Are predicted protein structures of any value for binding site prediction and virtual ligand screening?
Jeffrey Skolnick, Hongyi Zhou and Mu Gao

The recently developed field of ligand homology modeling (LHM) that extends the ideas of protein homology modeling to the prediction of ligand binding sites and for use in virtual ligand screening has emerged as a powerful new approach. Unlike traditional docking methodologies, LHM can be applied to low-to-moderate resolution predicted as well as experimental structures with little if any diminution in performance; thereby enabling ~75% of an average proteome to have potentially significant virtual screening predictions. In large scale benchmarking, LHM is able to predict off-target ligand binding. Thus, despite the widespread belief to the contrary, low-to-moderate resolution predicted structures have considerable utility for biochemical function prediction.

Address
Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, 250 14th Street NW, Atlanta, GA 30318, USA

Corresponding author: Skolnick, Jeffrey (skolnick@gatech.edu)

Current Opinion in Structural Biology 2013, 23:xx–yy
This review comes from a themed issue on Theory and Simulation
Edited by Jeffrey Skolnick and Richard Friesner

0959–440X/$ – see front matter, © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.sbi.2013.01.009

Introduction
Over the past decade, the field of protein structure prediction has matured to the point where a significant fraction of the proteins in a given proteome can be modeled at low-to-moderate resolution [1]. On the other hand, the biochemical function of many proteins in a proteome, most especially those associated with ligand binding and other intermolecular interactions, are only partially known [2]. For example, the metabolic enzymes of well-studied organisms such as yeast are not fully characterized [3,4]. Thus, a key question facing the field is can predicted protein structures be successfully employed for the prediction of protein function? Of course, function is multifaceted, but clearly the inference of biochemical function would be the most direct application of structural information. In this review, we focus on the utility of predicted protein structures in the identification of ligand binding sites, and having identified these sites, their usefulness in virtual ligand screening to assist in drug discovery. But, before embarking on a discussion of the utility of lower resolution structures, a brief summary of the status of the field when high-resolution structures are used is appropriate as it provides the standard by which newly developed approaches must be assessed.

Binding site detection in high-resolution structures
Having a three-dimensional structure in hand, one would like to identify its small molecule binding sites. Some approaches locate binding sites by a geometric match to three-dimensional descriptors or templates of biologically relevant sites [5,6]. More powerful is the evolutionary trace methodology that combines protein structure with conserved residue patterns mapped onto the protein’s surface [7,8,9]. There are also geometric methods that locate binding residues by searching for cavities/pockets in a protein’s structure [10,11]. Among the best pocket detection algorithms is LIGSITE [12] that calculates surface-accessibility on the protein’s Connolly surface [13] and then reranks the pockets by the degree of conservation of select surface residues. Other methods calculate titration curves [14] or identify electrostatically destabilized residues [15]. These methods strictly focus on the protein’s sequence and structural features and ignore the identity of the ligand, but they are a necessary first step.

Virtual ligand screening using high-resolution structures
Having identified a binding site in a structure, the next step is to identify its binding ligands. Most traditional approaches are docking-based and prioritize compounds by predicting their binding mode [16] and then binding affinity [17]. Here, high-resolution structures of the target protein receptor, preferably in its ligand-bound conformational state, are generally required [18]. There are many successful self-docking studies where the ligand is excised from its crystal structure and then redocked [19]. However, many proteins exhibit significant motion upon ligand binding [20,21], and even small motions diminish docking accuracy. For example, for trypsin, HIV-1 protease and thrombin, ~90% of initial docking accuracy is lost when the mean protein structural rearrangement exceeds 1.5 Å [22]. These results raise the following questions: are ligand binding sites really so structurally unique in nature and if not, why are high-resolution structures needed for ligand docking?
Does the need for high-resolution structures in binding site prediction and virtual screening reflect physical principles or is it just a technical limitation?

