Screening Drug-Like Compounds by Docking to Homology Models: A Systematic Study

Visvaldas Kairys,† Miguel X. Fernandes,† and Michael K. Gilson*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland 20850

Received June 13, 2005

In the absence of an experimentally solved structure, a homology model of a protein target can be used instead for virtual screening of drug candidates by docking and scoring. This approach poses a number of questions regarding the choice of the template to use in constructing the model, the accuracy of the screening results, and the importance of allowing for protein flexibility. The present study addresses such questions with compound screening calculations for multiple homology models of five drug targets. A central result is that docking to homology models frequently yields enrichments of known ligands as good as that obtained by docking to a crystal structure of the actual target protein. Interestingly, however, standard measures of the similarity of the template used to build the homology model to the targeted protein show little correlation with the effectiveness of the screening calculations, and docking to the template itself often is as successful as docking to the corresponding homology model. Treating key side chains as mobile produces a modest improvement in the results. The reasons for these sometimes unexpected results, and their implications for future methodologic development, are discussed.

INTRODUCTION

Virtual screening of candidate ligands by docking and scoring can only be carried out if a three-dimensional structural model is available for the target protein. The model usually is provided by X-ray crystallography or NMR spectroscopy, but sometimes no experimental structure is available. In such cases, comparative modeling can often be used to construct a structural model based upon a template protein, another protein whose amino acid sequence is similar and whose structure is known. Virtual screening based upon such comparative or, more loosely, homology models has helped discover small molecule ligands of, for example, falcipain-2[1], ρ-kinase,[2] and thyroid hormone nuclear receptor β[3]. This approach is likely to become increasingly prevalent as structural genomics projects elucidate the structures of suitable templates for many protein families. It is, therefore, of interest to assess how well homology models, as opposed to experimentally determined structures, work for virtual screening and to characterize the factors that determine the performance of a given homology model in this application.

Several studies have evaluated the applicability of homology models to virtual compound screening. Diller and Li focused on a series of protein kinases, reporting that some homology models perform as well as crystal structures, whereas others perform significantly worse.[4] Bissantz et al. demonstrated the successful use of homology models of three G-protein coupled receptors based upon a single template structure, bovine rhodopsin.[5] McGovern and Shoichet undertook a broader study comparing the efficacy of docking to a homology model, to a crystal structure of the target solved with no bound ligand, and to a crystal structure of the target solved with a bound ligand for 10 protein targets.[6] Overall, the homology models provided significant enrichment of known binders against a background of decoy compounds, but the best results were usually obtained with the crystal structure of the ligand-bound target. It was noted that some homology models performed poorly because of rather small structural errors, such as an incorrect side-chain rotamer. This observation is consistent with a study by DeWeese-Scott and Moult, who analyzed homology models submitted for the Comparative Assessment of Structure Prediction exercise[7] and found that a number of models would yield incorrect ligand−protein contacts due, primarily, to incorrect assignment of a few side-chain rotamers.[8] The study also touches on the relationship between the accuracy of the binding pocket and the sequence similarity between the target and the template but does not provide a systematic analysis because only a single homology model was tested for each target. Oshiro et al.[9] specifically addressed this relationship between sequence similarity and screening results by carrying out virtual screening with eight homology models for factor VIIa, based upon templates with binding-site sequence identities of 37−77%, and four homology models for cyclin dependent kinase 2, based upon templates with binding-site sequence identities of 43−60%. Their paper concludes that homology models based upon templates with greater than 50% sequence identity yield 5-fold better enrichment than homology models based upon templates of lower sequence identity. However, the generality of this observation is not entirely clear because only two systems were studied, and the high sequence identity models actually underperformed the low sequence identity models for one of the systems, cyclin dependent kinase 2. In our own previous study,[10] homology models of carboxypeptidase A and factor Xa were built on the basis of multiple templates of varying similarity to the targets. We found, somewhat surprisingly, that the screening results correlated poorly with sequence identity, as well as with other standard measures of protein similarity. However, the generality of this result is again unclear because of the small number of test cases.

