

Computational Method To Identify Druggable Binding Sites That Target Protein–Protein Interactions

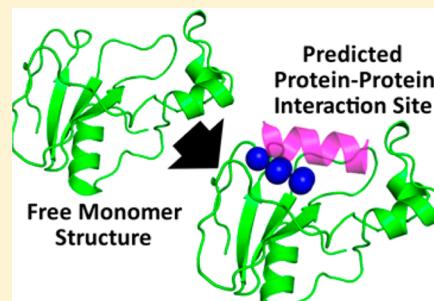
Hubert Li,[†] Vinod Kasam,[†] Christofer S. Tautermann,[‡] Daniel Seeliger,[‡] and Nagarajan Vaidehi*,[†]

[†]Division of Immunology, Beckman Research Institute of the City of Hope, 1500 E Duarte Road, Duarte, California 91010, United States

[‡]Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach, Germany

Supporting Information

ABSTRACT: Protein–protein interactions are implicated in the pathogenesis of many diseases and are therefore attractive but challenging targets for drug design. One of the challenges in development is the identification of potential druggable binding sites in protein interacting interfaces. Identification of interface surfaces can greatly aid rational drug design of small molecules inhibiting protein–protein interactions. In this work, starting from the structure of a free monomer, we have developed a ligand docking based method, called “*FindBindSite*” (FBS), to locate protein–protein interacting interface regions and potential druggable sites in this interface. *FindBindSite* utilizes the results from docking a small and diverse library of small molecules to the entire protein structure. By clustering regions with the highest docked ligand density from FBS, we have shown that these high ligand density regions strongly correlate with the known protein–protein interacting surfaces. We have further predicted potential druggable binding sites on the protein surface using FBS, with druggability being defined as the site with high density of ligands docked. FBS shows a hit rate of 71% with high confidence and 93% with lower confidence for the 41 proteins used for predicting druggable binding sites on the protein–protein interface. Mining the regions of lower ligand density that are contiguous with the high scoring high ligand density regions from FBS, we were able to map 70% of the protein–protein interacting surface in 24 out of 41 structures tested. We also observed that FBS has limited sensitivity to the size and nature of the small molecule library used for docking. The experimentally determined hotspot residues for each protein–protein complex cluster near the best scoring druggable binding sites identified by FBS. These results validate the ability of our technique to identify druggable sites within protein–protein interface regions that have the maximal possibility of interface disruption.



INTRODUCTION

Proteins often form complexes with other proteins or DNA for carrying out various cellular functions. Due to the importance of the protein complexes in cellular functions, targeting protein–protein interactions has been an attractive therapeutic strategy for various diseases.¹ Modulating protein–protein interactions through the use of antibodies, peptides, and small molecules is a major therapeutic focus.¹ Protein–protein interactions occur over a large surface area and therefore are particularly challenging to disrupt using small molecules. In addition to the large interaction surface area covered by protein–protein interfaces (PPIs) they also contain fewer features when compared to small molecule–protein interaction sites.^{2,3} Antibodies and peptides are the direct approach for designing therapeutics that can specifically target the interface region of protein–protein interactions. However, the difficulty with antibodies generally is that they are unable to access intracellular targets, and peptides have been in general poor drugs.^{1,4,5} Despite some promising results indicating successful targeting of PPI, few small molecule therapeutics for targeting PPI have progressed to the clinical phase.^{1,6,7}

Although PPIs cover a large surface area making them difficult to disrupt, there are certain residues in the PPI known as “hot spots” that show substantial contribution to the protein–protein free energy of binding.⁸ A hot spot has been defined as a 1 kcal/mol decrease in free energy of binding upon mutation of a specific residue, although the cutoff value for hot spots can vary depending on the study.⁹ Thus, identifying regions that contain clusters of hot spots in protein interfaces is an important component for designing small molecules that target the PPI. Combining computational methods that identify potential small molecule binding pockets and small molecule hits with experimental testing of hits can lead to savings in both time and cost when identifying protein–protein inhibitors. The wealth of protein structures available in the Protein Data Bank (www.rcsb.org) has facilitated the development of computational methods with the goal of identifying potential binding sites in protein structures. These computational methods could be classified into sequence/geometry based,^{10–20} energy-based,^{21–26} probe based,^{27–29} and ligand docking based

Received: December 17, 2013

Published: April 25, 2014

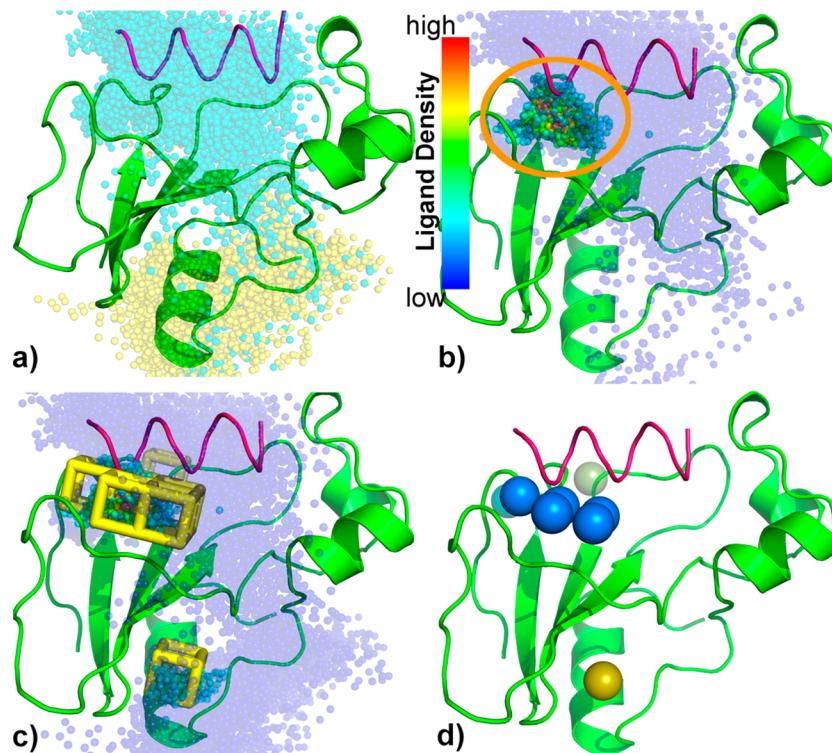


Figure 1. a) The target protein is shown in green and its interacting partner shown in pink. The individual spheres are representative of the heavy atoms located within these regions before the cutoff is applied. b) A single high density cluster is shown in the solid colored spheres highlighted in orange. The spheres are colored according to ligand density with regions of highest ligand density shown in red and lowest dark blue. c) The boxes formed around pockets of high ligand density are shown in yellow. d) The output from FBS shows the resulting clusters shown in the separate grouping the areas of high heavy atom density into the colored spheres on the protein surface.

approaches.^{30,31} Combinations of these methods have also been developed and used to identify PPI regions.^{21,32–34} Sequence and some geometry based methods rely on similarity in amino acid sequence or simplified geometry of the known small molecule binding sites in proteins.^{10–15} Therefore, the results of these methods are highly dependent on the quality of the training data sets available and hence are unreliable for those cases with no similarity available. Additionally, sequence based methods are unable to distinguish between conserved residues that are part of a binding site from conserved residues that are important to protein structure.¹⁴ Other geometry based methods attempt to locate cavities in protein surfaces that are characteristic of observed protein–protein or protein–ligand interactions.^{16–20} Energy-based methods dissect the nature of the interactions in the binding sites into various types of energetic contributions such as van der Waals and Coulombic interaction energies using probe molecules.¹⁴ The quality of the results from these methods is also dependent on how close the target protein is to the training set. The ligand docking method docks a small library of compounds to a preselected region of the protein to assess the druggability.³¹ This method does not examine all possible potential binding sites on the protein surface. Instead it focuses on a binding site preselected by the user to calculate druggability.

