Structure
Ways & Means

Finding Protein Targets for Small Biologically Relevant Ligands across Fold Space Using Inverse Ligand Binding Predictions

Gang Hu,[1] Jianzhao Gao,[1] Kui Wang,[1,*] Marcin J. Mizianty,[3] Jishou Ruan,[1,2] and Lukasz Kurgan[3,*]
[1]School of Mathematical Sciences and LPMC
[2]State Key Laboratory of Medicinal Chemical Biology
Nankai University, Tianjin, People's Republic of China 300071
[3]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada
*Correspondence: wangkui@nankai.edu.cn (K.W.), lkurgan@ece.ualberta.ca (L.K.)
http://dx.doi.org/10.1016/j.str.2012.09.011

SUMMARY

Inverse ligand binding prediction utilizes a few protein-ligand (drug) complexes to predict other secondary therapeutic and off-targets of a given drug molecule on a proteomic scale. We adapt two binding site predictors, FINDSITE and SMAP, to perform the inverse predictions and evaluate them on over 30 representative ligands. Use of just one complex allows the identification of other protein targets; the availability of additional complexes improves the results. Both methods offer comparable quality when using three complexes with diverse proteins. SMAP is better when fewer complexes are available, while FINDSITE provides stronger predictions for smaller ligands. We propose a consensus that combines (and outperforms) the two complementary approaches implemented by FINDSITE and SMAP. Most importantly, we demonstrate that these methods successfully find distant targets that belong to structurally different folds compared to the proteins in the input complexes.

INTRODUCTION

A diverse repertoire of protein functions is carried through their interactions with other molecules (Rausell et al., 2010) including proteins, nucleic acids, peptides, and a variety of small molecules. These interactions have been investigated and summarized in the past two decades (Luscombe et al., 2001; Jones and Thornton, 2004; Ellis et al., 2007; Zhu et al., 2008; Chen and Kurgan, 2009; Zhang et al., 2010b). We focus on the interactions with small organic ligands, which are defined as organic molecules with less than 100 non-hydrogen atoms (Chen et al., 2011a). These ligands constitute a significant majority of drugs approved by the US Food and Drug Administration (FDA; Wishart et al., 2008) and they play important roles in modulation of protein-protein interactions (González-Ruiz and Gohlke, 2006; Casey et al., 2009; Gao and Skolnick, 2012).

Despite continuing accumulation of protein-ligand complexes in the Protein Data Bank (PDB; Berman et al., 2000), only a small portion of these interactions is known. This problem is addressed through computational methods that are built using the known protein-ligand complexes and applied to predict interactions for uncharacterized protein structures. The feasibility of such predictors is motivated by the fact that ligands and associated protein sequences and structures have co-evolved (Goh et al., 2000; Dupont et al., 2006), and thus similarity in sequence and/or structure can be used to infer interactions with other targets by the same or similar ligands (Xie et al., 2011a). These methods are of two types. The first category includes methods that find and rank ligand-binding pockets on the protein surface without targeting a specific ligand. These methods, which include SURFNET (Laskowski, 1995), LIGSITE^{csc} (Huang and Schroeder, 2006), Fpocket (Le Guilloux et al., 2009), MetaPocket (Huang, 2009), and Concavity (Capra et al., 2009), use a relatively simple geometry-driven approach, and, in the case of Q-SiteFinder (Laurie and Jackson, 2005) approximation of binding energy between the protein and a hypothetical ligand. The second category involves methods that predict binding pockets for specific ligands. They include FINDSITE (Brylinski and Skolnick, 2008), which is a threading-based approach that uses a library of known protein-ligand complexes, and SMAP (Xie and Bourne, 2008; Ren et al., 2010), which utilizes a profile-profile alignment to predict binding pockets from known protein-ligand complexes. These methods are used to implement modern virtual screening-based rational drug design protocols (Brylinski and Skolnick, 2010; Sukumar and Das, 2011).

Recent studies indicate that cross-reactivity of ligands with proteins occurs beyond global sequence and structure homologs (Xie and Bourne, 2008; Nobeli et al., 2009; Petrey et al., 2009; Zhang et al., 2010a). This means that the same ligand might bind to proteins that belong to substantially different folds and thus the ligand binding predictors should be equipped to work across the fold space. That observation is particularly important in the context of the inverse ligand binding predictions, where only a handful of protein-ligand complexes is used to predict other protein targets on a proteomic scale. This is in contrast to classical predictors that use many complexes to predict an individual protein target. The inverse

predictions are an important element of rational drug discovery protocols, where they are used to find off-targets of a given drug or drug candidate molecule based on its known interaction with the therapeutic target (Xie et al., 2011a). A number of attempts have been already made to find the off-targets for specific drugs, including HIV protease inhibitors (Specker et al., 2005), Comtan (Kinnings et al., 2009), cholesteryl ester transfer protein inhibitors (Xie et al., 2009a), and nelfinavir (Xie et al., 2011b), and on a larger scale (Ji et al., 2006; Keiser et al., 2009). Early methods used an inverse docking-based approach (Chen and Zhi, 2001; Ji et al., 2006; Yang et al., 2009), while more recent approaches are integrative, in a sense that they attempt to combine homology detection, structural bioinformatics, protein-ligand docking, molecular dynamics simulations, free energy calculations, and biologic network analysis (Xie et al., 2011a). Here, we concentrate on the modern structural bioinformatics methods, which are implemented as the inverse ligand binding predictors and offer a computationally (substantially) less intensive alternative for the inverse docking, especially when considering large, proteomic-scale target sets.

Although the inverse ligand binding predictors were successfully used to predict the off-targets, they were never comprehensively evaluated, in particular on a proteomic scale and to investigate their quality when predicting across non-homologous folds. We adapt two ligand-specific predictors of binding pockets: FINDSITE (Brylinski and Skolnick, 2008) and SMAP (Xie and Bourne, 2008; Ren et al., 2010), to work as the inverse ligand binding predictors. We select three representative biologically relevant small organic ligands, NAG, ADP, and PLM, to perform detailed evaluation on a proteomic scale on two types of well-designed ligand-specific benchmark data sets: a redundant data set that includes all known ligand-binding proteins, and a nonredundant data set that includes a subset of diverse (in both sequence and structure) ligand-binding targets. Both data sets also include proteins that are unlikely to bind a given ligand and we use SCOP hierarchy (Murzin et al., 1995; Andreeva et al., 2008) to evaluate predictive quality when finding distant (low homology) targets, i.e., targets that belong to different folds compared to the proteins in the input/template complexes. Motivated by differences in the underlying methodologies implemented in FINDSITE and SMAP, we also propose a consensus-based approach that aims to provide improved predictive quality.

