JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Article

Prediction of substrates for glutathione transferases by covalent docking

GUANG QIANG DONG, Sara Calhoun, Hao Fan, Chakrapani Kalyanaraman, Megan C. Branch, Susan Mashiyama, Nir London, Matthew P Jacobson, Patricia Clement Babbitt, Brian K. Shoichet, Richard N. Armstrong, and Andrej Sali

J. Chem. Inf. Model., **Just Accepted Manuscript** • Publication Date (Web): 06 May 2014

Journal of Chemical Information and Modeling is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Prediction of substrates for glutathione transferases by covalent docking

Guang Qiang Dong[1], Sara Calhoun[1], Hao Fan[4], Chakrapani Kalyanaraman[3], Megan C. Branch[2], Susan T. Mashiyama[1], Nir London[3], Matthew P. Jacobson[3], Patricia C. Babbitt[1], Brian K. Shoichet[5], Richard N. Armstrong[2*], Andrej Sali[1*]

[1]*Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, California 94158, USA*

[2]*Departments of Biochemistry and Chemistry, Center in Molecular Toxicology, and Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37232-0146, USA*

[3]*Department Pharmaceutical Chemistry, California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, San Francisco, California 94158, USA*

[4]*Agency for Science, Technology and Research (A*STAR), Bioinformatics Institute 30 Biopolis Street, Matrix #07-01, Singapore, SG 1386715*

[5]*University of Toronto, Faculty of Pharmacy, 160 College St, Toronto, ON, CAN M5S 3E1*

[*]Corresponding authors:

Andrej Sali, UCSF MC 2552, Byers Hall at Mission Bay, Suite 503B, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94158, USA. Tel: +1 415 514 4227; fax: +1 415 514 4231. E-mail: *sali@salilab.org*

Richard Armstrong, Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. Tel: +1 615 343 2920. E-mail: *r.armstrong@vanderbilt.edu*.

Running title: Prediction of glutathione transferase substrates

Total number of pages, figures, tables: 31,10,4

TO BE Submitted to: *Journal of Chemical Information and Modeling*

Date: March 11, 2014

Keywords: GST; superfamily; covalent docking; virtual screening; comparative modeling.

ABSTRACT

Enzymes in the glutathione transferase (GST) superfamily catalyze the conjugation of glutathione (GSH) to electrophilic substrates. As a consequence they are involved in a number of key biological processes, including protection of cells against chemical damage, steroid and prostaglandin biosynthesis, tyrosine catabolism, and cell apoptosis. Although virtual screening has been used widely to discover substrates by docking potential non-covalent ligands into active site clefts of enzymes, docking has been rarely constrained by a covalent bond between the enzyme and ligand. In this study, we investigate the accuracy of docking poses and substrate discovery in the GST superfamily, by docking 6738 potential ligands from the KEGG and MetaCyc compound libraries into 14 representative GST enzymes with known structures and substrates using the PLOP program[1]. For X-ray structures as receptors, one of the top 3 ranked models is within 3Å all-atom RMSD of the native complex in 11 of the 14 cases; the enrichment LogAUC value is better than random in all cases, and better than 25 in 7 of 11 cases. For comparative models as receptors, near native ligand-enzyme configurations are often sampled, but difficult to rank highly. For models based on templates with the highest sequence identity, the enrichment LogAUC is better than 25 in 5 of 11 cases, not significantly different from the crystal structures. In conclusion, we show that covalent docking can be a useful tool for substrate discovery and point out specific challenges for future method improvement.

Availability: The scripts, comparative models, benchmark and the docking library are available at http://salilab.org/GST.

INTRODUCTION

The canonical glutathione transferases (also known as GSTs; EC 2.5.1.18) catalyze addition of an excellent nucleophile to an electrophilic center. They play important roles in the metabolism and detoxification of numerous endogenous and xenobiotic compounds, including oxidized lipids, drugs, and pollutants[2-5]. The canonical GSTs are a subset of the thioredoxin fold family of proteins[6, 7]. They consist of an N-terminal-thioredoxin domain and a C-terminal α-helical domain. Although a number of GSTs are known to have specific substrates, many if not most have no clearly assigned biological substrates. In addition, the general nature of their chemistry leads to enzymes that tend to be catalytically promiscuous even with respect to the transition state for the reaction[3]. As a consequence, the *de novo* prediction of enzyme function by computational methods becomes challenging.

Computational docking methods have been widely used in ligand discovery for many enzymes[8-12]. In particular, these methods can predict the docking pose of a known ligand (docking) and/or predict ligands in a large library of small molecules (virtual screening). Docking consists of searching through plausible binding modes of a compound and scoring each mode to distinguish a near-native binding pose from others (docking). Virtual screening consists of performing docking for each candidate ligand, followed by the ranking of the candidate ligand by their best docking scores.

Although this structure-based approach has been used successfully for the prediction of both substrates and other types of ligands (e.g., orthosteric inhibitors and allosteric modulators), docking for substrate discovery is most difficult, particularly when the enzyme may accommodate more than one type of reaction. In addition, only ground state or intermediate state complexes, not transition states, are typically accessible by standard docking procedures. The resulting complexes may or may not be directly competent for turnover in the absence of information on the "pre-organization" of the enzyme-substrate complex required for catalysis. Despite the difficulties involved in the docking of substrates, structure-based methods have been used successfully for substrate discoveries in several systems [13-18].

There are two principal challenges in the *de novo* prediction of the substrate preferences of enzymes in large, functionally diverse superfamilies. The first challenge is the availability of experimentally determined enzyme structures, which lags behind the number of known protein sequences by a factor of approximately 400 as of June 2013. To remedy this situation, virtual screening has also relied on comparative models, not only experimentally determined structures[19-26]. The relative utility of comparative models versus experimentally determined structures has been assessed[19, 27]. The second challenge is that standard docking procedures lack sufficient constraints that efficiently define the productive geometries between the reacting species (substrates) on the enzyme surface. This issue is addressed here by applying a covalent bond constraint between the sulfur of GSH and the electrophilic substrate. Importantly, the effectiveness of covalent docking has not been rigorously addressed yet.

The absence of a large-scale benchmarking of covalent docking using either X-ray structures or comparative models, and the need to predict substrates for GST enzymes, inspired us to investigate the following questions. Can covalent docking accurately predict docking poses of known ligands, given experimentally determined, homologous or modeled structures in the GST superfamily? Can covalent docking accurately predict ligands despite the catalytically promiscuity of many GST enzymes? What is the difference in the utility of apo, holo, comparative modeling, and homologous structures for virtual screening in the GST superfamily? Can the virtual screening be improved by consensus scoring, relying on independent screening against multiple holo, apo, comparative modeling, and homologous structures in the GST superfamily? If multiple models are calculated on the basis of different templates, can any of them outperform apo and even holo X-ray structures of the target? If so, can one reliably identify which model will do so, or even perform optimally among a set of modeled structures; are there sequence and/or structural attributes (i.e., the overall target-template sequence identity, the binding site target-template sequence identity, and the predicted accuracy of a model) that reliably predict the accuracy of ligand docking?

In this report, we attempt to answer these questions with the aid of a virtual library of compounds and 14 representative GST enzymes of known structure and function. The virtual library consists of known substrates for the selected GST proteins, and a large number of compounds selected from KEGG[28] and MetaCyc[29]. For each target, the entire virtual library of compounds was docked to the known X-ray structures, homologous structures, and comparative models of the protein. The results are analyzed by comparing the docking poses of native products to X-ray structures, and calculating the enrichment of known products with respect to the entire compound library.

We begin by describing the GST catalyzed reactions, the docking library, the selected GST proteins, the automated modeling pipeline, the docking pipeline, and methods to evaluate the accuracy of predicted docking poses for known ligands and the accuracy of virtual ligand screening (Methods). We then describe and compare the results of docking native ligands and virtual screening using apo, holo, comparative modeling, and homologous structures (Results). Finally, we discuss the implications of the current approach and answer the questions we asked above, given our modeling, docking, and benchmark (Discussions).

METHODS

We begin by listing the set of the GST catalyzed reactions used in this study, followed by a description of the corresponding products that comprise a virtual screening library. Next, we describe how we selected GST targets for docking and how we built their comparative models. Finally, we describe the covalent docking pipeline as well as the assessment criteria for evaluating the accuracy of docking and virtual screening.

Construction of the virtual screening library

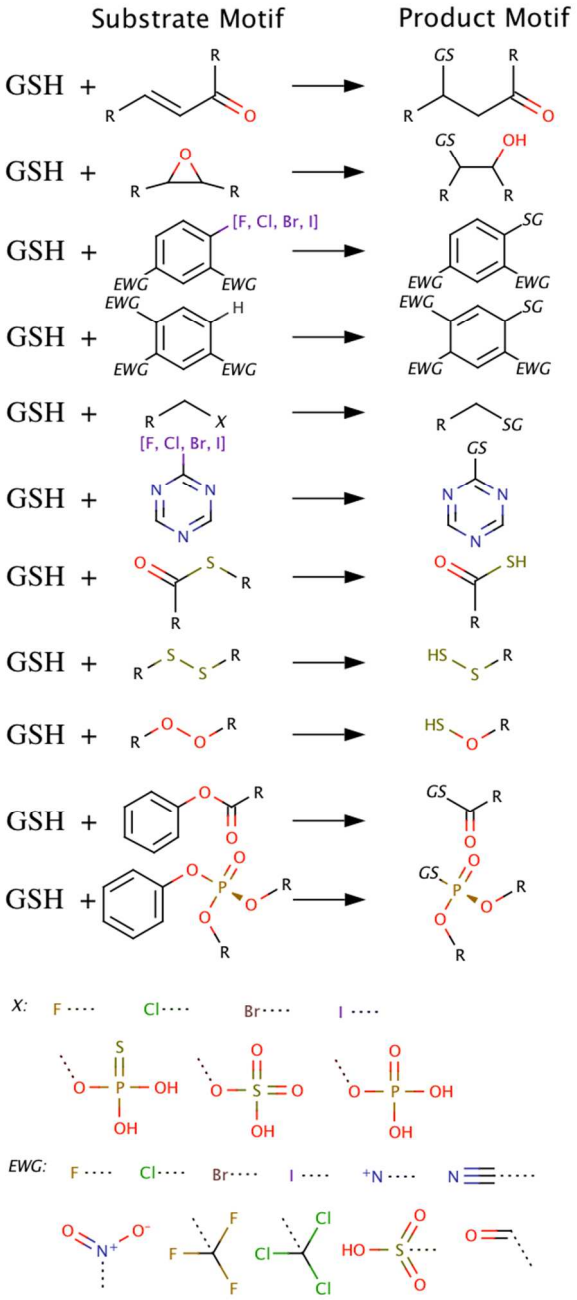


Figure 1. A partial list of chemical reactions catalyzed by GSTs. Each row defines a type of reaction based on the reactive functional group. *X* indicates a leaving group in the nucleophilic substitution reactions. *EWG* indicates an electron-withdrawing group.