Molecular dynamics (MD) simulations of proteins using a mixture of hydrophobic solvent and water have been performed to identify potential binding sites on a protein surface. Using this approach the binding sites were predicted by calculating the residence time of various types of solvent molecules in each potential cavity. This method has been tested on 5 proteins and could be limited by the length of the simulation time as well as

the diffusion rate of these hydrophobic solvent molecules into the crevices.³⁵ Summarizing, the drawbacks in any of these methods are 1) generality of the method for application to many types of proteins that lack similarity to the training set, 2) lack of information on all possible potential druggable binding sites in the protein, and 3) sorting all the possible binding sites not only by the property of the amino acids lining each pocket but also using properties of a large set of high affinity ligands.

In this paper we have developed a computational method called *FindBindSite* (FBS), for identifying 1) location of the PPI and 2) potential druggable binding sites in PPIs and rank these sites by their affinity to bind small drug-like molecules. Here we define the druggability of a binding site as measured by the binding affinity of small molecules to this site. The FBS method involves several steps: 1) docking a small library of drug-like small molecules or fragments of various sizes, shape, and other properties on to the entire protein surface generated using a free monomer structure wherever available (the monomer that forms the protein complex), 2) sorting the various binding pockets by the density of ligands docked, followed by the ligand docking energy, and the nature of the ligands that dock to each binding site, and 3) starting from the high ligand density binding pockets and concatenating the contiguous binding pockets to form a large potential binding surface, which is useful for predicting protein–protein interface. We have validated the ability of FBS to identify the site of the protein–protein interactions for 41 protein–protein complexes. When the interface regions are mapped starting from the high ligand density binding pocket and concatenating the contiguous binding pockets, we can correctly predict over 70% of the total protein–protein interacting surface area in 24 out of

41 test proteins. We have also demonstrated the ability of FBS to identify the druggable binding pockets within the PPI using the crystal structure of the free monomer wherever available. FBS identified cavities in the protein–protein interacting surface in 37 out of 41 protein complexes. The top scoring binding site from FBS was located in the protein–protein interacting surface in 28 out of 41 complexes tested, achieving an overall hit rate of 68%. FBS identifies the PPI or the known inhibitor binding site in 38 out of 41 cases in the top ten scoring binding sites. In the three cases where FBS failed out of 41 protein–protein complexes studied, the free monomer undergoes substantial conformational changes upon complex formation. FBS results are robust to variation in the size and nature of the ligand library docked to the monomer structure. Protein flexibility plays an important role in predicting druggable binding sites and also in predicting the location of the interacting interface. FBS does not account for protein flexibility. However, protein flexibility can be accounted for by using the FBS method on snapshots extracted from MD simulations, but this aspect has not been tested in this paper.

METHODS

Description of the FBS Method. The FBS method consists of several steps:

1. Generating an energy grid for docking to the whole protein
2. Docking a database of ligands or fragments to the entire protein surface
3. Analysis of the docked ligand data set.

The details of the first two steps are provided in the next sections under protein structure preparation and ligand database preparation. Here we describe the analysis used for identification and scoring of the multiple binding sites identified by docking. In this work we have used *Glide* (Schrödinger Inc.) for docking the ligand and fragment databases saving the top scoring pose by glide score for each ligand. Once the ligands were docked using *Glide*,^{36,37} the coordinates of the heavy atoms from all the docked ligands were extracted. The heavy atom coordinates are shown as spheres in Figure 1a. The centroid of the heavy atom coordinates from the region with highest ligand density (shown in Figure 1b) was used as the starting point. Using this centroid as the center, a cube of length 3.4 Å was placed around it, and the density of ligands within this cube was calculated. Cubes are then placed along each adjoining face, and the ligand density in each of the neighboring cubes was calculated. The neighboring cubes with greater than 3% ligand density were clustered with the original cube. This procedure is continued until a cube that does not meet the cutoff value was encountered, as shown in Figure 1c. The clustered cubes were then sorted by ligand density and then by ligand–protein interaction energy, calculated as the sum of van der Waals and hydrogen bond energies from *Glide*. FBS then outputs the rank of each site as well as the centers of the clusters that show ligand occupancy above the cutoff ligand density. These coordinates were then superimposed on the surface of the target protein as shown in Figure 1d.

Predicting Protein–Protein Interfaces Using FBS. One of our goals was to assess the performance of the FBS method in locating the PPI region on the surface of a monomer structure and to calculate the percentage of the protein–protein interacting surface area that gets covered by ligands docked in FBS. To do this we started with the top scoring region from

FBS and concatenated the low ligand density regions from FBS that are contiguous to the top scoring region. This resulted in identifying parts of a surface rather than a binding cavity within the PPI. To estimate the extent of coverage of the PPI detected by FBS, we calculated the ratio of the solvent accessible surface area (SASA) of the residues of the protein located within the concatenated clustered cubes described above and that of the entire protein–protein interacting interface from the structure of the protein–protein complex. The SASA was calculated using PyMOL software.^{38–40} This ratio gives the percentage coverage of the PPI surface that can be recovered by FBS compared to the total PPI surface area.

Validation of the FBS Results. Starting from the free monomer structures we have tested the FBS method for identifying locations on the protein–protein interface, or the protein–inhibitor interface, in the top two scoring sites for 41 different proteins. If the top two scoring sites were not located in these two interfaces, the entire list of FBS predicted binding sites was analyzed to verify if any of these sites were located in the protein interface. The predicted binding sites were sorted and evaluated by their positioning within PPIs. The predictions were considered successful with high confidence if either of the top two scoring FBS predicted binding sites was located in the protein–protein or protein–inhibitor interface. Predictions would be noted as a success with low confidence if the site scores lower than the second best scoring site by FBS. If none of the sites predicted by FBS were located in the PPI, then we would consider it unidentified.

Procedure for Generating Ligand Database for Docking. Five separate ligand databases from different sources and different sizes were created: 1) a diverse set of 60,000 compounds was selected from the National Cancer Institute's open database of compounds available at <http://cactus.nci.nih.gov/download/nci/>, 2) 60,000 compounds were randomly selected from the ZINC clean drug-like subset, 3) 20,000 compounds were randomly selected from the ZINC clean drug-like subset, 4) 60,000 compounds were randomly selected from the ZINC clean fragment subset, and 5) 20,000 compounds were randomly selected from the ZINC fragment-like subset.⁴¹ Each database was prepared using Schrödinger LigPrep, and the Epik package was used to generate ionization and tautomer states.^{42–45} For database 1, we used molecular charge, shape, and size of the molecules as properties to select a diverse set of molecules from the database.