Our results indicate that inverse ligand binding predictions are relatively accurate, even when just one input protein-ligand (drug) complex is used. As expected, availability of additional complexes with diverse proteins leads to improved predictions. We show that FINDSITE/SMAP performs well for smaller/larger ligands and that overall SMAP is better than FINDSITE when only one or two complexes are available. Based on a comprehensive test that uses a large and independent (from the three ligands used to design the consensus) set of 35 ligands, we show that the consensus of the two methods outperforms the individual predictors. Most importantly, we demonstrate that these three approaches are effective in finding structurally distant (from the proteins in the input complexes) protein targets.

EXPERIMENTAL PROCEDURES AND RESULTS

Selection of Representative Ligands and Construction of Benchmark Data sets

We select three representative biologically relevant small organic ligands to evaluate FINDSITE-, SMAP-, and consensus-based inverse ligand binding predictors. The selection criteria are that these ligands interact with sufficient number (to allow for statistically sound empirical evaluation) of nonredundant (both in sequence and structure spaces) targets for which complexes are available in the PDB, and that they represent major clusters of ligands in the PDB (they are dissimilar). We collect all biologically relevant small organic ligands (Dessailly et al., 2008) and their complexes in the PDB, reduce the set of target proteins for each ligand based on sequence and structure similarity, select and cluster the ligands with sufficient number of low similarity complexes, and choose three ligands, NAG, ADP, and PLM, that have a large number of complexes in the three resultant largest clusters. A detailed, step-by-step protocol to select these ligands is described in the Supplemental Experimental Procedures.

Empirical evaluation is performed on ligand-specific data sets, which are composed of one of two positive, ligand-binding, protein sets (redundant and non-redundant) and one negative set. The negative sets, one for each ligand, include proteins that are unlikely to bind the ligand. The first, redundant positive set includes all proteins collected from the PDB that are in complex with one of the three selected ligands. These are the proteins collected in step 1 of the ligand selection procedure. The second, nonredundant positive set is built using a subset of the redundant set with reduced sequence and structure similarity. This set includes proteins collected after step 4 of the ligand selection procedure, which means that the corresponding proteins are dissimilar at 25% sequence similarity and 0.4 structure similarity, which is measured using TM-score (Pandit and Skolnick, 2008). The negative sets are extracted as a subset of the culled PDB list generated by the PISCES server (Wang and Dunbrack, 2003) with proteins that are dissimilar to the proteins in the corresponding redundant positive sets and to other proteins that are known to interact with a given selected (or a similar) ligand. Specifically, using the PISCES server as of April 2011 we collected 2,214 proteins that have pairwise sequence similarity <25% and high-resolution structures (resolution <1.6 Å); we removed small proteins with <50 residues. Next, using BindingDB (Liu et al., 2007), we find ligands that are >90% similar to a given ligand and collect all their protein targets; in some cases, only their sequences are known. We combine these targets with all proteins from the nonredundant (structure-based) set for the same ligand, remove redundant proteins (leave one proteins from each set of proteins with identical sequences) and proteins with contact number <70. The resulting set represents proteins that are known to bind to a given selected ligand. Next, we align each sequence from the PISCES server to all sequences of the ligand-binding targets and we add a given chain to the negative set if its similarity to every target sequence is below 30%. As a result, the counts of protein chains in the redundant positive sets for NAG, ADP, and PLM are 1,753, 1,622, and 85, respectively; in the nonredundant set they are 59, 53, and 15, respectively; and in negative set they are 904, 607, and 177, respectively. We perform evaluations on two benchmark sets for each ligand: one that combines the redundant positive set and the negative set, and another with the nonredundant positive set and the negative set. The data sets are available at http://biomine.ece.ualberta.ca/ILbind/.

Evaluation of Predictive Quality

We rank all proteins according to the output of a given inverse ligand binding prediction method and plot the receiver operating characteristic (ROC) curves using this ranking. We count the number of proteins that bind the input ligand versus the number of proteins that not to bind among the top n proteins, when n varies from 0 to N, which is the number of proteins in the benchmark set. We use the area under the ROC curve (AUC) to evaluate the predictive quality.

We also evaluate significance of differences in AUC values for a given pair of predictors based on their paired results across a given set of ligands and templates. First, we determine normality of a given AUC measurement with the Anderson-Darling test at the 0.05 significance. For normal distributions, we use a paired t test; otherwise, we use the Wilcoxon rank-sum test.

Table 1. Average AUC Values for FINDSITE- and SMAP-Based Inverse Ligand Binding Predictors

		FINDSITE		SMAP	
Ligand	No. of Templates	Redundant	Nonredundant	Redundant	Nonredundant
NAG	5 individual templates	**0.65** ± 0.035	0.63 ± 0.036	0.60 ± 0.019	0.60 ± 0.011
	10 two-template sets	**0.69** ± 0.014	0.68 ± 0.018	0.59 ± 0.010	0.59 ± 0.006
	10 three-template sets	**0.70** ± 0.010	0.70 ± 0.013	0.59 ± 0.006	0.59 ± 0.003
ADP	5 individual templates	0.67 ± 0.012	0.77 ± 0.022	**0.75** ± 0.021	0.79 ± 0.011
	10 two-template sets	0.70 ± 0.008	0.82 ± 0.013	**0.79** ± 0.011	0.81 ± 0.006
	10 three-template sets	0.71 ± 0.005	0.84 ± 0.009	**0.81** ± 0.007	0.83 ± 0.002
PLM	5 individual templates	0.63 ± 0.024	0.53 ± 0.016	**0.71** ± 0.029	0.68 ± 0.034
	10 two-template sets	0.67 ± 0.014	0.56 ± 0.012	**0.76** ± 0.008	0.70 ± 0.011
	10 three-template sets	0.71 ± 0.012	0.59 ± 0.010	**0.79** ± 0.008	0.71 ± 0.008
Average	individual templates	0.65 ± 0.012	0.64 ± 0.070	0.69 ± 0.045	0.69 ± 0.055
	two-template sets	0.69 ± 0.009	0.69 ± 0.075	0.71 ± 0.062	0.70 ± 0.063
	three-template sets	0.71 ± 0.003	0.71 ± 0.072	0.73 ± 0.070	0.71 ± 0.069

The average (across the corresponding sets of templates) AUC values ± the corresponding standard errors for the FINDSITE- and SMAP-based inverse ligand binding predictors on the redundant and non-redundant benchmark data sets for the selected three representative small organic ligands: NAG, ADP, and PLM. The last "average" row shows AUC values that are averaged across all three ligands. The best results on the redundant data set for each ligand and number of templates are shown in bold.
See also Figure S1.