GSTs catalyze a range of different reactions. Here, we considered 11 different kinds of GST catalyzed reactions, each with a different substrate motif (Figure 1). For each substrate motif, we found all matching molecules in the KEGG and MetaCyc compound databases. The search was performed with the OECHEM tools[30] by matching a SMILES[31] string of a

database molecule to the SMARTS[31] strings representing the substrate motifs. We found 4,149 and 3,259 in KEGG and MetaCyc, respectively. We also added 64 substrates from the literature[32-53], for the total of 6,738 unique substrates.

Next, each of the substrates was converted to one or more products, as follows. The conversions corresponding to the 11 reactions were carried out using OECHEM's library generation function[30] with explicit hydrogens. The reactive functional groups in the substrate were identified by comparing the substrate SMILES string with the SMARTS string representing the substrate motif undergoing conversion to a product motif, for each reaction. Each match was then used to convert the substrate to a product that was added to the virtual screening library.

Finally, we prepared products for docking. For products with undefined stereocenters, we first enumerated the stereoisomers using OECHEM[30], with the maximum number of stereoisomers retained arbitrarily set to 16 for computational efficiency. Second, protonation states of each stereoisomer was enumerated within pH range 6 to 8, followed by generating a conformation for each protonation state, using Epik[54, 55] and LigPrep[56]. Finally, force field parameters for each product were generated using the hetgrp_ffgen utility (Schrödinger, LCC).

Selection of the GST targets for docking

A subset of GST structures were selected from the Protein Data Bank (PDB) for retrospective docking by maximizing the sequence and functional coverage of the GST superfamily, following three steps: First, all GST structures with a co-crystallized GSH-substrate conjugate were extracted from the PDB. Second, structures for the same protein were grouped together, resulting in fourteen groups with unique sets of ligands. Finally, for each group, the structure with the highest resolution, the lowest $R_{free,}$ and a unique ligand was selected, resulting in 14 target holo X-ray structures (Supporting Table S1).

Because GSTs function as dimers, we used for docking the biological dimer unit from the PDB. The substrate part of the GSH-substrate conjugate in holo-structures was removed, resulting in GST holo X-ray structures with only GSH bound. Each dimer was then prepared for docking using the Protein Preparation Wizard (Schrödinger, LCC) by designating each of the two GSHs as the reactive cofactor, resulting in two receptor binding sites (with the exception of 1GWC, which has only one site occupied by GSH); in most cases, the two binding sites are slightly different because symmetry is generally not imposed during crystallographic structure refinement.

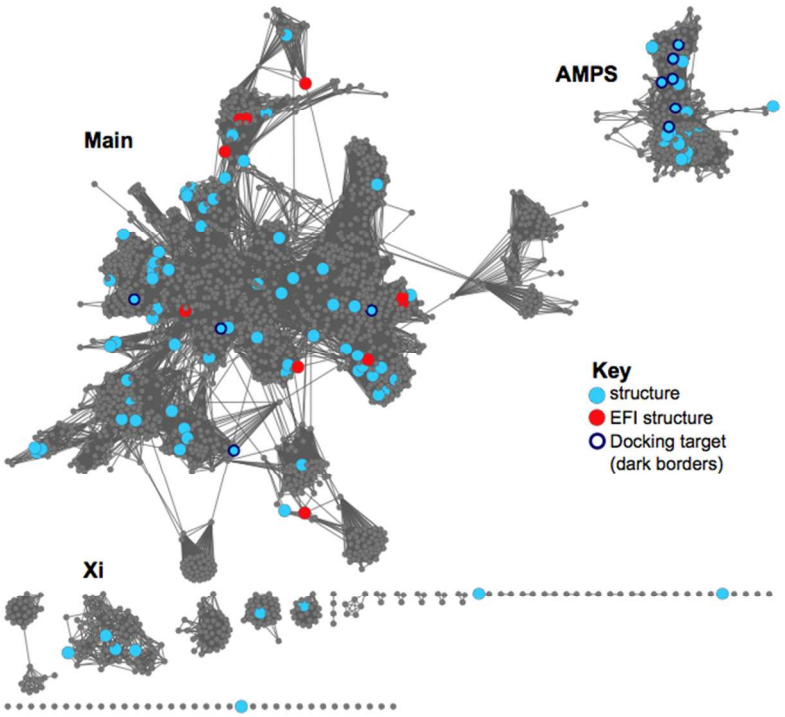


Figure 2. Representative network view of the GST superfamily showing where docking targets fall in this superfamily. In this view, 2,190 50% sequence identity filtered nodes ("ID50 node") represent 13,493 sequences are shown as dots, and each sequence similarity relationship with a BLAST *E*-value ≤ 1x10-13 is shown as a line or edge between representative nodes. The largest clusters are labeled by their subgroup in the SFLD[38]. The 13,493 sequences for cytosolic GSTs ("cytGSTs") is based on Pfam (v26.0)[35] sequences that were at least 100 residues in length and had scores above the gathering threshold for at least one Pfam hidden Markov model ("HMM") corresponding to the cytGST N-terminal domain ("GST_N," "GST_N2", or "GST_N3"; collectively referred to here as "GST_N*"). Also included in the dataset were 58 proteins that lacked matches to a GST_N* HMM, but were chosen as EFI[57] targets for experimental characterization because of their sequence similarities to cytGSTs. Sequence similarity was detected with an all-by-all BLAST search using blastp (v2.2.24)[36]. The data were viewed with Cytoscape (v2.8.3). Sequence identities are calculated using CD-HIT[37] ("ID50" set). The edges are shown using the organic layout where nodes that are more highly interconnected are clustered more tightly together. PDB structures associated with cytGST sequences (95% or more sequence identity to the PDB sequence) were identified using the Structure-Function Linkage Database ("SFLD")[38] interface. An ID50 node is shown as a blue dot if any member sequence of the node was associated with a PDB structure. Representative nodes where only an EFI structure was associated with its member sequences are shown as red dots. Retrospective docking targets are marked using dark blue borders; only 10 targets are shown because the rest are more than 50% identical to these 10 representatives shown.

Comparative modeling of the GST targets

Comparative models for each target were built by selecting template dimer structures from the PDB with either an unmodified GSH (apo) or a conjugate form (product) of GSH (holo). For each target, we aimed to include one holo and one apo template structure for each of the four sequence identity ranges: 25% to 40%, 40% to 55%, 55% to 70%, and 70% to 85%. The target-template sequence alignments were generated using HHalign[58], which in turn relied on target and template profile Hidden Markov Models (HMMs) generated using HHblits[58] with the UniProt20 database[59]. Five-hundred models were built for each target-template alignment, using the automodel protocol in MODELLER-9v10[60]. The coordinates for GSH atoms in the template structure were then copied to the target models, using the BLK function of MODELLER. Models were then assessed by the discrete optimized protein energy (DOPE)[61] function of MODELLER. Finally, the model with the lowest normalized DOPE score was used for docking.

To compare different comparative models and compare the models against X-ray structures, we calculated the binding site sequence identity of the templates to the X-ray structures, and the binding site non-hydrogen atom Root-Mean-Square Deviation (RMSD) between a comparative model and X-ray structure. The binding site was defined to contain atoms in all residues with at least one atom within 6 Å of any ligand atom in the X-ray structure. As expected, strong correlations were observed between the overall target-template sequence identity and the target-template binding site sequence identity, and between the target-template binding site sequence identity and model's non-hydrogen atom RMSD error (Figure 5A-B).

As designed by the choice of the template structures, the distribution of the non-hydrogen atom RMSD error of the binding site confirms that there is a range of accuracy across the set of modeled structures. The models with sub-Angstrom non-hydrogen atom RMSD error tend to be those based on templates with over 80% sequence identity in the binding site residues to the target sequence. Over a third of models had a binding site RMSD error over 4 Å. Thus, homology models cover a range of sequence identities and a range of accuracy.

Protocol for covalent docking

Covalent docking of a potential product to a receptor was started by placing it in the binding site. More specifically, the coordinates of the GSH part of the potential product were matched to the GSH molecule in the receptor, and the coordinates of the remaining atoms of the product were then built using OMEGA[62]. Up to 20 initial configurations were generated for each product molecule.

For each initial configuration, we then used PLOP's tether pred (Academic version 25.6) function[1] to rotate the rotatable bonds in the GSH-substrate conjugate. The rotatable bonds were identified using OECHEM. We sampled all the rotatable bonds in the substrate part of the GSH-substrate conjugate as well as the CA-CB and CB-SG bonds in the cysteine residue of GSH. Up to 50 configurations were generated by PLOP starting from each initial

configuration. These configurations were then scored by calculating the total potential energy of the product-receptor complex by PLOP (in the units of kcal/mol), which in turn relies on the OPLS force-field with a variable-dielectric generalized Born model[63]. The product-receptor distances alone were also scored by the atomic statistical potential PoseScore (in arbitrary units)[64]. While the PLOP potential energy does not contain the entropic contributions to the binding free energy, the PoseScore term does approximate the contribution of the interface to the binding free energy[65].

Finally, for virtual screening, a substrate was ranked using the median PLOP energy of its products' different stereoisomers and protonation states. The PLOP energy of a product in a specific stereoisomer form and protonation state was the median PLOP energy of its different configurations.

Assessment of docking and virtual screening

When a native product was docked (Supporting Table S2), the docking pose of the product was assessed for accuracy based on its non-hydrogen atom RMSD from the native configuration, after superposition of the receptor used for docking on the native structure of the receptor; GSH was excluded, except for the cysteine sulfur atom.

The accuracy of virtual screening was evaluated by the enrichment for the known products (Supporting Table S3) among the top scoring potential products. The enrichment curve was obtained by plotting the percentage of actual products found (y-axis) within the top ranked subset of all database compounds (x-axis on logarithmic scale). LogAUC, the area under the curve of the enrichment plot, was also calculated to indicate the accuracy of enrichment; random selection has a logAUC of 14.5.

RESULTS

We begin by evaluating the docking pose of the native ligands using holo and apo structures, followed by evaluating the docking pose of the native ligands using comparative models and homologous structures. Next, enrichment of the known ligands using X-ray structures is benchmarked and analyzed. Finally, we analyze the enrichment of the known ligands using comparative models.