Protein Structures Preparation and Docking. 41 proteins that are known to form protein–protein complexes and with crystal structures of both complex and monomers wherever available were selected to make up our test set. The protein–protein complex crystal structures are available for all the 41 proteins chosen here. For performing the FBS calculations, we gave preference to protein structures that had been crystallized as free monomers (18 out of 41), followed by inhibitor bound proteins (5 out of 41), and finally protein–protein complex bound forms (18 out of 41). Some peptide bound structures were also included to maximize the diversity of the test set. The resulting set represents a diverse set of protein–protein interactions including deep catalytic pockets, binding clefts, shallow interacting surfaces, α -helical bundles, and β -sheets.

Crystal structures were obtained from the Protein Data Bank (www.rcsb.org) and cleaned up to eliminate any multimeric units and ions. Monomers from the crystal structures were then subjected to protein preparation wizard in *Maestro* (Schrö-

dinger Inc.) to add missing side chains and hydrogens.^{43,46} Docking receptor grids were then generated using Glide suite with van der Waals radii scaled down to 0.5. Although this is not the default value in Glide, we have shown that using the default value of 0.8 yields similar results as described in the Results section.^{47,48} The docking grid was centered at the geometric center of the input structure, and the docking boxes were expanded to encompass the entire monomer protein unit with an additional 5 Å padding along each axis.^{49,50} Any structure that exceeded 10,000,000 grid points was divided into two equal parts, and FBS was performed on each of the component parts. The FBS results were concatenated and analyzed. Docking was performed using the High Throughput Virtual Screening setting with each given ligand subset in each reduced database. Ligand poses with docking energies of up to 100 kcals/mol were retained, and 1 pose per ligand was used for FBS analysis and ranking of binding sites. All other values not noted were default values set by Glide.

RESULTS

Performance of the FBS Method on 41 Protein Complexes. We tested the FBS procedure on 41 different protein–protein complexes, 37 of which have known small-molecule inhibitors. Only 25 of these 41 have crystal structures of the monomer with the inhibitor bound. Wherever available we performed the FBS calculations using the crystal structure of the free unbound monomer because this represents the realistic and most challenging scenario. We tested the results of FBS for various input protein structures in this order of preference: free monomer protein structures, followed by inhibitor bound monomer structures, and where the former two structures were not available we used the structures of the monomer extracted from the structures of the protein–protein complexes. This order was selected to test the FBS method in a real case scenario for inhibitor design, where only the structure of the monomer might be available. This also allowed us to study the performance of FBS in test cases where a conformational change occurs in the protein structure upon complexation.

Using these criteria we were able to obtain 18 unbound monomer structures for testing. For the remaining 23 structures we used the structure of the monomer extracted from the crystal structures of their respective protein–protein (18) or protein–inhibitor (5) complexes. FBS identified the cavity located in the protein–protein interacting surface (that is also the binding site of the known inhibitor wherever available) in the top ten druggable sites in 37 out of 41 protein complexes. This cavity identified by FBS was the inhibitor binding site for 22 out of 25 cases with known inhibitor bound monomer crystal structure. Using FBS we were able to identify the binding cavity in the PPI as the top scoring site, in 28 out of 41 complexes. Thus, achieving an overall hit rate of 68%. Here we have scaled the van der Waals radii of atoms while docking by 0.5. We tested the performance of the FBS method on predicting the binding site cavity in the protein–protein interacting surface, by increasing the van der Waals radii scaling factor to 0.8. We observed as shown in Table S1, only 7 out of 41 proteins showed a different FBS prediction with 0.5 as scaling factor. In 5 out of the 7 proteins where the results changed, FBS results with 0.5 scaling factor was better than with 0.8.

Q-SiteFinder is an example of an existing computational tool used for identifying druggable binding sites in proteins. We identified 11 proteins that are common to our test set and the

published test set for Q-SiteFinder, and hence we chose to compare these results.^{26,51} The Q-SiteFinder method showed a maximum of 70% area under the ROC curve by including the top 25% of the predicted binding sites.⁵¹ A comparison of the results to FBS for 11 overlapping complexes revealed that FBS is able to identify the correct binding in all 11 structures with a high degree of confidence Table S2. Q-SiteFinder tests were also performed using the monomer extracted from the crystal structures of the protein–protein complexes. The FBS method shows fairly robust predictions starting from a monomer structure which is an advancement over other existing methods. The improvement in FBS results can likely be attributed to the use of multiple fragment-like ligands that can search and rank multiple potential binding pockets.⁵² We theorize the success of FBS is due in part to the diversity provided by using a ligand library, whereas other docking based methods either use a small number of molecules or examine druggability by docking to a specific site in the protein.

FBS Hit Rate Starting from Free Monomer Structures.

In a real world drug design scenario only the structure of the free monomer might be available. Since we need at least this structure to start the FBS runs, we calculated the hit rate of FBS starting from just the free monomer structures. Only 18 out of 41 cases studied here have structures of free monomers, and Table 1 shows a summary of the results. The rankings of the

Table 1. Summary of FBS Analysis Performed on Free Monomers Crystal Structures^a

Complex	PDB id	Input protein	Binding site ranking
B2AR-Gs/Galpha	2RH1	B2AR	H
barnase/barstar	1A2P	barnase	H
Bcl-xL/Bak	1R2D	Bcl-xL	H
CheA-Che Y	1FWP	CheA	L
CheA-Che Y	3CHY	CheY	H
FTSZ/ZipA	1F46	ZipA	L
hen egg white lysozyme	2LYO	HEWL	H
HPV E1/HPV E2	1R6K	HPV E2	H
IL2/IL2R	3INK	IL2	H
IL4/IL4R	1HIK	IL4	L
LFA-1/ICAM-1	3F74	LFA-1	H
MDM2/P53	1Z1M	MDM2	H
PCNA/FEN1	1VYM	PCNA	NF
PCSK/LDLR	3BPS	PCSK	NF
pepsin/PI3	1PSN	pepsin	NF
peroxidase/cytochrome c	1BEK	peroxidase	L
TNF/TNF-R	1TNF	TNF	L
XIAP BIR3/CASPASE 9	1F9X	XIAP	H

^aBinding site ranking are ranked (H) predicted with high confidence, (L) predicted with low confidence, and (NF) binding site not found. High confidence sites are the top two ranking sites, and low confidence sites are those sites that are not ranked as the top two.

FBS results have been separated into three categories: H denotes a high confidence (within top two) binding site identified by FBS, L denotes a low confidence site identified in the protein–protein interaction region but not ranked as the two top ranking sites, and NF (not found) denotes a site that was completely missed by FBS. Using FBS, we were able to identify a pocket within the PPI with high confidence in 10/18 cases or 55% hit rate. This hit rate is comparable to the published accuracy of other methods on a different set of protein structures.^{51,52} If we relaxed the criteria to include sites

predicted with low confidence we correctly identified the binding site in 15/18 or 83% of the structures. One of the possible reasons for the failure of the scoring method could be the Coulombic interactions in the sites with many charged residues that dominated the ranking of the potential binding sites. These results suggest that further optimization of the scoring method is necessary to improve the hit rate with high confidence. The structures in which the PPI was not identified at all in any of the potential binding sites were free monomer structures that undergo significant conformational changes upon complexation.