Inverse Ligand Binding Predictors

We adapt FINDSITE and SMAP to perform inverse ligand binding prediction. The classical implementation of these methods uses a large library of diverse protein-ligand complexes to perform predictions. Here, we reduce their libraries to the selected (small) set complexes with a given ligand. Moreover, we analyze the outputs of these methods to select one output index for each method that provides the best predictive quality.

FINDSITE is based on binding-site similarity among superimposed groups of template structures identified using threading into the input/query protein structure (Brylinski and Skolnick, 2008). FINDSITE identifies binding residues using a consensus of the binding pockets in the selected superimposed threading templates. In our scenario, FINDSITE predicts binding pocket(s) on the query protein using one (a few) template complex with a given ligand. It returns detailed information about the predicted pocket(s) that allow ranking predictions across different query proteins. We identified seven indices/features that could be used to perform ranking: (1) TM-score with the target template protein structure; (2) RMSD of the C_α atoms in the aligned (between query and template structures) region; (3) alignment length (number of residues aligned with the template protein structure by fr-TM-Align (Pandit and Skolnick, 2008); (4) fraction of templates that share the detected pocket; (5) sequence identity calculated over the residues aligned by the fr-TM-Align; (6) number of predicted binding residues; and (7) number of detected pockets.

SMAP is based on a sequence order independent profile-profile alignment (SOIPPA) that was designed to find evolutionary and functional relationships across the fold space (Xie and Bourne, 2007, 2008; Xie et al., 2009b; Ren et al., 2010). SMAP utilizes a shape descriptor to characterize the structure of the protein template and the SOIPPA algorithm to detect and align similar pockets between the query and template proteins. Given a template complex and a query protein, SMAP outputs seven indices/features that can be used to rank predictions: (1) local structural alignment between ligand-binding sites in the query and template proteins, which is quantified with the number of predicted binding residues; (2) raw score; (3) p value; (4) volume coverage of binding pockets on the template (called target cover); (5) volume coverage of binding pockets on the query protein (called query cover); (6) Tanimoto coefficient; and (7) RMSD between the query and template proteins.

To comparatively evaluate predictive quality of the different outputs generated by FINDSITE and SMAP, we perform five predictions for each of the three selected ligands using a template (different each time) that is randomly selected from the corresponding nonredundant positive set; one template always corresponds to the largest cluster. These templates are substantially different in both sequence and structure; their PDB accession numberss are 1ZAG, 1NQL, 2CIY, 2WFO, and 3C45 for NAG; 1GZF, 3C9U, 1CQI, 2ZPA, and 3CNZ for ADP; and 2IU8, 3LSJ, 2IES, 3FYS, and 2G87 for PLM. These templates are also used in the subsequent sections. We predict the binding proteins on the redundant benchmark sets and average the AUC values across the five templates for each of the three selected ligands. We calculate seven averaged AUCs for SMAP and another seven for FINDSITE using each of their output features separately (see Figure S1 available online). The overall AUC values, which are averaged over the three ligands and the corresponding 15 templates, show that the top-performing index for SMAP is the raw score and for FINDSITE is the alignment length. The raw scores are rescaled to calculate the p values, so in fact these two indices are strongly correlated and provide virtually identical predictive quality. We observe that raw scores outperform all other SMAP-based features for all three ligands. They improve the overall AUC by 0.07 when compared with the second best SMAP-derived index, which is the predicted number of binding residues. Similarly, the alignment length provides the best results for the three ligands when considering the FINDSITE generated indices and it improves the overall AUC by 0.04 compared to the second best TM-Score. Therefore, we use the raw scores and the alignment length values to predict binding proteins using SMAP and FINDSITE, respectively. Besides the high overall AUC, this choice is motivated by the consistency of these two selected indices, which work equally well over the three diverse/representative ligands.

Evaluation of Inverse Ligand Binding Predictions with FINDSITE and SMAP

We evaluate the FINDSITE- and SMAP-based inverse ligand binding predictors on our benchmark data sets for the three selected ligands. For each ligand we generate five predictions using the five templates (listed in the Inverse Ligand Binding Predictors section) that are substantially different in both sequence and structure. We rank all proteins according to the best output of a given inverse ligand binding prediction method (alignment length generated by FINDSITE and raw score outputted by SMAP) and calculate the AUC values to evaluate the predictive quality. We predict the binding proteins on both the redundant and nonredundant benchmark data sets and report average AUC values across the five templates for each of the three selected ligands. Moreover, we empirically evaluate whether use of additional templates would lead to improved predictive performance. To this end, we generate predictions when using all ten combinations of two and ten combinations of three templates from the set of five templates and report the corresponding average AUCs. The results are summarized in Table 1.

Both FINDSITE and SMAP can be used to provide well-performing inverse ligand binding predictions. Using just one template, the average AUCs over the three representative ligands are 0.65 for FINDSITE and 0.69 for SMAP on the redundant data set, and 0.64 and 0.69 on the nonredundant data set, respectively. Although overall SMAP outperforms FINDSITE, the latter method is more accurate on NAG while SMAP is better on ADP and PLM. This might be explained by the size of ligands; of three ligands, NAG has only 15 heavy atoms, while ADP and PLM are larger and have 27 and 18 heavy atoms, respectively. The performance of FINDSITE stays relatively similar across the three ligands, while SMAP's performance varies more and correlates with the ligand size. FINDSITE predicts the binding pockets through threading while SMAP uses profile-profile alignments. While these alignments seem to outperform threading when ligands are larger (there are more binding residues to align, particularly in the case of the largest ADP), they may not work well when the ligand (and the corresponding pocket) are relatively small, which is when threading still predicts relatively well.