There is the widespread belief that predicted structures whose backbone RMSD ranges from 2 to 6 Å are useless for either ligand binding site prediction or for virtual ligand screening [22]. For example, the performance of the LIGSITECSC [12*] pocket detection algorithm deteriorates dramatically as one goes from crystal structures to predicted models in large-scale benchmark tests [20**]. However, local structural distortions are routine in nature [23]. For example, the binding sites of distantly related native proteins that bind very similar, if not identical ligands, with similar residues have an average pairwise backbone RMSD of 2.15 ± 0.77 Å [24]. As a specific example, for the subset of the kinome having holo crystal structures, the structural variation of the ‘conserved’ ATP-binding site is ~2.4 Å [25*]. Thus, there is significant structural plasticity of ligand binding sites [23,26]; it is unlikely that there is a unique ligand–protein conformation, with other nearby conformations having an entirely unfavorable binding free energy. The observed ensemble of native ligand binding conformations also suggests that low-resolution models might be useful for binding site identification/virtual screening provided that they capture the majority of the structural features and essential interactions.

Why then do extant docking methods [16,27–31] require high-resolution structures? One underlying cause is the fact that they are driven by steric and van der Waals interactions [32]. A slight conformational inaccuracy could cause a dramatic interaction change. If a ligand fits into the binding site, then ligand ranking is dominated by the molecular weight of the ligand, independent of whether the cognate ligand or a randomized version is used [32]. Thus, there is the need for a more accurate atomic force to be developed. However, if the resulting force field is too complex, it would have limited practical utility as it must be able to screen millions of compounds across the thousands of proteins in the human or other proteomes [33].

Ligand homology modeling: binding site detection and virtual ligand screening

To employ protein models requires approaches that can accommodate binding site structural variations without a significant diminution in accuracy. As a first approximation, one might imagine that global structural similarity between proteins would be sufficient to infer protein function [34], most especially, common binding sites. In a recent study [35], for structurally related proteins whose pairwise sequence identity is in the twilight zone, we concluded that even at quite high levels of structural similarity, less than 25% of the targets share a common binding pocket. Thus, structural similarity alone is insufficient to transfer binding site location. A class of methods that exhibits the desired insensitivity to receptor structure deformation and which allows one to infer binding site location and type of ligands bound is ligand homology modeling (LHM) [36,37,38,24,39,40*,41,42*]. LHM exploits the fact that the ideas of homology modeling, as applied to protein structure prediction [43], are applicable to functional inference, ligand binding pose prediction and virtual ligand screening. As shown schematically in Figure 1, LHM consists of six steps:

1. Functional relationships between evolutionarily distant proteins are detected by sequence profile-driven threading to identify common ligand binding pockets, functionally important residues and structural conservation (anchors) of their ligand binding modes [37*].
2. These conserved features are used to construct a ligand fingerprint profile from the identified template ligands [44].
3. Initial virtual screening of ligands is then done via fingerprint scanning.
4. The small molecule ligands are placed in the protein’s predicted binding site using the conserved ligand anchor regions identified in (1) [37*]. Interestingly, the pose of the anchor in the ligand binding site tends to be strongly conserved, as are the residues contacting the ligand. Furthermore, the B-factors of the residues touching the ligand’s anchor are lower than those outside the anchor region.
5. The ligand’s pose is readjusted to optimize its interactions with the protein’s structure [39]. We further found that the positions of the side chain functional groups in contact with the ligand anchor functional groups tend to be strongly conserved and act together as a structural unit [45]. Indeed, they can refine the backbone geometry. This is in contrast to traditional ligand docking where the protein’s structure is held fixed and the ligand conformation is adjusted to accommodate the protein’s structure [32,46*].
6. Using the refined conformations, the ligand library is then reranked via a machine learning procedure [37*,41].