When the test set was expanded to include monomer structures from protein complexes that are in complex with either an inhibitor or the partner protein, shown in Table 2, we were able to identify 18/23 or 78% as the top ranked site and identify binding pockets in 23/23 or 100% of potential binding sites with lower confidence.

Table 2. Summary of FBS Analysis Performed on Bound Crystal Structures^a

Protein/protein complex	PDB id	Ptn-Ptn/ Ptn-Inh	Input protein	Binding site ranking
carboxypeptidase/ LSI	1DTD	Ptn-Inh	carboxypeptidase	H
proteinase A/IA3	1DPJ	Ptn-Inh	proteinase A	H
thrombin/rhodniin	1TBQ	Ptn-Inh	thrombin	H
trypsin/tryp inhibitor	2UYY	Ptn-Inh	trypsin	H
camel H chain/ lysozyme	1MEL	Ptn-Ptn	lysozyme	H
cathepsin H/ stefin A	1NBS	Ptn-Ptn	cathepsin H	H
chagasin/papain	3E1Z	Ptn-Ptn	papain	H
chagasin/papain	3E1Z	Ptn-Ptn	chagasin	H
CTLA-4/B7-2	1I8S	Ptn-Ptn	CTLA-4	H
CTLA-4/B7-2	1I8S	Ptn-Ptn	B7-2	H
FKBP12/TGFR	1B6C	Ptn-Ptn	TGFR	H
FKBP12/TGFR	1B6C	Ptn-Ptn	FKBP12	L
HPV E1/E2	1TUE	Ptn-Ptn	HPV E2	L
HPV E1/E2	1TUE	Ptn-Ptn	HPV E1	H
Rap1A/cRaf1	1GUA	Ptn-Ptn	cRaf1	L
Rap1A/cRaf1	1GUA	Ptn-Ptn	Rap1A	L
subtilisin/eglin C	1CSE	Ptn-Ptn	subtilisin	H
subtilisin/eglin C	1CSE	Ptn-Ptn	eglin C	H
TCR/TAX peptide	1AO7	Ptn-Ptn	TCR	H
thrombin/hirudin	4HTC	Ptn-Ptn	thrombin	H
XIAP BIR3/ CASPASE 9	1NW9	Ptn-Ptn	CASPASE 9	H
XIAP BIR3/SMAC	1G73	Ptn-Ptn	SMAC	H
XIAP BIR3/SMAC	1G73	Ptn-Ptn	XIAP BIR3	L

^aThe type of complex is listed as either protein–protein (Ptn-Ptn) or protein–inhibitor (Ptn-Inh) depending on the type of the input crystal structure used.

Analysis of the FBS Results for Various Types of Protein–Protein Interaction Interface. The nature of the PPI can be roughly classified as surfaces that have a) deep binding cavities and b) shallow surfaces on the surface of a protein. In this section we highlight three examples of protein–protein complexes with different shapes of the interacting interface tested with FBS. The barnase-barstar complex represents an example of a protein with a well-defined binding cavity in the protein–protein interacting surface; the *Bcl-xL*/

BAK complex represents a case with a shallow binding groove and a shallow binding site; the LFA-1/ICAM-1 complex represents a protein with an allosteric PPI and an allosteric druggable site. These examples demonstrate the ability of FBS to identify interfaces and druggable sites on the proteins. Hotspot residues in PPI are those residues that upon mutation reduce the protein–protein interaction free energy significantly. Using 1 kcal/mol as the energy cutoff for the definition of hotspot residues,⁹ we have analyzed if the experimentally identified hotspot residues cluster around the top scoring binding site identified by FBS for various types of PPI.

Barnase-Barstar Complex: A Case with Well-Defined Cavity in the Protein–Protein Interacting Surface. The barnase-barstar complex has a well-defined cavity in the protein–protein interacting surface. The barnase-barstar complex is composed of the barnase ribonuclease isolated from *Bacillus amyloliquefaciens* and the intracellular inhibitor barstar. The catalytic site of barnase contains multiple basic residues giving it an overall positive charge that facilitates enzymatic activity with RNA. Barstar, the natural inhibitor of barnase complements the barnase interacting surface with an overall negative charge.

We used the 110 residue crystal structure of the wild type barnase (pdb ID: 1A2P) for FBS calculations. FBS identified the cavity in the protein–protein interaction site as the top scoring site shown as deep blue spheres in Figure 2. This site is

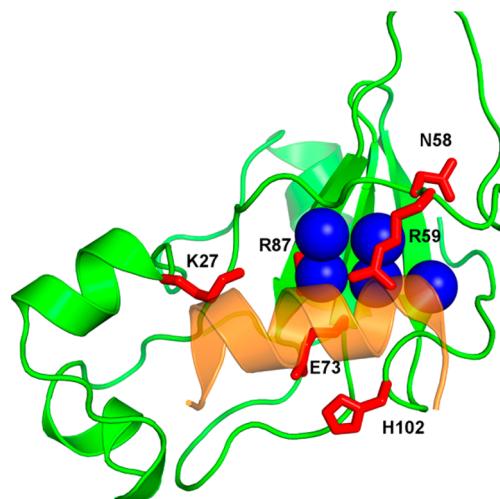


Figure 2. Barnase-barstar complex is shown here. A representation of the binding site predicted by FBS is shown in deep blue spheres. We also show in the orange helix representing the helical region of barstar that inhibits barnase. The hotspot residues of barnase are highlighted in red showing the close proximity of the predicted binding site to the hot spot residues.

also the inhibitor binding site of the natural inhibitor barstar (2ZA4) shown as a helix in Figure 2. The residues K27, I55, F56, S57, N58, R59, E60, G61, K62, E73, R87, L89, H102, Y103, Q104, T105, and F106 of barnase are within 5 Å of the top scoring binding pocket found using FBS. Experimental alanine scanning mutations have identified the residues K27, N58, R59, E73, R87, and H102 as hot spots in the barnase-barstar complex shown as the red sticks in Figure 2.⁵³ The top scoring binding site identified by FBS is located in the inhibitor binding site and covers part of the barnase/barstar interacting surface. The hotspot residues cluster in the FBS predicted top scoring binding site which confirms that this predicted site is a

druggable site. We have also tested FBS using the internally cross-linked barnase monomer structure (3KCH) and identified a potential binding site in the same location. The barnase-barstar complex test case is a standard test case for validating computational methods to identify potential druggable binding sites, since the barnase-barstar complex has a well-defined binding cavity in the protein–protein interface. Here we have demonstrated the ability of our FBS method to not only identify the correct binding site but also predict a binding site that overlaps with a cluster of hotspot residues that coincides with the inhibitor binding site. This shows that the FBS method can predict the PPI which is assumed to be the most druggable site in barnase.

Bcl-xL/Bak – A Case with a Defined Binding Groove.