As expected, the AUC values increase when the number of templates increases. This trend is true for both redundant and nonredundant data sets. Use of additional templates, which are diverse in sequence and structure, provides more information that is used to find targets that can be missed with fewer templates. We observe that while SMAP is better when fewer template complexes are available, both methods provide similar results when three templates are available. We evaluate the significance of differences between SMAP and FINDSITE when using one, two, and three templates across the three ligands. The corresponding p values are 0.16 (one template), 0.19 (two templates), and 0.30 (three templates) on the redundant data sets; and 0.16, 0.68, and 0.93 on the nonredundant data sets, respectively.

Table 2. Average AUC Values for the Consensus, FINDSITE-, and SMAP-Based Inverse Ligand Binding Predictors

Ligand	FINDSITE	SMAP	ILbind
NAG	0.646 ± 0.035	0.600 ± 0.019	**0.654** ± 0.031
ADP	0.668 ± 0.012	0.750 ± 0.021	**0.756** ± 0.016
PLM	0.632 ± 0.029	0.708 ± 0.024	**0.714** ± 0.021
35 independent ligands	0.666 ± 0.017	0.685 ± 0.024	**0.713** ± 0.022

The average (across the corresponding sets of five templates) AUC values ± the corresponding standard errors for the consensus-, FINDSITE-, and SMAP-based inverse ligand binding predictors on the redundant benchmark data sets for the selected three representative small organic ligands: NAG, ADP, and PLM. The last row shows the average AUC values ± the corresponding standard errors across the 35 ligands in the independent test set; ILbind predictor was build using the three ligands (NAG, ADP, and PLM) and tested on the 35 different ligands. The best results for each ligand are shown in bold.
See also Table S1 and Figure S2.

Proposed Consensus Approach

We designed a consensus-based approach motivated by the fact that FINDSITE and SMAP perform predictions using different underlying methodologies. The consensus combines selected outputs of SMAP and FINDSITE using an ensemble of machine learning predictors (Figure S2A).

We design and validate our method using cross validation, in which predictions on a given ligand are generated using a model established using outputs generated by SMAP and FINDSITE for the other two ligands; this is repeated three times, each time predicting a different ligand. This protocol prevents potential overfitting into the input data, i.e., training data (concerning a given ligand) and test data (concerning the other two ligands) are independent. The design includes two steps. First, we select suitable inputs from among the outputs generated by SMAP and FINDSITE, and next we build a consensus of predictors using these selected inputs.

We use support vector machine (SVM) predictors, which are widely used in recent related applications, such as prediction of catalytic residues (Zhang et al., 2008) and binding residues for ATP (Chauhan et al., 2009; Chen et al., 2011b), ADP, AMP, GTP, GDT (Chauhan et al., 2010; Chen et al., 2012), FAD (Mishra and Raghava, 2010), NAD (Ansari and Raghava, 2010), and heme (Liu and Hu, 2011). In our final design, which is deployed at http://biomine.ece.ualberta.ca/ILbind/, we combine 15 SVMs, where each is built using data concerning one of the five randomly selected (from the nonredundant data set) templates for each of the three ligands. In our cross validation-based empirical validation, we use ten SVMs to predict targets for a given ligand; they are built using data concerning the two remaining ligands, using five templates for each. The score generated by our method is computed as an average of the outputs generated by the ten (or 15) SVMs. The SVMs use linear kernel function with the complexity constant C = 1. The value of C was established by cross validation on the training data sets, where the test data were set aside.

The inputs to the SVM are selected among all 14 indices generated by SMAP and FINDSITE. The selection is performed separately for each of the three ligands. For a given ligand we follow a two-step process: (1) we sort the 14 indices by the average AUC values using the training data set (data concerning the remaining two ligands); and (2) we perform wrapper-based greedy best first search in which we start with the top ranked feature (index) and try to add one feature at the time by scanning the ranked list once; a given index is added into the feature set if its addition increases AUC value when compared with the feature set without this index, where AUC evaluates SVM-based prediction within the training data set (data concerning one training ligand is used to predict the other and vice versa). As a result, we obtain three feature sets for PLM (using training data from NAG and ADP), NAG (using data from ADP and PLM), and ADP (using data from PLM and NAG), respectively. The selected features for PLM are alignment length (from FINDSITE) and raw score (SMAP); for NAG are raw score (SMAP), alignment length (FINDSITE), TM-score (FINDSITE), and sequence identity (FINDSITE); and for ADP are alignment length (FINDSITE) and raw score (SMAP). The alignment length and raw score are included in all three selected feature sets and thus we use these two indices/features as the input to our SVM-based ensemble. These two indices have also the highest AUC when used independently (without SVM) to predict the binding (Figure S1). We compare the performance of these two features with the performance using all 14 features, i.e., combined outputs of FINDSITE and SMAP (Table S1A). The use of all versus the selected two features results in a lower predictive performance across all three ligands, with average AUCs of 0.67 and 0.71, respectively. We also considered additional features that are not related to the outputs of FINDSITE and SMAP. Because similarity in the sequence and the structure of the backbone are considered by FINDSITE and SMAP, we tried descriptors of the overall shape of the protein fold. They include radius of gyration, radius of cross section, coefficient of compactness, and normalized radius of gyration (Ivankov et al., 2009). We calculated the AUCs for these four features across the three selected ligands and the corresponding values are 0.52, 0.52, 0.56, and 0.51, respectively. As expected, these results are inferior to the outputs of SMAP and FINDITE. Their predictive value is marginally better than random and thus we did not consider them further when designing inverse ligand binding.