One of the advantages of LHM is that binding site detection is quite insensitive to structural quality. For example, consider the results when FINDSITE [20**] was applied to a representative benchmark set, none of whose templates are closer than 35% identical. We consider the prediction of a binding site to be successful when the centers of mass of the predicted and observed binding sites are <4 Å. Using crystal structures, for the best of top five predicted ligand-binding sites, the success rate for FINDSITE is 70.9% vs. 51.3% for LIGSITECSC. For TASSER [47] predicted models, FINDSITE has a
67.3% success rate, whereas LIGSITE\textsuperscript{CSC}’s success rate is 32.5%. Similar results have been reported for binding site detection by other LHM variants \cite{40,42,48}. LHM has also been applied to predict metal binding sites \cite{1,49}. For example, FINDSITE-metal identifies the metal binding site in TASSER models in 59.4% of the cases. Moreover, when the metal is iron, copper, zinc, calcium, and magnesium ions, the identity of the binding site metal can be predicted with 70–90% accuracy.

**What happens when holo templates are unavailable for the target of interest?**

While contemporary structure prediction approaches provide sufficiently accurate models for about 76% of the proteins in the human proteome \textless 1000 residues in length \cite{1}, because of the relative scarcity of solved holo template structures in the PDB \cite{50,51}, one can only infer ligand binding information for \textasciitilde 26% of the human proteome \cite{52}. Thus, methods that do not require holo template structures must be developed. To address this, FINDSITE\textsuperscript{X} \cite{52}, an extension of FINDSITE \cite{20**,} was developed that uses predicted structures for template proteins having experimental ligand binding information but which lack solved structures. Thus, pseudo holo templates are generated. To provide predicted protein structures, a fast and accurate version of TASSER\textsuperscript{VMT} \cite{53}, TASSER\textsuperscript{VMT-lite}, for template-based structural modeling was developed and tested, with comparable performance as the best CASP9 servers \cite{54}. Then, a hybrid approach that combines structure alignments with an evolutionary similarity score for identifying functional relationships between target and template proteins with binding data was formulated.

FINDSITE\textsuperscript{X} was applied to all identified human G-protein coupled receptors (GPCRs). First, TASSER\textsuperscript{VMT-lite} improved models of all previously modeled human GPCR structures \cite{55}. We then used these structures to screen against the ZINC8 \cite{56} nonredundant (Tanimoto coefficient \cite{57}, TC < 0.7) ligand set of 88,949 compounds combined with ligands from the GLIDA database \cite{58}. Testing FINDSITE\textsuperscript{X} (excluding GPCRs from the binding data library whose sequence identity > 30% to the target protein) on a 168 protein human GPCR set with known binders, the average enrichment factor in the top 1% of the compound library (EF\textsubscript{0.01}) is 22.7, with encouraging results for off-target interaction predictions. All 998 predicted human GPCR structures, virtual screening results and predicted off-target interactions are available at \cite{59}.
Combined LHM approaches to proteome scale virtual ligand screening

To combine the advantages of information provided by distant holo templates when they are available with experimental data and using pseudo holo templates when they are not, FINDSITE\(_{\text{comb}}\) was developed [60\textsuperscript{*}]. A significant component of FINDSITE\(_{\text{comb}}\), is an improved version of FINDSITE, FINDSITE\(_{\text{filt}}\) that filters out false positive ligands in threading identified templates by a better binding site detection procedure that includes binding site amino acid similarity. For virtual ligand screening, FINDSITE\(_{\text{comb}}\) combines FINDSITE\(_{\text{filt}}\) with FINDSITE\(_{\text{X}}\) that uses the ChEMBL [2] and DrugBank [61] ligand binding databases. The rank of each screened ligand is the best of its three ranks to ligands using fingerprints derived from the PDB, ChEMBL, and DrugBank libraries. In what follows, we summarize the results of FINDSITE\(_{\text{comb}}\) in benchmarking mode, where all template proteins with >30\% sequence identity to a target are excluded. We note that in large scale testing FINDSITE\(_{\text{comb}}\) produces significant virtual screening predictions for about 75\% of an average proteome [33].