Bcl-xL is an apoptosis regulator protein that is inhibited by the Bak peptide. This protein has been a therapeutic target for cancer because of its frequent mutations observed in cancer. The Bak peptide binds to a hydrophobic groove formed between three alpha helices of Bcl-xL. The interaction with the alpha helical region of Bak has been well studied, and inhibitors have been found which bind in this interface region.⁵⁴ The shape of the protein–protein interaction surface of this complex is not just a well-defined cavity but a larger groove.

We used the crystal structure (pdb ID: 1R2D) of the free monomer Bcl-xL for evaluation using FBS. As shown in Figure 3, the predicted binding site shown in deep blue spheres in the

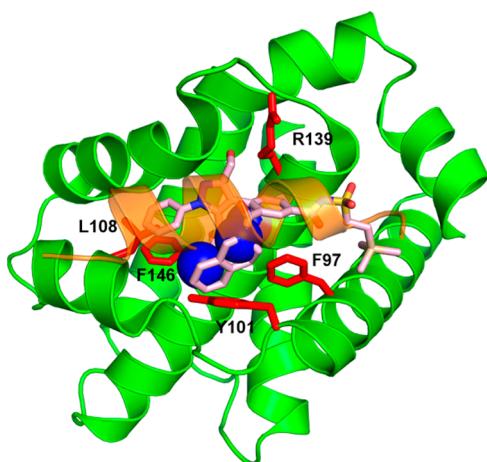


Figure 3. Bcl-xL/Bak complex is shown here. The Bcl-xL protein is shown in green and the predicted binding site represented by the blue spheres. The BAK peptide inhibitor in orange and a quinazoline sulfonamide inhibitor shown in pink show close proximity to the predicted binding site.

figure is the top ranked binding site in the PPI region. This site also overlaps with the binding site of the quinazoline sulfonamide family of inhibitors for this protein.⁵⁵ Out of the many residues that line the FBS predicted binding pocket, the residues F97, Y101, L108, R139, A142, and F146 have been shown to be hotspots to both ligand and peptide binding through mutational analysis.^{56,57} These hotspot residues shown as sticks in Figure 3 also cluster in the FBS identified binding site. Again using FBS we were able to identify a binding site that includes hotspot residues. FBS runs using the monomer structures from the small molecule inhibitor bound (pdb ID: 4EHR) structure and BAK peptide bound (pdb ID: 2LP8) structure also reveal that the top scoring binding sites are in the region of PPI.

LFA-1/ICAM-1 - A Case with an Allosteric Inhibitor Binding Site.

The LFA-1/ICAM-1 complex is an example, where the known inhibitor binds to a site that is allosteric to the protein–protein interaction surface. Lymphocyte function-associated antigen (LFA-1) plays an important role in inflammation by activating T-cells and promoting leukocyte migration. The activity of LFA-1 is modulated by the binding of ICAM-1. The drug Lovastatin inhibits the LFA-1 binding to ICAM-1 through an allosteric mechanism at the allosteric site on LFA-1 shown in Figure 4. Upon the binding of Lovastatin,

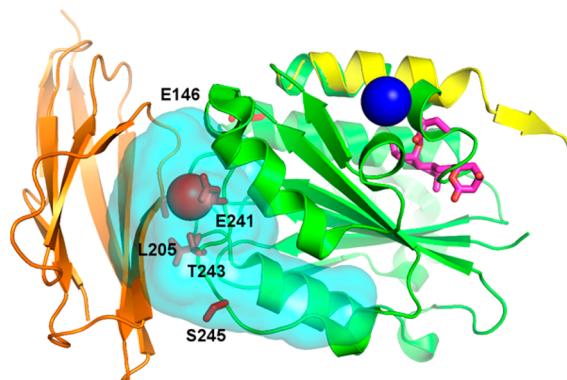


Figure 4. Two binding sites predicted by FBS, out of which one occupies the ICAM binding region of LFA1, and the other occupies the Lovastatin binding region. The red sphere represents a binding site predicted by FBS that occupies the ICAM (orange) binding region. The blue sphere occupies the allosteric Lovastatin binding site which was a lower ranked site predicted by FBS. Analysis of mutagenic binding assays reveals the cluster of residues shown in red that have been shown to be necessary for ICAM binding.

LFA-1 undergoes a significant conformational change, revealed by aligning the bound and unbound crystal structures. With this knowledge we selected this structure to highlight the ability of FBS to identify not only the PPI but also the druggable allosteric inhibitor binding site.

We performed FBS using the wild type LFA-1 free monomer structure (pdb ID: 3F74) and identified the top scoring binding site in the ICAM-1 binding interface (shown in red color sphere in Figure 4), and a second binding site was found in the region where the inhibitor Lovastatin binds (shown as blue color sphere in Figure 4). Despite the significant conformational changes undergone by the helix bordering the Lovastatin binding site we were able to identify this site as a the second best scoring site. For comparison we have aligned the inhibitor bound structure (pdb ID: 1CQP) to the unbound monomer structure. The helix shown in yellow is from the inhibitor bound structure, while the unbound structure is shown in green. Inspection of the inhibitor binding site shows that free monomer structure undergoes a conformational change in this region. Furthermore, the FBS identified site in the PPI interface consists of a cluster of charged residues in the interface region that make up the metal ion-dependent adhesion site. Mutations of M140, E146, L205, E241, T243, and S245 have been shown to disrupt metal binding and thus the LFA-1/ICAM-1 interaction.^{58,59} These results demonstrate the ability of FBS to not only identify a binding site with hotspots in the PPI but also an allosteric druggable site on the protein. It is important to note that although FBS identifies the ICAM-1 binding site as well as Lovastatin binding site, the Lovastatin binding site is a low confidence site among the FBS predicted binding sites.

Location of Hot Spot Residues in the FBS Predicted Binding Sites for Other PPI Complexes. In addition to the cases discussed above, experimentally identified hotspots have been found to cluster near the FBS predicted top scoring binding sites in other protein–protein complexes. Using the ASEdb and PINT databases we compiled the data for the measured change in the binding free energy for mutation of 20 residues in 5 different proteins that were a subset of the 41 proteins we have tested (Table S3).^{60,61} FBS identified binding sites in locations with hotspot residues within 7 Å in 15/20 (75%) observed residues (43). In the case of the IL-4/IL-4R complex, three residues are located just beyond the distance cutoff value of 7 Å. However, increasing the cutoff value to 8 Å brings all but one of these residues within proximity of the FBS predicted binding site. The hotspot residues in the IL-2/IL-2R complex are not clustered in the same region of the protein surface and the mutations involve distant noncontiguous regions. Despite this fact the predicted binding site was proximal to the hotspot residues, but these sites were ranked lower and have a lower density.

Testing the Robustness of the FBS Method with Alternative Ligand Database. The FBS results described so far have been tested by docking a 60,000 small molecule ligand database selected by ligand diversity from the NCI database of small molecules, as described in the Methods section. To test the sensitivity of the FBS results to the number and nature of ligands we generated three different databases of ligands: 1) a library with 60,000 fragments selected from the ZINC fragment database (<http://zinc.docking.org/>), 2) a small molecule library of 60000 selected from the Zinc drug-like small molecule database and 3) 20,000 small molecule drug-like ligands also from the ZINC small molecule database. The resulting libraries were then prepared as described in the Methods section using the same procedure used for the preparation of the NCI subset of compounds.