Evaluation of Consensus-Based Inverse Ligand Binding Predictions

The consensus-based inverse ligand binding predictor (ILbind) combines selected outputs of SMAP and FINDSITE using an ensemble of SVMs; a web server and a standalone version that implement ILbind are available at http://biomine.ece.ualberta.ca/ILbind/. The consensus-derived predictions are generated using a cross-validation protocol, in which predictions for a given ligand are computed using the SVM models that are developed using the other two ligands. We report the AUC values averaged over the five individual templates on the redundant benchmark data set for the three selected ligands; see Table 2. The AUCs for FINDSITE and SMAP are based on their best outputs, alignment length and raw score, respectively. The AUC for the consensus is calculated using the average of the scores generated by the ensemble of SVMs.

The overall AUCs averaged over the three ligands are 0.65, 0.69, and 0.71 for the FINDSITE-, SMAP-, and consensus-based predictors. The corresponding averaged (over the three ligands) ROC curves are shown in Figure S2B. The consensus outperforms its input SMAP- and FINDSITE-based methods on all three ligands. Although the improvements have relatively small magnitude compared to the best performing method on each ligand, they are consistent in contrast to the individual methods that outperform each other on different

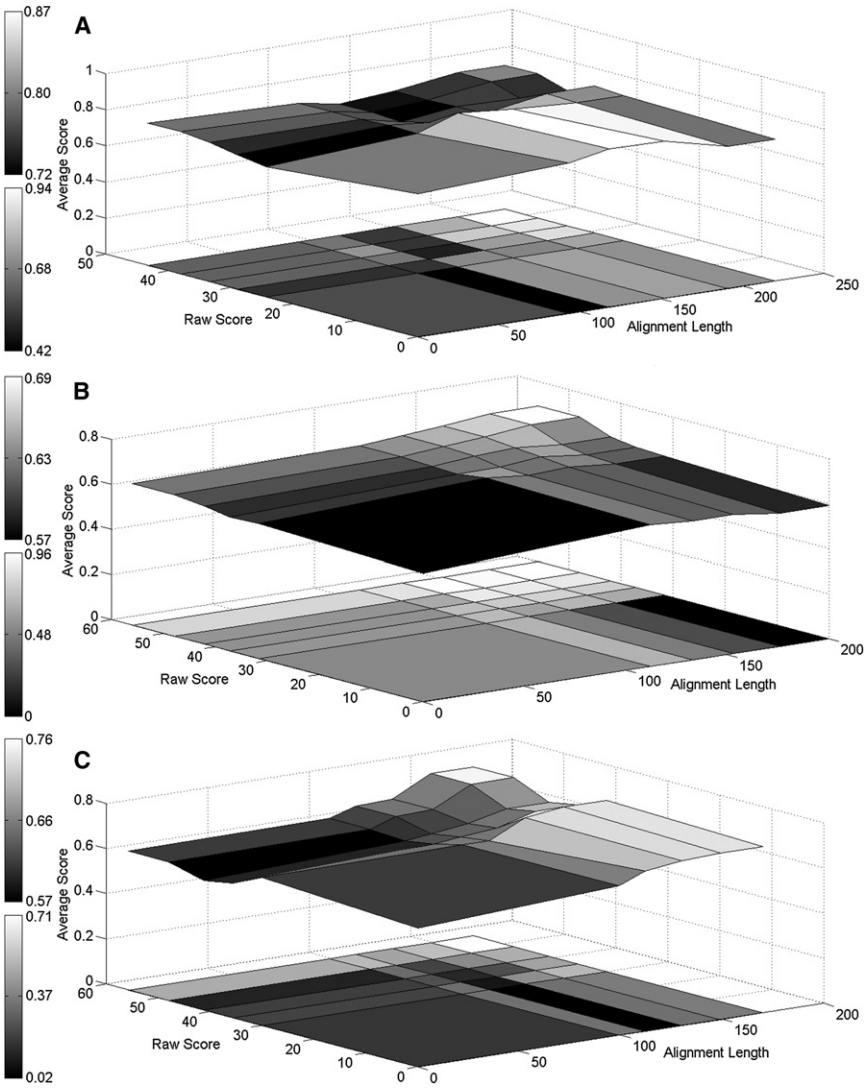

Figure 1. Values of Inputs of Consensus and the Probability of Binding

Relation between the values of the inputs of the consensus (alignment length and raw score shown on the x-y plane) and the probability of binding (color coded on the x-y plane) together with the scores generated by the consensus method (shown as a surface above the x-y plane) for NAG (panel A), ADP (panel B), and PLM (panel C).
See also Figure S1.

ligands. The differences in AUCs (across the corresponding paired results for the five templates and the three ligands) between the consensus and FINDSITE and between the consensus and SMAP are statistically significant at p values equal to 0.001 and 0.06, respectively. This means that the ensemble of SVMs leverages well-performing predictions, even if they come from one of the two methods. Overall, our consensus demonstrates that the input methods are complementary and that this can be exploited to improve predictive quality.

We also investigate the relation between the inputs of our consensus and the underlying probability of binding, which is quantified by a fraction of binding proteins, and show how this relation is modeled by our ensemble of SVMs. For each of the three ligands, we create a five-by-five size grid that divides the proteins in the redundant data set into approximately equal-sized subsets using values of the raw score and alignment length; we sort proteins separately by each index to determine endpoints of the five intervals so that the number of proteins in each interval is the same. Data for each ligand are represented by two plots: one on the x-y plane, where colors of the nodes on the grid represent proportion of native binding proteins (darker colors correspond to lower fractions of binding proteins); and the other is a surface where colors denote the average scores generated by the consensus method (darker colors correspond to lower scores); see Figure 1. Overall, the plots suggest that proteins with longer alignment length or larger raw score are more likely to bind the ligand, which agrees with the interpretation of these indices. However, this is not always the case. For instance, Figure 1B (which considers ADP) shows that proteins with high alignment length may include those that are unlikely to bind, depending on their raw scores. The same is true when considering proteins with high raw scores and the entire spectrum of the alignment length values. Combining both indices leads to better discrimination. Figure 1B demonstrates that proteins with relatively high raw score and alignment length (the far corner on the x-y plane) are more likely to bind. Similar trends can be observed for the other two ligands. These observations suggest that SMAP and FINDSITE provide complementary predictions and explain why the consensus outperforms the two individual methods. The surfaces on the z-axis, which show outputs of the consensus, reveal that predictions from the SVMs relatively well follow the native probability of binding. The color patterns of these surfaces are similar to the patterns on the x-y planes, which means that the scores generated by the consensus provide good predictive quality.