Docking of products into X-ray structures

As the easiest test, we first applied covalent docking by PLOP to reconstruct binding poses of the crystallographic products in the corresponding holo X-ray structures (Supporting Table S1). In all 14 cases, a docking pose within 3Å of the native structure was sampled (within 2Å for 13 cases), though not necessarily recognized as such by the scoring function (Supporting Table S4); the 3Å threshold on the all-atom RMSD subjectively distinguishes near-native and non-native poses. However, the top 3 ranked structure was within 3Å to the native structure in 9 out of 14 cases using the PLOP energy function, and in 11 out of 14 cases using PoseScore (Figure 3).

Next, we performed a slightly more difficult, but still easy test. For targets with more than one crystallographic product, we docked each product to the non-native holo X-ray structure of the target (cross-docking; Supporting Table S5). In all 12 cases, a docking pose within 3Å to the native structure was sampled (within 2Å for 11 cases). The top 3 ranked complex was within 3Å to the native structure in 7 out of 12 cases for the PLOP energy function, and in 8 out of 12 cases for PoseScore. To test whether combining the two scoring functions can further improve the result, we considered an optimal linear combination of the PoseScore and the PLOP potential energy, using the average all-atom RMSD of the top ranked docking poses as the optimization criterion; the optimal linear combination has zero weight for the PLOP potential energy.

Finally, to test whether or not apo structures can be used for predicting the native product docking pose, we docked the two products of GSTM1_HUMAN (GDN and GTD) to its two available apo structures (1XW6 and 1YJ6). The results were comparable to those from cross-docking (Supporting Table S6).

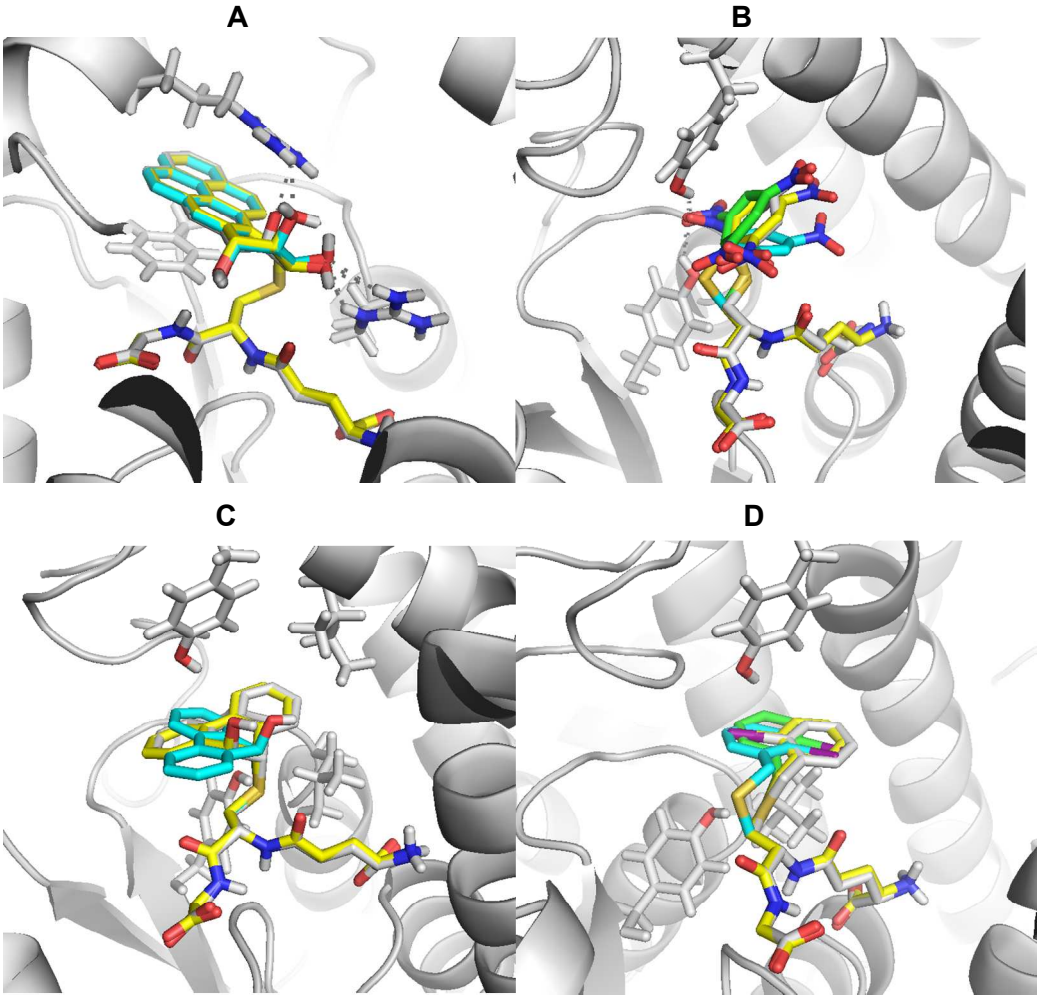


Figure 3. Comparison of the predicted complex between the native product and its holo X-ray structure with the native pose. A) 1F3B with GBX. B) 2F3M with GTD. C) 2GST with GPS. D) 1M9B with IBG. The surface of the receptor is shown, with red and blue corresponding to oxygen and nitrogen atoms. The product is shown in the native configuration (grey), the most accurate sampled configuration (yellow), and the top ranked configuration by the PLOP energy function (light blue) and PoseScore (green).

Docking of native products into comparative models and homologous crystallographic structures

To test whether or not comparative models can be used for docking pose prediction of native products, a total of 62 homology models were generated for the 14 GST targets (Supporting Table S7). We performed native product docking against these models as well as some of their template structures.

To find out to what degree the success of docking pose prediction was limited by sampling *versus* scoring, we plotted the RMSD error of the top ranked docking pose *versus* the RMSD error of the best sampled docking pose (Figure 4C). As expected there was a significant correlation between these two RMSDs; the RMSD error of the best sampled pose is a lower

bound of the RMSD error of the top ranked pose. Even when accurate poses were sampled, the scoring function often did not recognize them (Figure 4C). In 56 out of 62 cases, a docking pose within 3Å was sampled, but only in 22 out of 62 cases, it was ranked in top 3. Therefore, the limiting factor for docking pose prediction is the accuracy of the scoring function used. In this application, the inaccuracy of the scoring function may be exacerbated by errors in the model of the binding site (Figure 6). Despite the difficulties in ranking a near-native pose in the context of comparative model, for 6 of the 14 targets (1GWC, 2AB6, 3LJR, 3O76, 2C4J, 1BYE, Figure 5), the RMSD errors of the top ranked poses by PoseScore were comparable to those against holo X-ray structures.

We then asked the question whether or not there were model properties that predict the best model for docking pose prediction, among multiple models based on different template structures. We found no such predictors. There was essentially no correlation between model's binding site RMSD error and the RMSD error of the top ranked docking pose (Figure 4D). There were only weak correlations between the target-template sequence identity and the RMSD error of the best-sampled docking pose (Figure 4F, R=-0.19 for apo tempalates, R=-0.20 for holo templates). We also calculated the correlation between the RMSD errors of the model binding site and the best sampled docking pose. There was no significant correlation between these two RMSD errors. We also investigated whether or not holo and apo structures performed differently for docking pose prediction. On average, models based on holo template structures were not noticeably better for docking pose prediction than models from apo template structures.

Finally, we tested the accuracy of docking pose prediction using a homologous structure itself, instead of a comparative target model based on it. The test was performed with the two known products of GSTM1_HUMAN (GDN and GTD) and its homologous structure (2C4J, binding site SI and RMSD). The docking pose of GTD had the RMSD error of 2.2Å, and the docking pose of GDN had the RMSD error of 4.7Å, on average not significantly different from that using the corresponding comparative model (3.0Å for GTD and 3.5Å for GDN, binding site RMSD error).

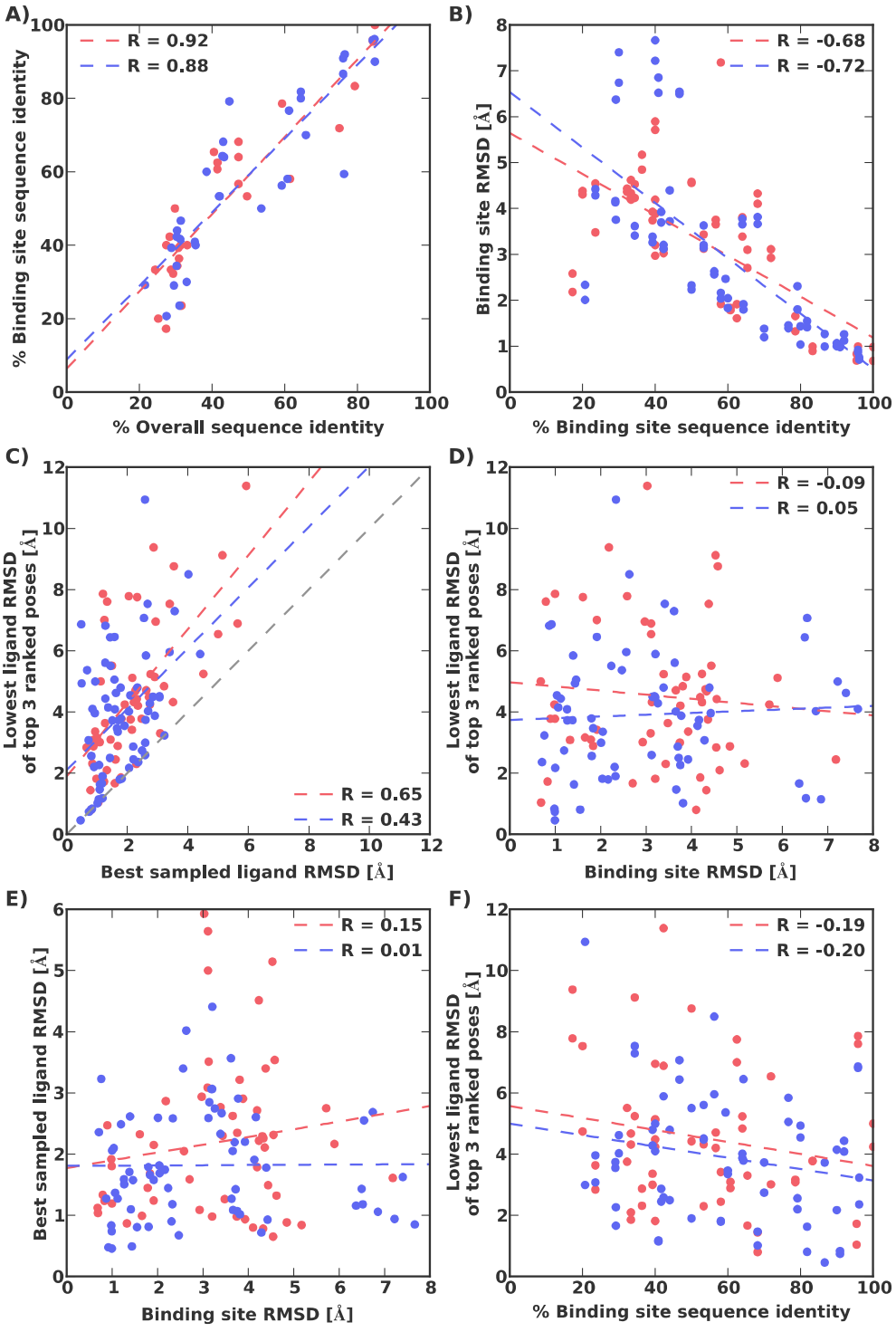


Figure 4. Comparison of comparative model properties and the RMSD errors of the docked poses. Each point represents the property of a single comparative model and/or the RMSD error of the docked pose of its corresponding crystallographic product against this model. Results for models built based on apo and holo template structures are shown in red and