We first tested the results of using an alternative compound library by comparing the results with the two 60,000 compound and fragment databases. These databases were used for docking against the same set of 41 proteins, and the results are summarized in Table 3 and Figure 5. The overall hit rate for the

Table 3. Summary of Results from FBS Using Each Database

	top ranked in PPI (%)	BS in PPI (%)	not found (%)
NCI database-subset 60k	28 (68)	38 (93)	3 (7)
ZINC database-drug-like subset 20k	28 (68)	37 (90)	4 (10)
ZINC database-drug-like subset 60k	26 (63)	37 (90)	4 (10)
ZINC database-fragment-like subset 20k	24 (59)	37 (90)	4 (10)
ZINC database-fragment-like subset 60k	24 (59)	37 (90)	4 (10)

drug-like 60,000 molecule Zn database is 26/41 (63%) and the 60,000 fragment-like database is (59%) which is less than that of the 60,000 NCI molecules database (71%). This shows that there is some level of sensitivity to the size and nature of the ligand database size using the current ranking system. We performed a *t*-test assuming equal variance to assess the significance of the difference of results between the NCI, drug-like, and fragment-like databases. We did not observe a statistically significant difference between the NCI and drug-like

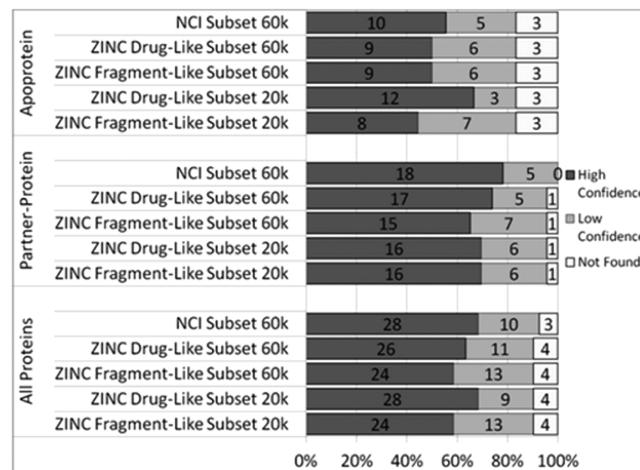


Figure 5. Overall results from FBS are summarized here in chart form. A breakdown of all the protein structures shows limited sensitivity to both the database size and nature of the database used. The groups are separated by the input protein type with a summary of the overall results and then further divided by the type of the library used.

databases ($p = 0.08$); however when comparing the results using the NCI and fragment-like database ($p = 0.02$) one was observed. However, the hit rate is the same if we ignore the FBS ranking of the sites and include sites predicted with low confidence. The results indicate that ranking of sites shows some sensitivity to the molecular weights of the compounds used because the fragment-like database was made up of compounds in the range of 120–250 Da and the drug-like compounds were in the weight range of 150–500 Da. A detailed list of which proteins were identified with low confidence, high confidence, or missed is given in Table S4 of the Supporting Information. Using the fragment database whether it is 20,000 or 60,000 fragments, we observed that the FBS ranking performed poorly due to certain interior cavities in the protein structure that accommodate small fragments being scored higher than cavities in the PPI. The reduction of the hit rate when using the 20,000 ligand or fragment database is due to ligand density cutoff that was used in scoring the sites. A lower ligand density cutoff increases the hit rate for the 20,000 molecule or fragment database. The prediction of the druggable binding site was missed uniformly for three proteins using the structures of the free monomers. These protein monomers undergo substantial conformational changes upon protein–protein complex formation.

Testing the Ability of FBS To Map PPI Surfaces. We observed that the high ranking FBS predicted binding sites, where the docked ligand densities were high, correlated strongly to the protein–protein interaction interface. We further wanted to assess the extent of coverage of the protein–protein interacting interface covered by the high scoring regions in FBS. Starting from the top ranking FBS predicted binding site, we expanded this site by concatenating adjacent lower ligand density sites identified by FBS. We then calculated the solvent accessible surface area covered by the residues in this extended site. The ratio of this SASA to the total protein–protein interacting interface area was calculated. This ratio gives us an estimate of the percentage coverage of the PPI interface area captured by FBS. An example of the surface area covered by FBS predicted binding regions is shown in Figure 6. The figure shows the interaction surface of the

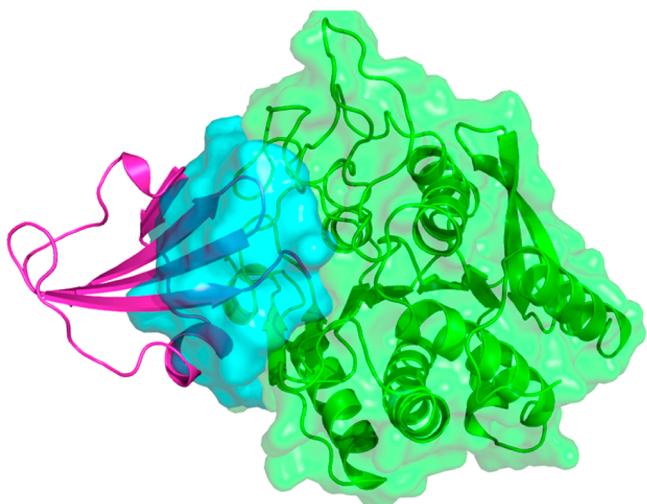


Figure 6. Carboxypeptidase/LSI complex is shown in this figure. The crystal structure is shown in the green cartoon and surface with the metallocarboxypeptidase inhibitor shown in purple. The interaction surface mapped by FBS is shown here in cyan and represents a complete mapping of the surface occluded by the complex structure.

carboxypeptidase/LSI complex. FBS predicted interaction surface covers most of the protein–protein interaction surface as shown in cyan surface representation in the figure. We calculated the percentage of the PPI surface covered by FBS for the 41 proteins in our test set. We observe that over 50% of the interaction area was captured by the FBS method in 30/41 structures and over 70% of the interaction area in 24/41 structures shown in Figure 7. The raw values for the overlapping surface areas are reported in Table S5. We found that the false positive rate in our sample set was less than 25% in 27/41 (66%) structures.