Evaluation on Independent Set of Ligands

Step 4 of the ligand selection procedure returns 38 ligands with a sufficient number of low similarity complexes. We use 35 of them, excluding the three selected ligands (NAG, ADP, and PLM), to build an independent, with respect to the data used to design ILbind, set of ligands that is used to evaluate SMAP-based, FINDSITE-based, and consensus-based ILbind predictors. We randomly selected a single complex for each of these ligands as the template for FINDSITE and SMAP. Next, we collected positive and negative data sets for each ligand and use them to evaluate predictions from FINDSITE, SMAP, and ILbind. The positive data set contains all complexes with this ligand in the PDB and the negative data set is constructed in the same as for the three selected ligands, i.e., all proteins collected from PISCES server with less than 0.3 sequence similarity to every protein in positive data set of this ligand. As a result, we have 35 positive and 35 negative data sets.

Consistent with the evaluations on the three ligands, the AUCs for FINDSITE and SMAP are calculated using the alignment length and raw score, respectively. The AUCs of the consensus-based ILbind are based on the average probability outputted by the SVM models. The overall AUCs averaged over the 35 ligands are 0.666, 0.685, and 0.713 for the FINDSITE, SMAP, and ILbind, respectively; see Table 2. Detailed results are given in Table S1B and the corresponding average (across the 35 ligands) ROC curves are shown in Figure S2C. The evaluation of the statistical significance of the paired differences over the 35 ligands reveals that ILbind outperforms SMAP and FINDSITE with p values at 0.001 and 0.0003, respectively. The SMAP is better than FINDSITE with a p value of 0.25. These results are consistent with the results on the three selected ligands. Moreover, they also confirm our finding that the performance of these methods is relative to the size of the ligand (Table S1C). We divided the 35 ligands into four equally sized size-based bins: nine small ligands with the number of heavy atoms $\leq$ 14,

Table 3. Average AUC Values for Consensus-Based ILbind, FINDSITE-, and SMAP-Based Inverse Ligand Binding Predictors

		Test on the Same		Test on the Different		
Ligand	Methods	Class	Fold	Classes	Folds	Test on all SCOP-Annotated Proteins
NAG	ILbind	0.66 ± 0.04	1.00 ± 0.00	0.67 ±0.02	0.65 ± 0.04	0.66 ± 0.03
	FINDSITE	0.65 ± 0.07	1.00 ± 0.00	0.65 ± 0.03	0.63 ± 0.05	0.65 ± 0.03
	SMAP	0.64 ± 0.03	1.00 ± 0.00	0.61 ± 0.02	0.60 ± 0.02	0.62 ± 0.02
ADP	ILbind	0.83 ± 0.02	0.82 ± 0.14	0.72 ± 0.01	0.74 ± 0.01	0.76 ± 0.02
	FINDSITE	0.74 ± 0.04	0.85 ± 0.15	0.65 ± 0.02	0.67 ± 0.02	0.68 ± 0.02
	SMAP	0.80 ± 0.03	0.83 ± 0.10	0.72 ± 0.01	0.72 ± 0.01	0.75 ± 0.02
PLM	ILbind	0.95 ± 0.04	1.00 ± 0.00	0.79 ± 0.05	0.79 ± 0.02	0.82 ± 0.03
	FINDSITE	0.87 ± 0.04	1.00 ± 0.00	0.78 ± 0.03	0.76 ± 0.03	0.80 ± 0.04
	SMAP	0.77 ± 0.22	1.00 ± 0.00	0.75 ± 0.06	0.76 ± 0.03	0.80 ± 0.05

The average (across the corresponding sets of five templates) AUC values ± the corresponding standard errors for the consensus-based ILbind, FINDSITE-, and SMAP-based inverse ligand binding predictors on the benchmark data sets annotated using SCOP hierarchy for the selected three representative small organic ligands: NAG, ADP, and PLM. The tests were performed on a subset of benchmark proteins that belong to the same/different SCOP fold and class when compared with the fold and class of a given template protein. The right column includes results on all SCOP-annotated proteins.

nine small-medium ligands with the number of atoms between 15 and 26, eight medium-large with sizes between 27 and 31, and nine large ligands with more than 31 heavy atoms. We find that FINDSITE outperforms the other methods, including a statistically significant improvement over SMAP with a p value of 0.02, on the small ligands. At the same time, SMAP and ILbind outperform FINDSITE on the large ligands, with p values at or below 0.01. Finally, ILbind provides the highest average AUC values for the medium-sized ligands and matches SMAP for the large ligands.

Inverse Ligand Binding Predictions across the Fold Space

As discussed in a number of studies (Xie and Bourne, 2008; Nobeli et al., 2009; Petrey et al., 2009; Zhang et al., 2010a), the same ligand may have protein partners that belong to substantially different folds. This motivates an evaluation of the ability of the considered inverse ligand binding predictors to find structurally distant (belonging to a substantially different fold compared to the template proteins) binding proteins.

To this end, we constructed four subsets of the redundant data sets based on the SCOP annotations (Murzin et al., 1995; Andreeva et al., 2008), with the same or different SCOP classes or SCOP folds when compared with the template protein; proteins that lack SCOP annotations were removed from this evaluation. The SCOP-annotated benchmark data sets include 957 positive and 383 negative proteins for NAG, 773 positive and 250 negative for ADP, and 37 positive and 75 negative for PLM. For each template complex that constitutes input to a given inverse ligand binding prediction method, we constructed the four subsets of the SCOP-annotated benchmark data sets. The first subset includes the proteins that are in the same SCOP class as the template protein. The second subset includes the proteins that are in different SCOP classes compared with the class of the template protein. Analogously, the third (fourth) subsets include proteins from the same (different) SCOP fold compared to the fold of the input template proteins.