blue, respectively, with the least-squares linear fits shown as dashed lines. The Pearson correlation coefficients R are also shown. The grey dashed line indicates lower bound of the RMSD of the top 3 ranked poses.

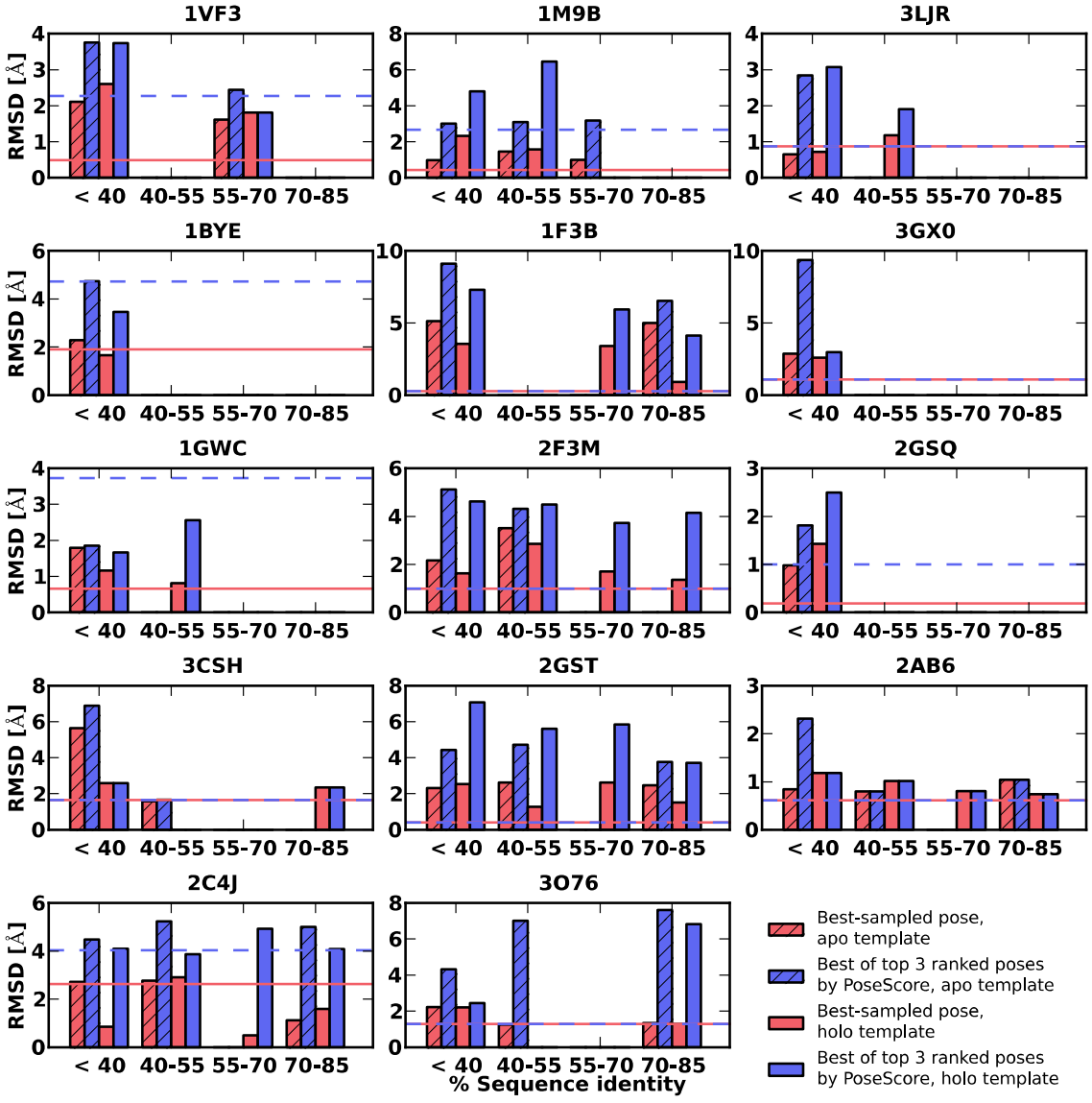


Figure 5. The non-hydrogen atom RMSD error of the docked poses, for the crystallographic product against its corresponding comparative models with different sequence identities. For each target, the RMSD errors of the best-sampled poses using apo templates (stripped red bar), the RMSD errors of the top-ranked poses by PoseScore using apo templates (stripped blue bar), the RMSD errors of the best-sampled poses using holo templates (red bar), and the RMSD errors of the top-ranked pose by PoseScore using apo templates (blue bar) are shown. For comparison, the RMSD errors of the docked poses using holo X-ray

structures are shown as horizontal lines (blue solid line for the top ranked pose by PoseScore, red dashed line for the best-sampled pose).

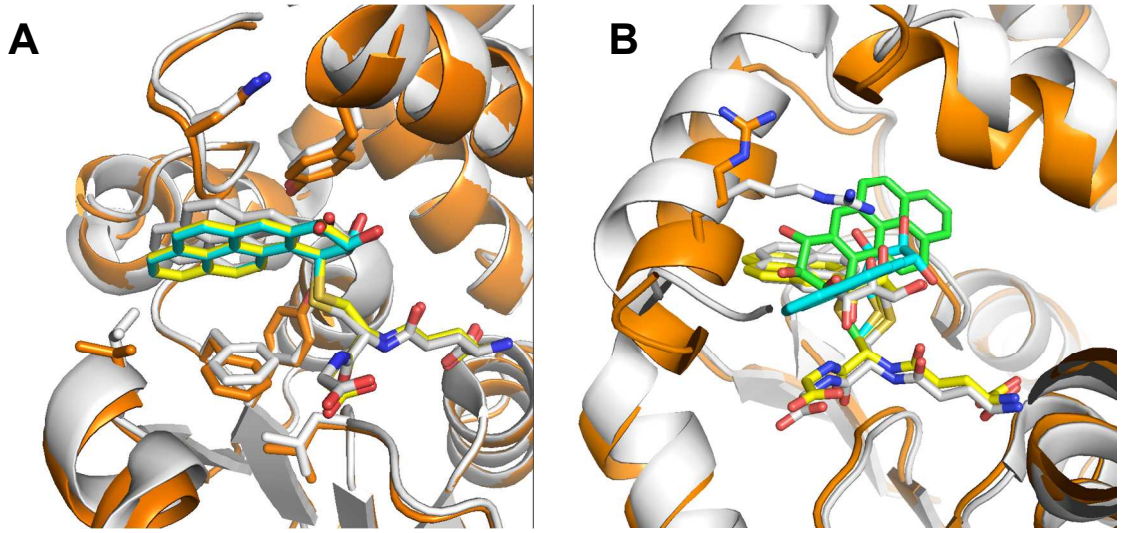


Figure 6. Docking of native products against comparative models. A) Successful docking of GBX onto a model of 3CSH, based on 85% sequence identity to the template 3O76 (Chain A). The ribbon representation of the crystal structure is shown in light grey and that for the model in orange. The C-α RMSD error of the model is 0.64Å, and the non-hydrogen atom RMSD error of the binding site is 0.71Å. The ligand is shown in the stick representation. The native pose is shown in light grey, the best sampled pose (RMSD of 0.90Å) is shown in yellow, the pose top ranked by PLOP (0.92Å) is shown in cyan, and the pose top ranked by POSESCORE (0.92Å) is shown in green (not visible because it is hidden behind the cyan structure). B) Failed docking of GPS onto a model of 1F3B, based on 76% sequence identity to the template 1YDK (Chain B). The ribbon representation of the crystal structure is shown in light grey and that for the model in orange. The C-α RMSD error of the model is 2.42Å, and the non-hydrogen atom RMSD error of the binding site is 2.47Å. The ligand is shown in the stick representation. The native pose is shown in light grey, the best-sampled pose (RMSD of 5.38Å) is shown in yellow, the pose top ranked by PLOP (6.58 Å) is shown in cyan, and the pose top ranked by POSESCORE (0.92Å) is shown in green.

Enrichment for known products by virtual screening against X-ray structures

We now address the question whether or not holo and apo X-ray structures can be used for virtual product screening by covalent docking.

We first applied PLOP to virtual substrate discovery by docking the whole product library against holo structures of 11 GSTs with more than 3 known substrates; this is the minimal number of known ligands needed for computing an enrichment curve. We docked the

products onto each chain of the GST dimers independently, and analyzed the docking results separately for each chain as well as collectively for all chains, by relying on our consensus scoring scheme[27]; consensus scoring combines docking scores from independent virtual screen against multiple chains/structures/models of the same target. In all 11 cases, the enrichment result was better than that from random selection (Figure 7). In 7 of the 11 cases, LogAUC was better than 25. In contrast to non-covalent docking benchmark[27], consensus scoring did not improve LogAUC for any target.

Next, we tested the enrichment of known ligands using two apo structures of GSTM1_HUMAN (1YJ6 and 1XW6; Supporting Figure S2). The averaging logAUC (32.6) is slightly worse than that using the corresponding holo structure (33.3).