Comparison of LHM to traditional docking approaches

The DUD set is designed to help test docking algorithms by providing challenging decoys [62\textsuperscript{*}]. For each active, there are 36 decoys with similar physical properties (e.g. molecular weight, calculated log \(P\)) but dissimilar chemical topology. Table 1 compares the relative performance of FINDSITE\(_{\text{comb}}\) with traditional docking methods, including AUTODOCK Vina [46\textsuperscript{*}] and DOCK 6 [31], in cross-docking (a realistic scenario), where all 97,974 nonredundant DUD ligands are screened against all targets, as well as in non-cross-docking, where screening is just done against the experimentally determined active and inactive molecules. Results are presented for crystal structures and TASSER\(_{\text{VMT-lite}}\) modeled structures. For both cases, each FINDSITE component performs better than AUTODOCK Vina or DOCK 6. While the performance of traditional methods deteriorates when models are used, FINDSITE-based approaches do not. Finally, in [63], several docking programs were compared for virtual screening accuracy in non-cross-docking on experimental structures on DUD. FINDSITE\(_{\text{comb}}\), whose mean average area under the ROC curve, the Area Under the Accumulation Curve, AUAC = 0.77, performs as well as the best performing Glide (v4.5) [28] (mean AUAC = 0.72). FINDSITE\(_{\text{comb}}\) performs better than all other compared methods: DOCK 6 (mean AUAC = 0.55), FlexX [30] (mean AUAC = 0.61), ICM [27] (mean AUAC = 0.63), PhDOCK (mean AUAC = 0.59) [29,64,65] and Surflex [63] (mean AUAC = 0.66). Table 2 shows the AUAC values using both experimental and modeled structures for FINDSITE\(_{\text{comb}}\) with AUAC = 0.77 and 0.75, respectively, as well as its constituent components for both experiment and modeled structures. As in Table 1, the dominant contribution to the success of FINDSITE\(_{\text{comb}}\) is due to FINDSITE\(_{\text{filt}}\) whose AUAC = 0.74 for experimental and modeled structures is the same.

In addition to being broadly applicable, FINDSITE\(_{\text{comb}}\) is considerably faster than traditional docking methods. On a single state of the art CPU, for a 325 residue protein screened against 100,000 compounds, FINDSITE\(_{\text{comb}}\) is ~30 times faster than AUTODOCK Vina [46\textsuperscript{*}] and ~160 times faster than DOCK 6 [31]. Thus, FINDSITE\(_{\text{comb}}\) can be applied to screen millions of compounds on a proteomic scale. Despite the fact that predicted models rather than high-resolution crystal structures are used, LHM methods are very strongly competitive with traditional docking approaches.

<table>
<thead>
<tr>
<th>Method</th>
<th>Cross-docking</th>
<th>Non-cross-docking</th>
<th>DOCK 6</th>
<th>AUTODOCK Vina</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINDSITE(_{\text{X}})</td>
<td>16.89</td>
<td>20.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINDSITE(_{\text{Filt}})</td>
<td>22.32</td>
<td>21.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINDSITE(_{\text{Comb}})</td>
<td><strong>27.69</strong></td>
<td><strong>23.10</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUTODOCK Vina</td>
<td>8.92</td>
<td>2.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOCK 6</td>
<td>3.14</td>
<td>3.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\text{EF}_{0.01}\) is the enrichment factor relative to random for the top 1\% of ranked molecules.

Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>DUD non cross docking</th>
<th>DrugBank targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental structures</td>
<td>Modeled structures</td>
</tr>
<tr>
<td>FINDSITE(_{\text{Comb}})</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>FINDSITE(_{\text{Filt}})</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>FINDSITE(_{\text{X}})</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>FINDSITE</td>
<td>–</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Large-scale benchmarking tests on drug target proteins and the prediction of off-target interactions