Table 3 reports the AUC values averaged over the five individual templates for the four SCOP-annotated subsets and all SCOP-annotated proteins and for each of the three selected ligands: NAG, ADP, and PLM. The key observation is that all three approaches relatively accurately find structurally distant (sharing low homology with the template protein) binding proteins. When focusing on finding targets that belong to a different SCOP fold FINDSITE-based predictor obtains AUCs between 0.63 and 0.76, SMAP-based method between 0.6 and 0.76, and our consensus between 0.65 and 0.79, depending on the ligand. To compare, the overall predictive quality (using all SCOP-annotated proteins) across the three methods is somewhat similar and ranges between 0.62 and 0.82. The predictions for proteins that are in the same fold or class as the template are, as expected, characterized by relatively high predictive performance, with AUCs between 0.82 and 1 in case of the SCOP fold. We also observe consistent improvements, across different levels of homology and different ligands, offered by the consensus-based ILbind method. Most importantly, these results suggest that the three considered methods can be used to relatively accurately predict the off-targets (other binding proteins) that share low homology with the template protein. This conclusion justifies the use of these inverse ligand binding predictors on a proteomic scale, even if only a few template complexes are available to them.

DISCUSSION

Effective inverse ligand binding predictors would be helpful for identifying the off-targets of a specific ligand/drug on a proteomic scale. We provide comparative empirical evaluations of the predictive quality of these methods. We adapt two recently developed binding site predictors, FINDSITE and SMAP, to perform the inverse ligand binding prediction and evaluate their predictive quality for three representative small biologically relevant ligands using well-designed benchmark data sets. We show that both approaches offer certain advantages. FINDSITE seems to provide better predictive quality for smaller ligands, while SMAP performs better for bigger ligands and when fewer complexes with a given ligand are available. Our results demonstrate that availability of additional structurally/in-sequence diverse templates leads to improvements and that both methods provide similar predictive quality when at least three templates are used. We also propose a consensus method that combines FINDSITE and SMAP using an ensemble of SVMs. This consensus is empirically shown to provide improved predictive performance when compared with FINDSITE and SMAP. Our approach leverages good-quality predictions, even if they come from one of the two methods; this comes from the fact that it provides improvements across the three selected ligands and across the set of the 35 independent ligands, while the two individual predictors outperform each other on different ligands. Most importantly, we empirically demonstrate that the three considered methods relatively accurately predict the off-targets that share low homology with the template protein(s). This interesting conclusion motivates the use of these inverse ligand binding predictors on a proteomic scale where some of the protein targets share low similarity with the template complexes. As example applications, these methods could be used to improve

current virtual screening-based rational drug design protocols and to perform protein function annotations across the fold space.

The predictive quality offered by the considered approaches is far from being perfect. However, we believe that it is relatively good considering that there are a number of factors that limit the ability of these methods to find certain targets. The target protein may undergo a large conformation change upon binding (Gunasekaran and Nussinov, 2007) and such binding events cannot be captured by these methods. Moreover, some ligands bind at the interface between multiple protein chains or proteins (Gao and Skolnick, 2012), which is virtually impossible to predict when only a single template chain (or a few templates) is available. These issues should be addressed by a new generation of the inverse ligand binding predictors.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2012.09.011.

ACKNOWLEDGMENTS

We thank Drs. Lei Xie and Michal Brylinski for advice on using their software. This work was supported in part by National Science Foundation of China (NSFC) Grants 31050110432 and 31150110577 to L.K. and J.R., Discovery Grant 298328 from NSERC (National Science and Engineering Research Council) Canada to L.K., and National Science Foundation of China (NSFC) Grants 11101226 and 20836005 to G.H. and K.W. J.R. was also supported by the International Development Research Center, Ottawa, Canada. (No. 104519-010) and Tianjin science and technology support project 08ZCHHZ00200. G.H., K.W., and J.G. were supported by the Fundamental Research Funds for the Central Universities Grants 65011441, 65011521, and 65011491. M.J.M. was supported by the Izaak Walton Killam Memorial Scholarship.

Received: June 7, 2012
Revised: August 14, 2012
Accepted: September 16, 2012
Published: November 6, 2012

REFERENCES

Andreeva, A., Howorth, D., Chandonia, J.-M., Brenner, S.E., Hubbard, T.J.P., Chothia, C., and Murzin, A.G. (2008). Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res. *36*(Database issue), D419–D425.

Ansari, H.R., and Raghava, G.P. (2010). Identification of NAD interacting residues in proteins. BMC Bioinformatics *11*, 160.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Brylinski, M., and Skolnick, J. (2008). A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc. Natl. Acad. Sci. USA *105*, 129–134.

Brylinski, M., and Skolnick, J. (2010). Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening. J. Chem. Inf. Model. *50*, 1839–1854.

Capra, J.A., Laskowski, R.A., Thornton, J.M., Singh, M., and Funkhouser, T.A. (2009). Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput. Biol. *5*, e1000585.

Casey, F.P., Pihan, E., and Shields, D.C. (2009). Discovery of small molecule inhibitors of protein-protein interactions using combined ligand and target score normalization. J. Chem. Inf. Model. *49*, 2708–2717.

Chauhan, J.S., Mishra, N.K., and Raghava, G.P. (2009). Identification of ATP binding residues of a protein from its primary sequence. BMC Bioinformatics *10*, 434.

Chauhan, J.S., Mishra, N.K., and Raghava, G.P. (2010). Prediction of GTP interacting residues, dipeptides and tripeptides in a protein from its evolutionary information. BMC Bioinformatics *11*, 301.

Chen, K., and Kurgan, L. (2009). Investigation of atomic level patterns in protein—small ligand interactions. PLoS ONE *4*, e4473.

Chen, K., Mizianty, M.J., Gao, J., and Kurgan, L. (2011a). A critical comparative assessment of predictions of protein-binding sites for biologically relevant organic compounds. Structure *19*, 613–621.

Chen, K., Mizianty, M.J., and Kurgan, L. (2011b). ATPsite: sequence-based prediction of ATP-binding residues. Proteome Sci. *9*(*Suppl 1*), S4.

Chen, K., Mizianty, M.J., and Kurgan, L. (2012). Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. Bioinformatics *28*, 331–341.

Chen, Y.Z., and Zhi, D.G. (2001). Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins *43*, 217–226.

Dessailly, B.H., Lensink, M.F., Orengo, C.A., and Wodak, S.J. (2008). LigASite—a database of biologically relevant binding sites in proteins with known apo-structures. Nucleic Acids Res. *36*(Database issue), D667–D673.