Finally, we tested whether or not consensus scoring using multiple structures could improve enrichment for known products by combining virtual screening results from either two apo or two holo structures of GSTM1_HUMAN. Using the two apo structures (1YJ6 and 1XW6), the logAUC for consensus scoring was 34.5. Similarly, using holo structures (2F3M and 1XWK) resulted in logAUC value of 33.6. In both cases, logAUC for consensus scoring was between the best and worst logAUC values of individual virtual screening runs.

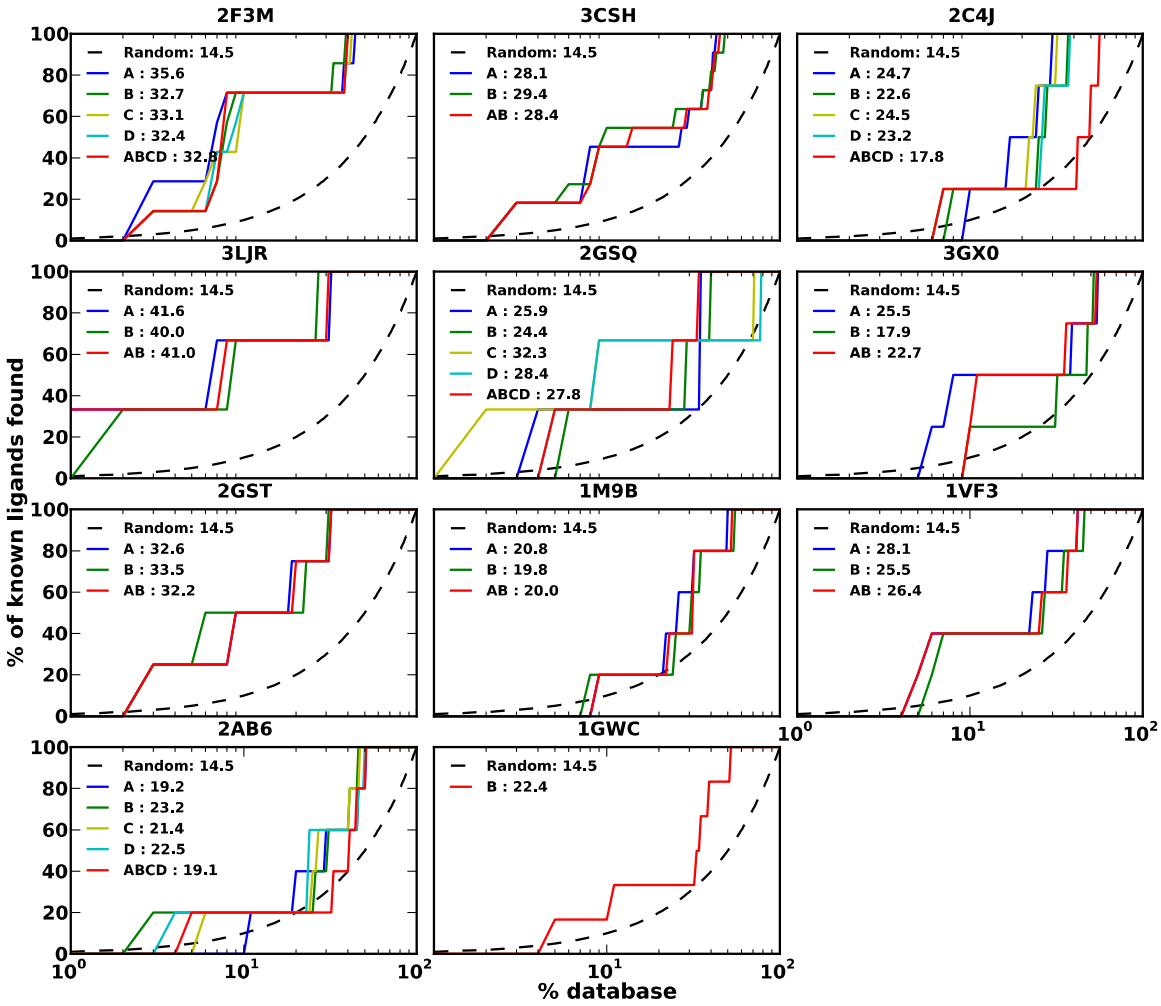


Figure 7. Enrichment curves for the 11 GST targets with 3 or more known products, using holo X-ray structures. For each target, the enrichment curves for the individual chains (blue, green, yellow, and cyan lines), the consensus scoring for multiple chains (red line), and random selection (black dashed line) are shown.

(blue line), the consensus enrichment for multiple models (red line), and random selection (dotted line).

Enrichment for known products by virtual screening against comparative models and homologous crystallographic structures

To evaluate the utility of homology models for discovering substrates of GSTs, the enrichment of known products was determined and compared to that for the crystal structures.

We first tested whether or not there was an apparent correlation between the enrichment and the target-template sequence identity by virtual screening against the 6 comparative models of the target 2F3M. No correlation was observed between logAUC and the target-template sequence identity (Figure 9). The absence of such a correlation was also observed in an extensive non-covalent docking benchmark[27].

Given the absence of a correlation between the enrichment and the target-template sequence identity, we next benchmarked the utility of the best comparative models for virtual screening, for the 11 targets with more than 3 known substrates; the best comparative model is that based on the highest target-template sequence identity. We docked the products onto each chain of the GST dimers independently, and analyzed the docking results separately for each chain as well as collectively for all chains, by relying on our consensus scoring scheme[27]. In all cases, the enrichment result was better than that from random selection (Figure 10). In 5 of the 11 cases, LogAUC was better than 25. Again, the consensus scoring did not improve LogAUC for any target.

We further tested whether or not consensus docking using multiple comparative models can improve the enrichment by combining the docking results from the 6 comparative models for target 2F3M (Figure 9). Consensus scoring resulted in a logAUC of 40.4, which was slightly better than the best logAUC value for individual models (40.3), and was better than logAUC for the holo X-ray structure.

Finally, we tested whether or not homologous X-ray structures could be used in place of comparative models for virtual screening, using a homolog of GSTM1_HUMAN (2C4J). The corresponding logAUC is 30.9 for chain A, 31.6 for chain B, and 31.5 for consensus scoring using chain A and B, which is significantly worse than that for the comparative model based on 2C4J (34.9 for chain A, 35.8 for chain B, and 38.7 for consensus scoring). A similar finding was obtained in the non-covalent docking benchmark[27].

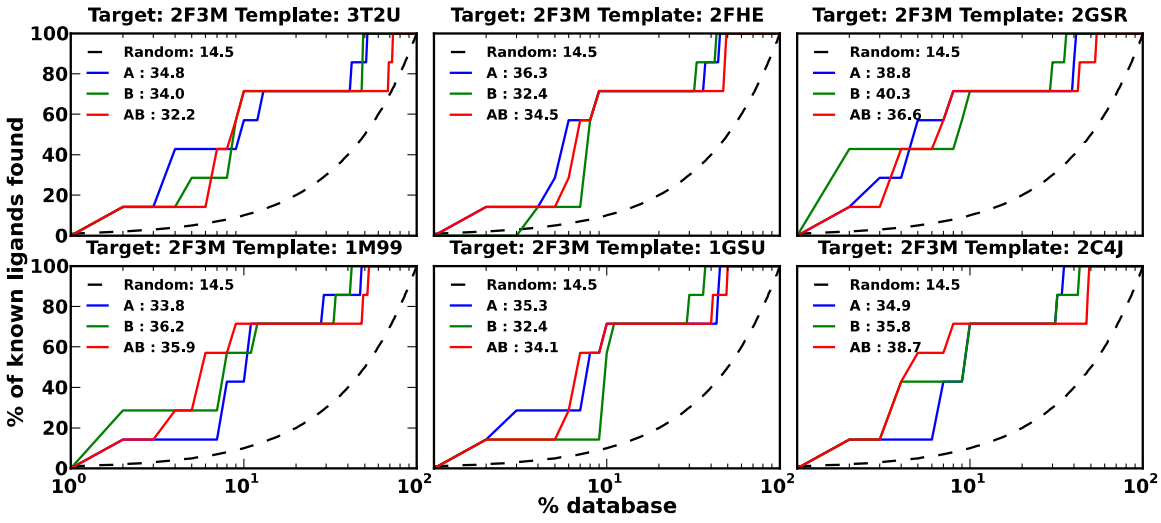


Figure 8. Enrichment curves for target 2F3M using its 6 comparative models based on different apo and holo templates. For each model, the enrichment curves for the individual chains (blue, green, yellow, and cyan lines), the consensus scoring for multiple chains (red line), and random selection (black dashed line) are shown. Target-template sequence identities are as follows: The 2F3M-3T2U and 2F3M-2FHE sequence identities are 33% and 50% respectively (apo templates). Target-template sequence identities are as follows: 33% (2F3M-3T2U, apo), 50% (2F3M-2FHE), 33% (2F3M-2GSR, holo), 42% (2F3M-1M99, holo), 66% (2F3M-1GSU, holo), and 85% (2F3M-2C4J, holo). Binding site RMSD errors are as follows: 5.9Å (2F3M-3T2U, apo), 3.1Å (2F3M-2FHE), 7.4Å (2F3M-2GSR, holo), 3.1Å (2F3M-1M99, holo), 1.4Å (2F3M-1GSU, holo), and 1.1Å (2F3M-2C4J, holo).

