other hints for successful active analogue searches with feature count descriptors. *Journal of chemical information and modeling* 53(7):1543-1562.

175.   Gowthaman R*, et al.* (2015) DARC: Mapping Surface Topography by Ray-Casting for Effective Virtual Screening at Protein Interaction Sites. *J. Med. Chem.*:DOI: 10.1021/acs.jmedchem.1025b00150.

176.   Willett P (2003) Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* 31(Pt 3):603-606.

177.   Brozell SR*, et al.* (2012) Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* 26(6):749-773.

178.   Lang PT*, et al.* (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* 15(6):1219-1230.

179.   Morris GM*, et al.* (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* 30(16):2785-2791.

180. McGovern SL & Shoichet BK (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* 46(14):2895-2907.

181. Vajda P, Felinger A, & Guiochon G (2013) Evaluation of surface excess isotherms in liquid chromatography. *J Chromatogr A* 1291:41-47.

182. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, & Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling* 52(7):1757-1768.

183. Truchon JF & Bayly CI (2007) Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *Journal of chemical information and modeling* 47(2):488-508.

184. Bazzoli A, Kelow SP, & Karanicolas J (2015) Enhancements to the Rosetta energy function enable improved identification of small molecules that inhibit protein-protein interactions. *PloS one* 10(10):e0140359.

185. Gowthaman R, Deeds EJ, & Karanicolas J (2013) Structural properties of non-traditional drug targets present new challenges for virtual screening. *J Chem Inf Model* 53(8):2073-2081.

186. Cosconati S*, et al.* (2010) Virtual Screening with AutoDock: Theory and Practice. *Expert opinion on drug discovery* 5(6):597-607.

187. Cerqueira NM*, et al.* (2015) Receptor-based virtual screening protocol for drug discovery. *Archives of biochemistry and biophysics* 582:56-67.

188. Schonherr H & Cernak T (2013) Profound methyl effects in drug discovery and a call for new C-H methylation reactions. *Angewandte Chemie* 52(47):12256-12267.

189. Martin YC, Kofron JL, & Traphagen LM (2002) Do structurally similar molecules have similar biological activity? *J. Med. Chem.* 45(19):4350-4358.

190. Ajay D & Sobhia ME (2011) Simplified receptor based pharmacophore approach to retrieve potent PTP-LAR inhibitors using apoenzyme. *Current computer-aided drug design* 7(3):159-172.

191. Carlson HA*, et al.* (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43(11):2100-2114.

192. Damm KL & Carlson HA (2007) Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* 129(26):8225-8235.

193. Grohmann D*, et al.* (2008) Small molecule inhibitors targeting HIV-1 reverse transcriptase dimerization. *ChemBioChem* 9(6):916-922.

194. Bowman GR & Geissler PL (2012) Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci U S A* 109(29):11681-11686.

195. Gorre ME*, et al.* (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* 293(5531):876-880.

196. Ma B, Shatsky M, Wolfson HJ, & Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.* 11(2):184-197.

197. Xia Y*, et al.* (Submitted) Rationally designing inhibitors of the Musashi protein-RNA interaction by hotspot mimicry.