Dupont, C.L., Yang, S., Palenik, B., and Bourne, P.E. (2006). Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. Proc. Natl. Acad. Sci. USA *103*, 17822–17827.

Ellis, J.J., Broom, M., and Jones, S. (2007). Protein-RNA interactions: structural analysis and functional classes. Proteins *66*, 903–911.

Gao, M., and Skolnick, J. (2012). The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. Proc. Natl. Acad. Sci. USA *109*, 3784–3789.

Goh, C.S., Bogan, A.A., Joachimiak, M., Walther, D., and Cohen, F.E. (2000). Co-evolution of proteins with their interaction partners. J. Mol. Biol. *299*, 283–293.

González-Ruiz, D., and Gohlke, H. (2006). Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. Curr. Med. Chem. *13*, 2607–2625.

Gunasekaran, K., and Nussinov, R. (2007). How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. J. Mol. Biol. *365*, 257–273.

Huang, B. (2009). MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS *13*, 325–330.

Huang, B., and Schroeder, M. (2006). LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. BMC Struct. Biol. *6*, 19.

Ivankov, D.N., Bogatyreva, N.S., Lobanov, M.Y., and Galzitskaya, O.V. (2009). Coupling between properties of the protein shape and the rate of protein folding. PLoS ONE *4*, e6476.

Ji, Z.L., Wang, Y., Yu, L., Han, L.Y., Zheng, C.J., and Chen, Y.Z. (2006). In silico search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction. Toxicol. Lett. *164*, 104–112.

Jones, S., and Thornton, J.M. (2004). Searching for functional sites in protein structures. Curr. Opin. Chem. Biol. *8*, 3–7.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. Nature *462*, 175–181.

Kinnings, S.L., Liu, N., Buchmeier, N., Tonge, P.J., Xie, L., and Bourne, P.E. (2009). Drug discovery using chemical systems biology: repositioning the

 ©2012 Elsevier Ltd All rights reserved

safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS Comput. Biol. *5*, e1000423.

Laskowski, R.A. (1995). SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graph. *13*, 323–330, 307–308.

Laurie, A.T., and Jackson, R.M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics *21*, 1908–1916.

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

Liu, R., and Hu, J. (2011). HemeBIND: a novel method for heme binding residue prediction by combining structural and sequence information. BMC Bioinformatics *12*, 207.

Liu, T., Lin, Y., Wen, X., Jorissen, R.N., and Gilson, M.K. (2007). BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. *35*(Database issue), D198–D201.

Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. *29*, 2860–2874.

Mishra, N.K., and Raghava, G.P. (2010). Prediction of FAD interacting residues in a protein from its primary sequence using evolutionary information. BMC Bioinformatics *11*(*Suppl 1*), S48.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Nobeli, I., Favia, A.D., and Thornton, J.M. (2009). Protein promiscuity and its implications for biotechnology. Nat. Biotechnol. *27*, 157–167.

Pandit, S.B., and Skolnick, J. (2008). Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. BMC Bioinformatics *9*, 531.

Petrey, D., Fischer, M., and Honig, B. (2009). Structural relationships among proteins with different global topologies and their implications for function annotation strategies. Proc. Natl. Acad. Sci. USA *106*, 17377–17382.

Rausell, A., Juan, D., Pazos, F., and Valencia, A. (2010). Protein interactions and ligand binding: from protein subfamilies to functional specificity. Proc. Natl. Acad. Sci. USA *107*, 1995–2000.

Ren, J., Xie, L., Li, W.W., and Bourne, P.E. (2010). SMAP-WS: a parallel web service for structural proteome-wide ligand-binding site comparison. Nucleic Acids Res. *38*(Web Server issue), W441–W444.

Specker, E., Böttcher, J., Lilie, H., Heine, A., Schoop, A., Müller, G., Griebenow, N., and Klebe, G. (2005). An old target revisited: two new privileged skeletons and an unexpected binding mode for HIV-protease inhibitors. Angew. Chem. Int. Ed. Engl. *44*, 3140–3144.

Sukumar, N., and Das, S. (2011). Current trends in virtual high throughput screening using ligand-based and structure-based methods. Comb. Chem. High Throughput Screen. *14*, 872–888.

Wang, G., and Dunbrack, R.L., Jr. (2003). PISCES: a protein sequence culling server. Bioinformatics *19*, 1589–1591.

Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. *36*(Database issue), D901–D906.

Xie, L., and Bourne, P.E. (2007). A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. BMC Bioinformatics *8* (*Suppl 4*), S9.

Xie, L., and Bourne, P.E. (2008). Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. Proc. Natl. Acad. Sci. USA *105*, 5441–5446.

Xie, L., Li, J., Xie, L., and Bourne, P.E. (2009a). Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. PLoS Comput. Biol. *5*, e1000387.

Xie, L., Xie, L., and Bourne, P.E. (2009b). A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. Bioinformatics *25*, i305–i312.

Xie, L., Xie, L., and Bourne, P.E. (2011a). Structure-based systems biology for analyzing off-target binding. Curr. Opin. Struct. Biol. *21*, 189–199.

Xie, L., Evangelidis, T., Xie, L., and Bourne, P.E. (2011b). Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. PLoS Comput. Biol. *7*, e1002037.

Yang, L., Chen, J., and He, L. (2009). Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. PLoS Comput. Biol. *5*, e1000441.

Zhang, Q.C., Petrey, D., Norel, R., and Honig, B.H. (2010a). Protein interface conservation across structure space. Proc. Natl. Acad. Sci. USA *107*, 10896–10901.

Zhang, T., Zhang, H., Chen, K., Ruan, J., Shen, S., and Kurgan, L. (2010b). Analysis and prediction of RNA-binding residues using sequence, evolutionary conservation, and predicted secondary structure and solvent accessibility. Curr. Protein Pept. Sci. *11*, 609–628.

Zhang, T., Zhang, H., Chen, K., Shen, S., Ruan, J., and Kurgan, L. (2008). Accurate sequence-based prediction of catalytic residues. Bioinformatics *24*, 2329–2338.

Zhu, H., Sommer, I., Lengauer, T., and Domingues, F.S. (2008). Alignment of non-covalent interactions at protein-protein interfaces. PLoS ONE *3*, e1926.

 ©2012 Elsevier Ltd All rights reserved