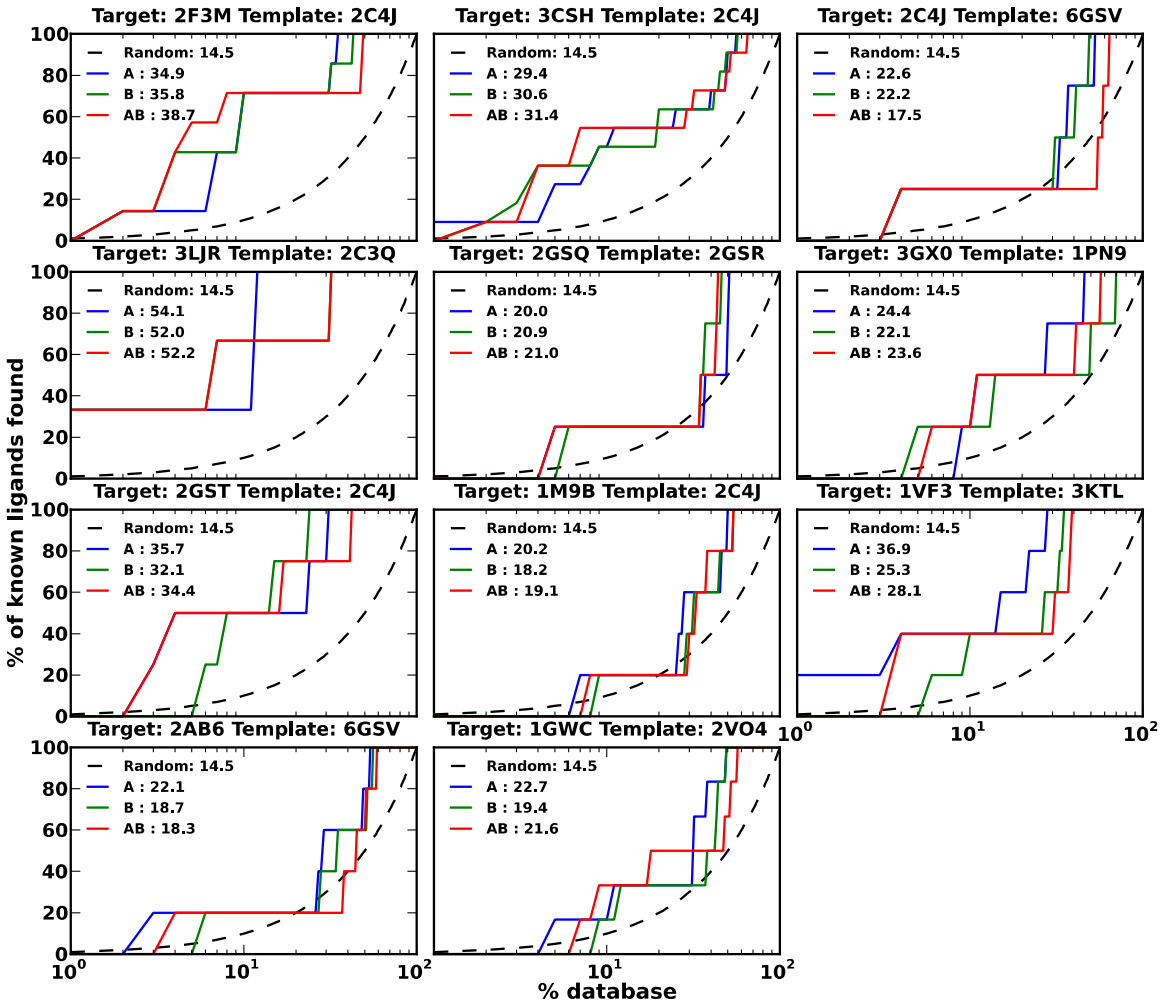


Figure 9. Enrichment curves for the 11 GST targets with 3 or more known products, using their comparative models based on templates with the highest sequence identities. For each target, the enrichment curves for the individual chains (blue, green, yellow, and cyan lines), the consensus scoring for multiple chains (red line), and the random selection (black dashed line) are shown. Target-template sequence identities are as follows: 85% (2F3M-2C4J), 85% (3CSH-2C4J), 77% (2C4J-6GSV), 54% (3LJR-2C3Q), 33% (2GSQ-2GSR), 28% (3GX0-1PN9), 76% (2GST-2C4J), 44% (1M9B-2C4J), 64% (1VF3-3KTL), 76% (2AB6-6GSV), and 46% (1GWC-2VO4).

DISCUSSION

We developed a comparative modeling and covalent docking pipeline for predicting docking poses of known products and for the virtual discovery of substrates, as well as benchmarked it on the GST enzymes. Either a holo or apo structures can be used for predicting docking poses relatively accurately in a majority of cases. While using homology models for predicting docking poses is more difficult, a homology model can also be useful. Specifically, for native holo X-ray structures as receptors, one of the top 3 ranked models is within 3 Å all-

atom RMSD of the native complex in 11 of the 14 test cases (in 8 of 12 cases for cross docking in which a product from another holo structure of the same enzyme is docked; Supporting Tables S4 and S5). For comparative models based on more than 30% sequence identity, one of the top 3 ranked models is within 3Å all-atom RMSD of the native complex in 22 of the 62 test cases (Figure 5). For virtual screening against holo X-ray structures, the enrichment logAUC value is better than random (14.5) in all cases, and better than 25 in 7 of 11 cases (Figure 7). For models based on templates with the highest sequence identity (approximately 30% to 85%), the logAUC is better than 25 in 5 of 11 cases (Figure 10), not significantly different from the crystal structures. In conclusion, we show that covalent docking can be a useful tool for substrate discovery.

Next, we discuss three points in turn. First, we assess to what degree the accuracy of the docking poses and ranking of alternative products is limited by the degree of structure sampling and the accuracy of scoring. Second, we discuss the utility of comparative models for docking and virtual screening. Third, we conclude by commenting on docking transition state conformations of ligands.

Sampling *versus* Scoring

To predict an accurate docking pose, two conditions need to be satisfied: (i) a near-native structure needs to be generated by the sampling procedure and (ii) an accurate sample structure needs to be scored better than the inaccurate structures by the scoring function. Using X-ray structures as receptors, a near-native docking pose (<3Å) was sampled in all cases (Supporting Tables S4 and S5); thus, the degree of sampling does not limit the prediction accuracy. But in 21% of the native docking cases and in 33% of the cross docking cases, a near-native model was not ranked in top 3 by PoseScore. Using homology models as receptors, a near-native docking pose was sampled in 56 of 62 cases but was only ranked in top 3 in 22 of 62 cases (Figure 5). The reason is that our scoring function is not always accurate enough to differentiate between accurate and inaccurate docking poses.

To rank different substrates, we used the median PLOP energy of their product-enzyme complexes. We chose the median instead of the lowest PLOP energy because it resulted in a significantly higher average logAUC (27.3 vs 18.9). A likely reason is that the energy estimate of a single configuration is noisier than the median, compensating for the lack of accounting for the proper physics of the problem. In the future, other function such as the Boltzmann average of the sampled energies will be explored.

Although PoseScore re-ranking of the docking poses generated by optimizing the PLOP energy further improved the accuracy of the top scoring docked poses, this result does not imply that PoseScore is more accurate than the PLOP energy. It is possible that re-ranking by PLOP would also improve the accuracy of the poses generated by optimizing PoseScore; for technical reasons, it is not straightforward to actually make this test. Our result indicates merely that the two scoring functions are different, imperfect, and that PoseScore contains

helpful information not present in PLOP, while being silent on whether or not PLOP also contains information not present in PoseScore.

In contrast to non-covalent docking[27], on average, consensus scoring did not improve virtual screening accuracy. A combination of the following two assumptions explains this finding: (i) covalent docking is limited more by scoring than by sampling compared to non-covalent docking; and (ii) consensus scoring works best when sampling is a limiting factor in docking. A major difference between non-covalent and covalent docking is that the latter is constrained by a covalent bond, resulting in an easier sampling problem all other things considered equal. Using multiple receptor structures to mimic receptor flexibility was already shown to lead to more accurate sampled docking poses in non-covalent docking[27]. Thus, the disadvantage of a larger number of decoys produced by using multiple templates, which increases the burden on the scoring function for identifying the most accurate pose, is outweighed by the more accurately sampled poses that are more likely to be ranked highly by the scoring function. In contrast, if scoring is limiting and sampling is not, as is the case in covalent docking (cf, our native docking and cross-docking tests), considering poses from multiple receptor structures is not likely to improve the results, as observed."

Given the demonstrated limitations of current scoring, the accuracy of ranking ligand poses and ligands can be improved by more accurate scoring functions. On the one hand, the ability of the PLOP potential energy to rank the native state (or at least a near native state) may be improved by estimating the free energy of the native state[66]. On the other hand, the accuracy of the statistical potential could be improved by applying the Bayesian framework for inferring Statistically Optimized Atomic Potentials (SOAP)[67]. This framework (1) uses multiple data-driven "recovery" functions based on probability theory, without recourse to questionable statistical mechanical assumptions about statistical potentials, (2) restrains the relative orientation between two covalent bonds instead of a simple distance between two atoms, in an effort to capture orientation-dependent interactions such as hydrogen bonds, (3) performs Bayesian smoothing for estimating the underlying smooth distributions from noisy observations, (4) applies Bayesian inference for calculating parameter values that maximize the posteriori probability, and (5) benefits from Bayesian model selection based on Bayesian predictive densities, in an effort to improve the generalizability of the derived statistical potentials.

Comparative models for docking and virtual screening

Using comparative models for predicting docking poses often produces inaccurate models (Figures 4 and 5). One possible reason is that covalent docking is sensitive to minor changes in the binding site, especially for residues close to the GSH sulfur atom due to the substrate-receptor covalent bond constraint (Figure 6). Another possible reason is that the active sites of GST enzymes are often defined in part by the C-terminus, which is often flexible or missing in a template structure (Figure 6). Moreover, there was no significant correlation between the binding site non-hydrogen atom RMSD and the docking pose RMSD (Figure 4). The lack

of this correlation may also result from the disproportionate impact of residues close to the GSH sulfur atom on the docking pose.

Virtual screening using comparative models generated results comparable to the X-ray structures (average logAUC of 28.2 *versus* 27.3, better than 25 in 5 *versus* 7 out of 11 cases), consistent with what has been observed when using comparative models for virtual screening by non-covalent docking (average logAUC of 28.7 *versus* 30.6)[27], even though comparative models performed much worse than X-ray structures for docking pose prediction (average non-hydrogen atom RMSD of 3.98Å *versus* 1.83Å). This finding may be rationalized by comparatively worse scoring of both true and decoy products (average PoseScore of -4.2 and 4.1 *versus* -60.0 and -52.2) when docking against inaccurate binding sites in comparative models.

In prospective applications, if some substrates for the target of interest are known, virtual screening using comparative models can be improved by selecting the most enriching model for known substrates[68-70]. If no substrates are known for the target of interest, it might be advantageous to build and dock into multiple structurally diverse models, and analyze the virtual screening results by hand[15]. An experienced user may be able to identify an accurate docking result based on the totality of their knowledge about the proteins of interest.

Reactant, product, intermediate, and transition states

Generally, a substrate of an enzyme is required to have a sufficiently high kcat/Km; a typical threshold is 10^4 $(mol/L)^{-1}s^{-1}$. An enzymatic reaction involves interconversions between pairs of stable states, separated by a transition state; the stable states includes the enzyme and substrate without interaction (E+S), an enzyme-substrate complex intermediate (ES), an enzyme-product intermediate (EP), and the enzyme and product without interaction (E+P)[71]. In general, kcat/Km depends on the free energies of all states. Thus, virtual ligand screening should in principle consider all states, which is generally not the case.