■ DISCUSSION

FBS has been shown to be a useful method for identifying potential druggable sites for disrupting protein–protein interactions. Using the structures of free monomers, and docking a set of 60,000 diverse small molecule ligands from

NCI small molecule database, we were able to identify top ranked druggable pockets in 11/18 starting from free monomer structures. Further analysis showed that experimentally identified hotspot residues are located within 5 Å of identified druggable sites. Such a prediction for a druggable binding site is potentially useful for selecting small molecules or fragments that anchor at this site and aid the disruption of the protein–protein interactions. When using crystal structures with interfaces already primed for protein–protein or protein–ligand interactions, we were able to discern with a high degree of confidence, the potential binding sites in PPIs. The FBS method can be useful in identifying sites and/or residues that might be involved in protein–protein interactions in an unknown case. The binding sites identified by FBS could also be used for designing inhibitors to protein–protein interactions. In our previous work we demonstrated the use of FBS to identify a new small molecule inhibitor for the protein STAT3 dimer. Our goal was to identify a small molecule inhibitor for the dimerization of the transcription factor STAT3 and not disrupt the closely related dimerization of STAT1. Using molecular dynamics simulations and FBS we identified an allosteric site that could inhibit the dimerization of the protein STAT3 implicated in cancer.⁶² Subsequently, using the virtual ligand screening technique, we identified a small molecule compound that selectively inhibited the dimerization of STAT3 and not the related STAT1.^{62,63}

FBS shows limited sensitivity to the number of compounds used in the ligand library, and we observed that a database with 60,000 compounds gives slightly better results than a database of 20,000 compounds. There was some sensitivity to the size of the small molecules chosen, and this was observed in the reduction of the hit rate when using the fragment-like databases. This reduction in hit rate with the fragment database was attributed to fragments docking in the interior of proteins that are not accessible to full compounds. Docking to internal cavities may be the result of using reduced van der Waals radii (scaling factor 0.5) or of the weights for hydrophobic enclosure and desolvation penalties used to score the docked poses. Additionally, Coulombic interactions dominate the top ranking binding site due to a large proportion of charged residues in certain sites. To address this we are currently investigating methods to improve the energy function by attenuating the Coulombic energy and including a desolvation penalty. The preference FBS shows for drug-like compounds over fragment-like compounds is interesting and in contrast from the small probe-like molecules used in other methods.^{16,31,35} The preference may be caused by the inability of our clustering method to bridge the areas in between fragment clusters and the chemical space of sampled by our compound libraries limiting the ability of docking to regions of low affinity. Overall these issues may highlight a dependence on the ability of the docking program to correctly score individual compounds accurately.

We observed that FBS is an effective method for identifying druggable sites in PPIs. Energetic methods like Q-SiteFinder are very effective for identifying properties that correlate to potential binding sites in protein–ligand complexes. However, its accuracy is markedly reduced when used to identify pockets in protein–protein interactions.⁵² Many low ranked pockets are included reducing the overall precision of the method. SiteMap is a method developed by Schrodinger Inc., and it leverages the same properties generated by docking compounds in FBS. SiteMap uses grid properties to identify possible interaction

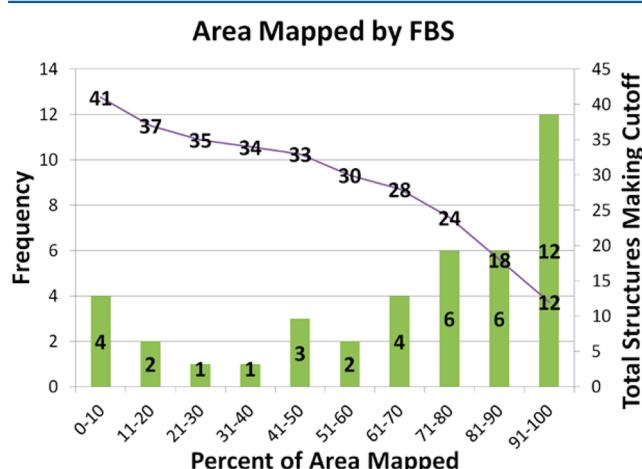


Figure 7. Results from the surface mapping of FBS are summarized here in chart form. The line shows the total number of structures falling within the cutoff values, and the bars represent the number of structures residing in each cutoff.

sites on a protein surface.²¹ However, SiteMap has not been shown to be effective in predicting sites in PPIs. FBS does leverage the grid properties used by SiteMap in the compound docking, so exploring the robustness of our method using other docking methods will be investigated in the future. A review of other methods shows hit rates in the range between 60 and 90%; FBS is able to predict the correct binding site with low confidence in 93% of the structures and the correct binding site with a high degree of confidence in 68% of the structures. Furthermore, the ability to map over 70% of the PPI surface area in 24/41 (58%) of structures demonstrates a unique capability in FBS that is not present in many of the other methods to the best of our knowledge. The extent to which this calculation overestimates the PPI surface area is typically less than 25% of the proteins' surface area.

CONCLUSIONS

FBS utilizes a large diverse ligand database to identify druggable sites within protein–protein interfaces. We observed that by clustering and mapping the sites occupied by these drug-like ligands we were able to identify druggable sites in the PPI. Using 41 proteins as a test set, FBS predicted the protein interacting surface with a high degree of confidence in 68% of complexes and 93% of complexes with low confidence. We are also able to map over 70% of the interaction surface in 59% of the structures using these predicted sites. The sensitivity of the FBS results to the size of the ligand database used was minimal. However, when using a compound library containing lower molecular weight fragments a decrease in precision was observed. Despite these fluctuations the accuracy of the method is unchanged suggesting some further ranking optimizations must be made. These results are comparable to or improvement upon other PPI binding site detection methods. We demonstrate that FBS has the potential to be an effective means for identifying druggable sites in PPIs.

ASSOCIATED CONTENT

Supporting Information

Tables S1, S2, and S4 contain breakdowns of the individual results from FBS on each protein complex. Table S3 contains hotspot residues with known $\Delta\Delta G$ values in our test set. Table S5 contains the calculated SASA values for all the complexes in our test set using the NCI database. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: 626-301-8408. E-mail: nvaidehi@coh.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Jianping Lin, Dr. Spencer Hall, and Mr. Allen Mao for helping the beginnings of this project. We thank Ms. Reshma Patel for generating the reduced NCI database for docking. We thank Boehringer-Ingelheim, Biberach, Germany for funding this project.

REFERENCES

- (1) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001–1009.
- (2) Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177–2198.
- (3) Smith, R. D.; Hu, L.; Falkner, J. A.; Benson, M. L.; Nerothin, J. P.; Carlson, H. A. Exploring protein–ligand recognition with Binding MOAD. *J. Mol. Graphics Modell.* **2006**, *24*, 414–425.
- (4) Perot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A. C.; Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discovery Today* **2010**, *15*, 656–667.
- (5) Arkin, M. R.; Whitty, A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Curr. Opin. Chem. Biol.* **2009**, *13*, 284–290.
- (6) Macarron, R. Critical review of the role of HTS in drug discovery. *Drug Discovery Today* **2006**, *11*, 277–279.
- (7) Whitty, A.; Kumaravel, G. Between a rock and a hard place? *Nat. Chem. Biol.* **2006**, *2*, 112–118.
- (8) Clackson, T.; Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383–386.
- (9) Kortemme, T.; Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14116–14121.
- (10) Peters, K. P.; Fauck, J.; Frommel, C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.* **1996**, *256*, 201–213.
- (11) Zhong, S.; MacKerell, A. D., Jr. Binding response: a descriptor for selecting ligand binding site on protein surfaces. *J. Chem. Inf. Model.* **2007**, *47*, 2303–2315.
- (12) Binkowski, T. A.; Naghibzadeh, S.; Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **2003**, *31*, 3352–3355.
- (13) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinf.* **2009**, *10*, 168.
- (14) Ghersi, D.; Sanchez, R. Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *J. Struct. Funct. Genomics* **2011**, *12*, 109–117.
- (15) Geppert, T.; Hoy, B.; Wessler, S.; Schneider, G. Context-based identification of protein-protein interfaces and "hot-spot" residues. *Chem. Biol.* **2011**, *18*, 344–353.
- (16) Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 516–29.
- (17) Glaser, F.; Morris, R. J.; Najmanovich, R. J.; Laskowski, R. A.; Thornton, J. M. A method for localizing ligand binding pockets in protein structures. *Proteins: Struct., Funct., Bioinf.* **2006**, *62*, 479–488.
- (18) Till, M. S.; Ullmann, G. M. McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Graphics Modell.* **2010**, *16*, 419–429.
- (19) Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinf.* **2006**, *7*, 316.
- (20) Brady, G. P., Jr.; Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383–401.
- (21) Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146–148.
- (22) Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (23) Seco, J.; Luque, F. J.; Barril, X. Binding site detection and druggability index from first principles. *J. Med. Chem.* **2009**, *52*, 2363–2371.
- (24) An, J.; Totrov, M.; Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell. Proteomics* **2005**, *4*, 752–761.
- (25) Ghersi, D.; Sanchez, R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. *Proteins: Struct., Funct., Bioinf.* **2009**, *74*, 417–424.