FINDSITE\textsuperscript{comb} was tested in benchmarking mode on all 3576 DrugBank [66] targets <1000 residues in length. Target and template structures are modeled with TASSER\textsuperscript{MTH}-lite [52]. The screened compound library consists of all 6507 drugs (the true binders of all targets) plus 67,871 ZINC8 nonredundant (culled to TC < 0.7) compounds [56] as background. The results of FINDSITE\textsuperscript{comb} along with its component methods and the original FINDSITE [20\textsuperscript{**}] are compiled in Table 3. FINDSITE\textsuperscript{comb} is better than any of its component methods. Table 3 also shows that FINDSITE\textsuperscript{filt} is better than FINDSITE [20\textsuperscript{**}] by a significant ~45\% for EF\textsubscript{0.01} (46.0 vs. 31.7), as well as in its coverage of targets with EF\textsubscript{0.01} > 1 (58\% vs. 43\%). FINDSITE\textsuperscript{comb} has an average EF\textsubscript{0.01} of 52.1 and is better than random (EF\textsubscript{0.01} > 1) for 65\% of the targets. Finally, Table 2, column 4 shows the AUAC results for FINDSITE FINDSITE\textsuperscript{comb} where AUAC = 0.87 and its constituent components. As in the DVD benchmark, the performance of FINDSITE\textsuperscript{comb} is dominated by FINDSITE\textsuperscript{filt}.

Another application of the LHM approach was in the structural and functional characterization of the entire human kinome [25\*]. Encouraging virtual screening results were presented for ligands predicted to bind to the conserved ATP-binding pocket [56]. In a more rigorous test, crossreactivity virtual profiling of the human kinome was done. For almost 70\% of the inhibitors, their alternate molecular targets can be effectively identified in the human kinome with a high (>0.5) sensitivity, yet relatively low false positive rate (<0.5) [67].

Conclusions

Just as the field of protein structure prediction has greatly benefited by the development of template-based approaches [54,68], we argue that LHM [1,20\textsuperscript{**},37,58,40,41,42] has matured to the point where LHM is a powerful method for the prediction of ligand binding sites and virtual ligand screening. It offers the advantages that predicted as well as high-resolution structures can be successfully used, with minor diminution in performance. While certainly not perfect, in virtual screening LHM results are often considerably better than random and could be used to guide experimental screening approaches.

As noted by Bourne and coworkers [69,70]\* and is evident from an analysis of DrugBank [61] targets, the binding of ligand to a protein target other than the one for which the drug was designed is quite common [66,71]. Moreover, in PDB structures, very similar binding sites are found in globally unrelated proteins [72\*]. The challenge will be to extend these observations to predicted low-to-moderate resolution protein structures and then to apply them on a proteomic scale. If so, LHM could be a powerful tool to help repurpose FDA approved drugs and could help with the elucidation of metabolic pathways [73]. These and other related applications will undoubtedly be pursued in the near future.

Table 3

| Performance of different FINDSITE based methods for 3576 DrugBank targets |
|---------------------------------|-----------------|-----------------|
| Method                          | Average EF\textsubscript{0.01} | # (% of targets having EF\textsubscript{0.01} > 1) |
| FINDSITE (PDB)                  | 31.7            | 1526 (43\%)     |
| FINDSITE\textsuperscript{X}     | 36.6            | 1714 (48\%)     |
| FINDSITE\textsuperscript{filt} (PDB) | 46.0            | 2080 (58\%)     |
| FINDSITE\textsuperscript{comb}  | 52.1            | 2333 (65\%)     |

Acknowledgement

This work was supported in part by grant Nos. GM-48835, GM-37048 and GM-084222 of the Division of General Medical Sciences of the National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


6 Theory and Simulation


43. An LHM approach to binding site identification that improves upon FINDSITE by better describing the ligand binding site and that performed well in CASP8.


48. A widely used traditional approach to molecular docking and virtual ligand screening.


Are predicted protein structures of any value for binding site prediction and virtual ligand screening? Skolnick, Zhou and Gao


A combined, next generation, LHM approach that uses both holoto templates and pseudo holoto templates for virtual ligand screening that gives quite good performance on the DUD benchmark as well as for DrugBank drug targets.


A much needed and widely used benchmark set to assess the utility of ligand docking algorithms.


Argues that the off-targeting of drugs is the norm rather than the exception and describes proteome scale approaches to drug repurposing.


A convincing and important paper which demonstrates that the notion of the ‘one-gene—one drug—one disease’ paradigm of traditional drug discovery significantly underestimates the likelihood that a single drug interacts with multiple targets.