When conversion from ES to EP is the rate-limiting step, kcat is mostly determined by the transition state energy barrier for this step. Enzymes lower this barrier by creating a binding site that is complementary to the transition state of the substrate. Thus, docking a substrate (or a product) in its transition state conformation may be more appropriate than docking a substrate (or a product) in its ground state. Although Quantum Mechanics calculations can model the transition state and estimate the activation energy barrier, running these calculations for the entire virtual screening library is not computationally feasible[72]. Instead, we docked here the reaction product states and achieved relative success. However, the enrichment results could possibly be greatly improved if we can dock the transition state, or a transition state like structure, without recourse to Quantum Mechanics. For GST substrate discovery, compounds similar to the high energy intermediates used for docking to amidohydrolases[73] could be constructed, provided some technical challenges are solved (e.g., building C atoms with 2 partial bonds for a total of 5 bonds). Although enzymes presumably maximize complementarity to the transition state, other states during the

reaction must also be compatible with the potentially flexible binding site of the enzyme. Thus, substrate discovery might also be improved by consideration of docking of all ligand states to the corresponding state(s) of the enzyme.

CONCLUSIONS

Can covalent docking accurately predict docking poses of known ligands, given experimentally determined, homologous or modeled structures in the GST superfamily? Yes, either a holo or apo structures can be used for predicting docking poses accurately in a majority of cases, while using homology models is less successful. Specifically, for native holo X-ray structures as receptors, one of the top 3 ranked models is within 3 Å all-atom RMSD of the native complex in 11 of the 14 cases (in 8 of 12 cases for cross docking). For comparative models as receptors, one of the top 3 ranked models is within 3 Å all-atom RMSD of the native complex in 22 of the 62 test cases.

Can covalent docking accurately predict ligands despite the catalytically promiscuity of many GST enzymes? Often.

What is the difference in the utility of apo, holo, comparative modeling, and homologous structures for virtual screening in the GST superfamily? Holo and apo structures, homologous structures, and comparative models can all enrich for known products. For virtual screening against holo X-ray structures, the enrichment logAUC value is better than random in all cases, and better than 25 in 7 of 11 cases (average logAUC 27.3; Figure 7). Using apo X-ray structures of 2F3M, the average logAUC (32.6) is slightly worse than that using the corresponding holo structure (33.3). For models based on templates with the highest sequence identity (approximately 30% to 85%), the logAUC is better than 25 in 5 of 11 cases (average logAUC 28.2; Figure 10), which is comparable to that for X-ray structure. Using homologous structure of 2F3M, the logAUC (31.3) is significantly worse than that using comparative models (36.5).

Can the virtual screening be improved by consensus scoring, relying on independent screening against multiple holo, apo, comparative modeling, and homologous structures in the GST superfamily? Rarely. Consensus scoring using multiple chains from the same structure/model or multiple structures/models of the same target often does not improve the virtual screening accuracy. Although consensus scoring improves logAUC in comparison to using a single receptor in some cases, it makes it worse in others.

If multiple models are calculated on the basis of different templates, can any of them outperform apo and even holo X-ray structures of the target? If so, can one reliably identify which model will do so, or even perform optimally among a set of modeled structures; are there sequence and/or structural attributes (i.e., the overall target-template sequence identity, the binding site target-template sequence identity, and the predicted accuracy of a model) that reliably predict the accuracy of ligand docking? For docking pose prediction, there are a few cases where comparative models

outperformed the corresponding holo X-ray structure (Figure 5). However, we did not find a predictor that could reliably predict the accuracy of the docked pose. For virtual screening, some models also outperform the corresponding holo X-ray structure (Figures 9 and 10), but we did not find a correlation between logAUC and the binding site sequence identity or the binding site RMSD errors.

ACKNOWLEDGMENTS

We are grateful to Dr. Ben Webb for his help with computing infrastructure. This work was supported by the NIH grant U54 GM093342 to AS, RNA, BKS, MPJ, and PCB.

Supporting Information

The list of GST targets used for docking, their crystal ligands and known substrates, the detailed results of using crystal structures and comparative models for docking pose prediction, and the enrichment curves for GSTM1_HUMAN using its apo structures, the target-template pairs used to generate comparative models, the correlations between the Z-DOPE score and the model's RMSD are included as Supporting Information. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., and Friesner, R. A. (2004) A hierarchical approach to all-atom protein loop prediction, *Proteins 55*, 351-367.
2. Mannervik, B., and Danielson, U. H. (1988) Glutathione transferases--structure and catalytic activity, *CRC critical reviews in biochemistry 23*, 283-337.
3. Armstrong, R. N. (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases, *Chemical research in toxicology 10*, 2-18.
4. Armstrong, R. N. (1998) Mechanistic imperatives for the evolution of glutathione transferases, *Current opinion in chemical biology 2*, 618-623.
5. Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005) Glutathione transferases, *Annu. Rev. Pharmacol. Toxicol. 45*, 51-88.
6. Atkinson, H. J., and Babbitt, P. C. (2009) Glutathione transferases are structural and functional outliers in the thioredoxin fold, *Biochemistry 48*, 11108-11116.
7. Copley, S. D., Novak, W. R., and Babbitt, P. C. (2004) Divergence of function in the thioredoxin fold suprafamily: evidence for evolution of peroxiredoxins from a thioredoxin-like ancestor, *Biochemistry 43*, 13981-13995.
8. Abagyan, R., and Totrov, M. (2001) High-throughput docking for lead generation, *Current opinion in chemical biology 5*, 375-382.

9. Apostolakis, J., Pluckthun, A., and Caflisch, A. (1998) Docking small ligands in flexible binding sites, *Journal of computational chemistry 19*, 21-37.
10. Cavasotto, C. N., and Orry, A. J. (2007) Ligand docking and structure-based virtual screening in drug discovery, *Current topics in medicinal chemistry 7*, 1006-1014.
11. Schroder, J., Klinger, A., Oellien, F., Marhofer, R. J., Duszenko, M., and Selzer, P. M. (2013) Docking-Based Virtual Screening of Covalently Binding Ligands: An Orthogonal Lead Discovery Approach, *Journal of medicinal chemistry*.
12. Chen, Y., and Shoichet, B. K. (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery, *Nature chemical biology 5*, 358-364.
13. de Graaf, C., Oostenbrink, C., Keizers, P. H. J., van der Wijst, T., Jongejan, A., and Vemleulen, N. P. E. (2006) Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking, *Journal of medicinal chemistry 49*, 2417-2430.
14. Gohda, K., Teno, N., Wanaka, K., and Tsuda, Y. (2012) Predicting subsite interactions of plasmin with substrates and inhibitors through computational docking analysis, *Journal of enzyme inhibition and medicinal chemistry 27*, 571-577.
15. Fan, H., Hitchcock, D. S., Seidel, R. D., 2nd, Hillerich, B., Lin, H., Almo, S. C., Sali, A., Shoichet, B. K., and Raushel, F. M. (2013) Assignment of pterin deaminase activity to an enzyme of unknown function guided by homology modeling and docking, *Journal of the American Chemical Society 135*, 795-803.
16. Wallrapp, F. H., Pan, J. J., Ramamoorthy, G., Almonacid, D. E., Hillerich, B. S., Seidel, R., Patskovsky, Y., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Poulter, C. D. (2013) Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily, *Proceedings of the National Academy of Sciences of the United States of America 110*, E1196-1202.
17. Carlsson, J., Coleman, R. G., Setola, V., Irwin, J. J., Fan, H., Schlessinger, A., Sali, A., Roth, B. L., and Shoichet, B. K. (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure, *Nature chemical biology 7*, 769-778.
18. Schlessinger, A., Geier, E., Fan, H., Irwin, J. J., Shoichet, B. K., Giacomini, K. M., and Sali, A. (2011) Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET, *Proceedings of the National Academy of Sciences of the United States of America 108*, 15810-15815.
19. Jacobson, M., and Sali, A. (2004) Comparative protein structure modeling and its applications to drug discovery, *Annu Rep Med Chem 39*, 259-276.
20. Bissantz, C., Bernard, P., Hibert, M., and Rognan, D. (2003) Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets?, *Proteins-Structure Function and Genetics 50*, 5-25.
21. Evers, A., and Klebe, G. (2004) Ligand-supported homology modeling of G-protein-coupled receptor sites: Models sufficient for successful virtual screening, *Angewandte Chemie-International Edition 43*, 248-251.

22. Evers, A., and Klebe, G. (2004) Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model, *Journal of medicinal chemistry 47*, 5381-5392.
23. Radestock, S., Weil, T., and Renner, S. (2008) Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring, *Journal of chemical information and modeling 48*, 1104-1117.
24. Kalyanaraman, C., Imker, H. J., Federov, A. A., Federov, E. V., Glasner, M. E., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2008) Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening, *Structure (London, England : 1993) 16*, 1668-1677.
25. Diller, D. J., and Li, R. X. (2003) Kinases, homology models, and high throughput docking, *Journal of medicinal chemistry 46*, 4638-4647.
26. Nguyen, T. L., Gussio, R., Smith, J. A., Lannigan, D. A., Hecht, S. M., Scudiero, D. A., Shoemaker, R. H., and Zaharevitz, D. W. (2006) Homology model of RSK2 N-terminal kinase domain, structure-based identification of novel RSK2 inhibitors, and preliminary common pharmacophore, *Bioorganic & Medicinal Chemistry 14*, 6097-6105.
27. Fan, H., Irwin, J. J., Webb, B. M., Klebe, G., Shoichet, B. K., and Sali, A. (2009) Molecular Docking Screens Using Comparative Models of Proteins, *Journal of chemical information and modeling 49*, 2512-2527.
28. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic acids research 40*, D109-114.
29. Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P., and Karp, P. D. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, *Nucleic acids research 40*, D742-753.
30. OEChem, T. (2010) version 1.7. 4.3; OpenEye Scientific Software Inc, *Santa Fe, NM*.
31. James, C. A., Weininger, D., and Delany, J. (2004) Daylight theory manual, *Daylight chemical information systems 3951*.
32. Prade, L., Huber, R., and Bieseler, B. (1998) Structures of herbicides in complex with their detoxifying enzyme glutathione S-transferase - explanations for the selectivity of the enzyme in plants, *Struct Fold Des 6*, 1445-1452.
33. Cardoso, R. M., Daniels, D. S., Bruns, C. M., and Tainer, J. A. (2003) Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from Schistosoma japonicum, *Proteins 51*, 137-146.
34. Gu, Y., Singh, S. V., and Ji, X. (2000) Residue R216 and catalytic efficiency of a murine class alpha glutathione S-transferase toward benzo [a] pyrene 7 (R), 8 (S)-diol 9 (S), 10 (R)-epoxide, *Biochemistry 39*, 12552-12557.