- (26) Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics* **21**, 1908–1916.
- (27) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (28) Harris, R.; Olson, A. J.; Goodsell, D. S. Automated prediction of ligand-binding sites in proteins. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1506–1517.
- (29) Ruppert, J.; Welch, W.; Jain, A. N. Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci.* **1997**, *6*, 524–533.
- (30) Chang, D. T.; Oyang, Y. J.; Lin, J. H. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.* **2005**, *33*, W233–238.
- (31) Huang, N.; Jacobson, M. P. Binding-site assessment by virtual fragment screening. *PLoS One* **2010**, *5*, e10109.
- (32) Huang, B.; Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **2006**, *6*, 19.
- (33) Nayal, M.; Honig, B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 892–906.
- (34) Rossi, A.; Marti-Renom, M. A.; Sali, A. Localization of binding sites in protein structures by optimization of a composite scoring function. *Protein Sci.* **2006**, *15*, 2366–2380.
- (35) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2435–2447.
- (36) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (37) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (38) Eslami, H.; Mojahedi, F.; Moghadasi, J. Molecular dynamics simulation with weak coupling to heat and material baths. *J. Chem. Phys.* **2010**, *133*, 084105.
- (39) *The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.0*; Schrödinger, LLC: 2010.
- (40) *The PyMOL Molecular Graphics System, Version 1.3r1*; Schrödinger, LLC: 2010.
- (41) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.
- (42) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
- (43) Maestro, 9.2; Schrödinger, LLC: New York, NY, 2011.
- (44) Suite 2012: LigPrep, 2.5; Schrödinger, LLC: New York, NY, 2012.
- (45) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 591–604.
- (46) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.
- (47) Bhattacharya, S.; Hall, S. E.; Li, H.; Vaidehi, N. Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation. *Biophys. J.* **2008**, *94*, 2027–2042.
- (48) Suite 2012: Glide, 5.8; Schrödinger, LLC: New York, NY, 2012.
- (49) Floriano, W. B.; Vaidehi, N.; Goddard, W. A., III; Singer, M. S.; Shepherd, G. M. Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10712–10716.
- (50) Freddolino, P. L.; Kalani, M. Y.; Vaidehi, N.; Floriano, W. B.; Hall, S. E.; Trabanino, R. J.; Kam, V. W.; Goddard, W. A., III Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2736–2741.
- (51) Burgoyne, N. J.; Jackson, R. M. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006**, *22*, 1335–1342.
- (52) Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M. Predicting druggable binding sites at the protein-protein interface. *Drug Discovery Today* **2009**, *14*, 155–161.
- (53) Vaughan, C. K.; Buckle, A. M.; Fersht, A. R. Structural response to mutation at a protein-protein interface. *J. Mol. Biol.* **1999**, *286*, 1487–1506.
- (54) Bruncko, M.; Oost, T. K.; Belli, B. A.; Ding, H.; Joseph, M. K.; Kunzer, A.; Martineau, D.; McClellan, W. J.; Mitten, M.; Ng, S. C.; Nimmer, P. M.; Oltersdorf, T.; Park, C. M.; Petros, A. M.; Shoemaker, A. R.; Song, X.; Wang, X.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J. Med. Chem.* **2007**, *50*, 641–662.
- (55) Sleefs, B. E.; Czabotar, P. E.; Fairbrother, W. J.; Fairlie, W. D.; Flygare, J. A.; Huang, D. C.; Kersten, W. J.; Koehler, M. F.; Lessene, G.; Lowes, K.; Parisot, J. P.; Smith, B. J.; Smith, M. L.; Souers, A. J.; Street, I. P.; Yang, H.; Baell, J. B. 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.* **2011**, *54*, 1914–1926.
- (56) Sattler, M.; Liang, H.; Nettesheim, D.; Meadows, R. P.; Harlan, J. E.; Eberstadt, M.; Yoon, H. S.; Shuker, S. B.; Chang, B. S.; Minn, A. J.; Thompson, C. B.; Fesik, S. W. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **1997**, *275*, 983–986.
- (57) Manion, M. K.; O'Neill, J. W.; Giedt, C. D.; Kim, K. M.; Zhang, K. Y.; Hockenberry, D. M. Bcl-XL mutations suppress cellular sensitivity to antimycin A. *J. Biol. Chem.* **2004**, *279*, 2159–2165.
- (58) Edwards, C. P.; Fisher, K. L.; Presta, L. G.; Bodary, S. C. Mapping the intercellular adhesion molecule-1 and -2 binding site on the inserted domain of leukocyte function-associated antigen-1. *J. Biol. Chem.* **1998**, *273*, 28937–28944.
- (59) Huang, C.; Springer, T. A. A binding interface on the I domain of lymphocyte function-associated antigen-1 (LFA-1) required for specific interaction with intercellular adhesion molecule 1 (ICAM-1). *J. Biol. Chem.* **1995**, *270*, 19008–19016.
- (60) Kumar, M. D.; Gromiha, M. M. PINT: Protein-protein Interactions Thermodynamic Database. *Nucleic Acids Res.* **2006**, *34*, D195–198.
- (61) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1–9.
- (62) Lin, J.; Buettner, R.; Yuan, Y. C.; Yip, R.; Horne, D.; Jove, R.; Vaidehi, N. Molecular dynamics simulations of the conformational changes in signal transducers and activators of transcription, Stat1 and Stat3. *J. Mol. Graphics Modell.* **2009**, *28*, 347–356.
- (63) Buettner, R.; Corzano, R.; Rashid, R.; Lin, J.; Senthil, M.; Hedvat, M.; Schroeder, A.; Mao, A.; Herrmann, A.; Yim, J.; Li, H.; Yuan, Y. C.; Yakushijin, K.; Yakushijin, F.; Vaidehi, N.; Moore, R.; Gugiu, G.; Lee, T. D.; Yip, R.; Chen, Y.; Jove, R.; Horne, D.; Williams, J. C. Alkylation of cysteine 468 in Stat3 defines a novel site for therapeutic development. *ACS Chem. Biol.* **2011**, *6*, 432–443.