35. Thom, R., Cummins, I., Dixon, D. P., Edwards, R., Cole, D. J., and Lapthorn, A. J. (2002) Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification, *Biochemistry 41*, 7008-7020.
36. Norrgard, M. A., Ivarsson, Y., Tars, K., and Mannervik, B. (2006) Alternative mutations of a positively selected residue elicit gain or loss of functionalities in enzyme evolution, *Proceedings of the National Academy of Sciences of the United States of America 103*, 4876-4881.
37. Federici, L., Lo Sterzo, C., Pezzola, S., Di Matteo, A., Scaloni, F., Federici, G., and Caccuri, A. M. (2009) Structural basis for the binding of the anticancer compound 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol to human glutathione s-transferases, *Cancer research 69*, 8025-8034.
38. Patskovsky, Y., Patskovska, L., Almo, S. C., and Listowsky, I. (2006) Transition state model and mechanism of nucleophilic aromatic substitution reactions catalyzed by human glutathione S-transferase M1a-1a, *Biochemistry 45*, 3852-3862.
39. Rowe, J. D., Patskovsky, Y. V., Patskovska, L. N., Novikova, E., and Listowsky, I. (1998) Rationale for reclassification of a distinctive subdivision of mammalian class Mu glutathione S-transferases that are primarily expressed in testis, *The Journal of biological chemistry 273*, 9593-9601.
40. Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M., and Jornvall, H. (1985) Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties, *Proceedings of the National Academy of Sciences of the United States of America 82*, 7202-7206.
41. Ji, X., von Rosenvinge, E. C., Johnson, W. W., Armstrong, R. N., and Gilliland, G. L. (1996) Location of a potential transport binding site in a sigma class glutathione transferase by x-ray crystallography, *Proceedings of the National Academy of Sciences of the United States of America 93*, 8208-8213.
42. Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N., and Gilliland, G. L. (1995) Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods, *Biochemistry 34*, 5317-5328.
43. Ji, X., Armstrong, R. N., and Gilliland, G. L. (1993) Snapshots along the reaction coordinate of an SNAr reaction catalyzed by glutathione transferase, *Biochemistry 32*, 12949-12954.
44. Ji, X., Johnson, W. W., Sesay, M. A., Dickert, L., Prasad, S. M., Ammon, H. L., Armstrong, R. N., and Gilliland, G. L. (1994) Structure and function of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene, *Biochemistry 33*, 1043-1052.
45. Reinemer, P., Dirr, H. W., Ladenstein, R., Huber, R., Lo Bello, M., Federici, G., and Parker, M. W. (1992) Three-dimensional structure of class pi glutathione S-

transferase from human placenta in complex with S-hexylglutathione at 2.8 A resolution, *Journal of molecular biology 227*, 214-226.

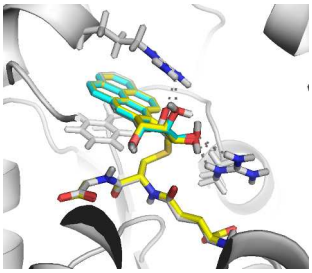
46. Parker, L. J., Ciccone, S., Italiano, L. C., Primavera, A., Oakley, A. J., Morton, C. J., Hancock, N. C., Bello, M. L., and Parker, M. W. (2008) The anti-cancer drug chlorambucil as a substrate for the human polymorphic enzyme glutathione transferase P1-1: kinetic properties and crystallographic characterisation of allelic variants, *Journal of molecular biology 380*, 131-144.
47. Ji, X., Tordova, M., O'Donnell, R., Parsons, J. F., Hayden, J. B., Gilliland, G. L., and Zimniak, P. (1997) Structure and function of the xenobiotic substrate-binding site and location of a potential non-substrate-binding site in a class pi glutathione S-transferase, *Biochemistry 36*, 9690-9702.
48. Liu, L. F., Liaw, Y. C., and Tam, M. F. (1997) Characterization of chicken-liver glutathione S-transferase (GST) A1-1 and A2-2 isoenzymes and their site-directed mutants heterologously expressed in Escherichia coli: identification of Lys-15 and Ser-208 on cGSTA1-1 as residues interacting with ethacrynic acid, *The Biochemical journal 327 (Pt 2)*, 593-600.
49. Wadington, M. C., Ladner, J. E., Stourman, N. V., Harp, J. M., and Armstrong, R. N. (2009) Analysis of the structure and function of YfcG from Escherichia coli reveals an efficient and unique disulfide bond reductase, *Biochemistry 48*, 6559-6561.
50. Kanai, T., Takahashi, K., and Inoue, H. (2006) Three distinct-type glutathione S-transferases from Escherichia coli important for defense against oxidative stress, *Journal of biochemistry 140*, 703-711.
51. Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G., and Parker, M. W. (1998) Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site, *Structure (London, England : 1993) 6*, 309-322.
52. Tan, K. L., Chelvanayagam, G., Parker, M. W., and Board, P. G. (1996) Mutagenesis of the active site of the human Theta-class glutathione transferase GSTT2-2: catalysis with different substrates involves different residues, *The Biochemical journal 319 (Pt 1)*, 315-321.
53. McManus, G., Costa, M., Canals, A., Coll, M., and Mantle, T. J. (2011) Site-directed mutagenesis of mouse glutathione transferase P1-1 unlocks masked cooperativity, introduces a novel mechanism for 'ping pong' kinetic behaviour, and provides further structural evidence for participation of a water molecule in proton abstraction from glutathione, *The FEBS journal 278*, 273-281.
54. Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution, *J Comput Aid Mol Des 24*, 591-604.
55. Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007) Epik: a software program for pK (a) prediction and

protonation state generation for drug-like molecules, *J Comput Aid Mol Des 21*, 681-691.

56. Ligprep, V. 2.3,(2009) Schrodinger, *LLC, New York, NY*.
57. Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The Enzyme Function Initiative, *Biochemistry 50*, 9950-9962.
58. Soding, J. (2005) Protein homology detection by HMM-HMM comparison, *Bioinformatics 21*, 951-960.
59. Apweiler, R., Martin, M. J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Alpi, E., Antunes, R., Arganiska, J., Casanova, E. B., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chan, W. M., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dimmer, E., Fazzini, F., Gane, P., Fedotov, A., Castro, L. G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Jacobsen, J., Jones, R., Legge, D., Liu, W. D., Luo, J., MacDougall, A., Mutowo, P., Nightingale, A., Orchard, S., Patient, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G. Y., Rosanoff, S., Sawford, T., Sehra, H., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Corbett, M., Donnelly, M., van Rensburg, P., Goujon, M., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Binz, P. A., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., de Castro, E., Cerutti, L., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., James, J., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lew, J., Lieberherr, D., Martin, X., Masson, P., Morgat, A., Neto, T., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Zerara, M., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C. M., Chen, Y. X., Huang, H. Z., Kukreja, A., Laiho, K., McGarvey, P., Natale, D. A., Natarajan, T. G., Roberts, N. V., Suzek, B. E., Vinayaka, C. R., Wang, Q. H., Wang, Y. Q., Yeh, L. S., Yerramalla, M. S., Zhang, J., and Consortium, U. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013, *Nucleic acids research 41*, D43-D47.
60. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *Journal of molecular biology 234*, 779-815.
61. Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures, *Protein Science 15*, 2507-2524.
62. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *Journal of chemical information and modeling 50*, 572-584.

63. Zhu, K., Shirts, M. R., and Friesner, R. A. (2007) Improved methods for side chain and loop predictions via the protein local optimization program: Variable dielectric model for implicitly improving the treatment of polarization effects, *Journal of Chemical Theory and Computation 3*, 2108-2119.
64. Fan, H., Schneidman-Duhovny, D., Irwin, J. J., Dong, G., Shoichet, B. K., and Sali, A. (2011) Statistical potential for modeling and ranking of protein-ligand interactions, *Journal of chemical information and modeling 51*, 3078-3092.
65. Sippl, M. J. (1993) Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures, *Journal of computer-aided molecular design 7*, 473-501.
66. Chang, C. E., Chen, W., and Gilson, M. K. (2007) Ligand configurational entropy and protein binding, *Proceedings of the National Academy of Sciences of the United States of America 104*, 1534-1539.
67. Dong, G. Q., Fan, H., Schneidman-Duhovny, D., Webb, B. M., and Sali, A. (2013) Optimized atomic statistical potentials: Assessment of protein interfaces and loops, *Manuscript submitted for publication*.
68. Kamat, S. S., Bagaria, A., Kumaran, D., Holmes-Hampton, G. P., Fan, H., Sali, A., Sauder, J. M., Burley, S. K., Lindahl, P. A., Swaminathan, S., and Raushel, F. M. (2011) Catalytic mechanism and three-dimensional structure of adenine deaminase, *Biochemistry 50*, 1917-1927.
69. Kamat, S. S., Fan, H., Sauder, J. M., Burley, S. K., Shoichet, B. K., Sali, A., and Raushel, F. M. (2011) Enzymatic deamination of the epigenetic base N-6-methyladenine, *Journal of the American Chemical Society 133*, 2080-2083.
70. Goble, A. M., Fan, H., Sali, A., and Raushel, F. M. (2011) Discovery of a cytokinin deaminase, *ACS chemical biology 6*, 1036-1040.
71. Nelson, D. L., Lehninger, A. L., and Cox, M. M. (2008) *Lehninger principles of biochemistry*, Macmillan.
72. Dykstra, C., Frenking, G., Kim, K., and Scuseria, G. (2011) *Theory and Applications of Computational Chemistry: the first forty years*, Access Online via Elsevier.
73. Hermann, J. C., Ghanem, E., Li, Y., Raushel, F. M., Irwin, J. J., and Shoichet, B. K. (2006) Predicting substrates by docking high-energy intermediates to enzyme structures, *Journal of the American Chemical Society 128*, 15882-15891.

For Table of Contents Use Only



Prediction of substrates for glutathione transferases by covalent docking

Guang Qiang Dong, Sara Calhoun, Hao Fan, Chakrapani Kalyanaraman, Megan C. Branch, Susan T. Mashiyama, Nir London, Matthew P. Jacobson, Patricia C. Babbitt, Brian K. Shoichet, Richard N. Armstrong*, Andrej Sali*