* Corresponding author phone: (240) 314-6217; fax: (240) 314-6255; e-mail gilson@umbi.umd.edu.

† Current address: Centro de Química da Madeira, Universidade da Madeira, Campus da Penteada, 9000-390 Funchal, Portugal.

10.1021/ci050238c CCC: $33.50
Published on Web 12/13/2005

Table 1. Enrichment from Virtual Docking for Various Models of the Target Protein Carboxypeptidase[a]

target/template	PDB code	seq. ID (%)	FASTA E score	bind site seq. ID	bind site seq. simil.	overall RMSD	bind site RMSD	⟨rank⟩	⟨rank⟩ docking to template
carboxypeptidase A	**1cps**[33]	**100**		**100**	**100**	**0.0**	**0.0**	**9.8**	
procarboxypeptidase A2	*1aye*[34]	66.6	2.9×10^{-87}	90	100	1.2	2.6	17.3	25.8
procarboxypeptidase B	*1kwm*[35]	46.5	2.1×10^{-58}	100	100	0.7	2.3	15.6	20.5
carboxypeptidase T	*1obr*[36]	31.8	1.4×10^{-16}	90	100	0.6	2.6	8.5	10.6
procarboxypeptidase A	*1jqq*[37]	30.9	3.4×10^{-28}	90	100	1.4	2.4	12.9	17.4

[a] Top row: results for crystal structure of the target (bold font). Subsequent rows: homology models built from specified templates. target/template: Name of the crystal structure used for docking (top row) or the template used for the homology model (other rows). seq. ID: Percent FASTA sequence identity of the template to target. bind site seq. ID: Percent sequence identity of the residues forming the binding site. bind site seq. simil.: Percent similarity of the residues forming the binding site. overall RMSD: Root-mean-square deviation (Å) of all α-carbon coordinates of the homology model from the crystal structure of the target. bind site RMSD: Root-mean-square deviation (Å) of all binding-site atoms from the crystal structure of the target. ⟨rank⟩: Average rank (percent) of known binders in the ranked list of binders and decoys (see Methods) for docking to the specified structure. ⟨rank⟩ docking to template: Average rank (percent) of known binders in the ranked list of binders and decoys for docking to the template structure itself, rather than to the homology model. An italic PDB code in the second column indicates a structure solved with no bound ligand.

Table 2. Enrichment from Virtual Docking for Various Models of the Target Protein Factor Xa[a]

target/template	PDB code	seq. ID (%)	FASTA E score	bind site seq. ID	bind site seq. simil.	overall RMSD	bind site RMSD	⟨rank⟩	⟨rank⟩ docking to template
factor Xa	**1g2l**[38]	**100**		**100**	**100**	**0.0**	**0.0**	**2.6**	
factor IXa	1rfn[39]	43.8	6.10×10^{-43}	68.4	89.5	1.8	2.1	3.1	3.7
α-thrombin	1doj[40]	40.7	5.80×10^{-17}	73.7	89.5	1.9	1.3	1.9	7.1
factor VIIa	1kli[41]	38.8	9.00×10^{-41}	63.2	84.2	2.2	2.0	6.0	4.4
protein C	1aut[42]	34.9	2.80×10^{-35}	73.7	84.2	2.3	1.0	2.8	13.4

[a] See Table 1 for details.

Table 3. Enrichment from Virtual Docking for Various Models of the Target Protein PPARα[a]

target/template	PDB code	seq. ID (%)	FASTA E score	bind site seq. ID	bind site seq. simil.	overall RMSD	bind site RMSD	⟨rank⟩	⟨rank⟩ docking to template
PPARα	**1i7g**[28]	**100**		**100**	**100**	**0.0**	**0.0**	**11.7**	
PPARγ	2prg[43]	62.0	2.00×10^{-71}	60.0	80.0	1.0	1.8	9.5	12.3
RORα	1n83[44]	30.8	1.40×10^{-15}	13.3	46.7	2.5	4.4	23.6	13.1
LXRβ	1pq9[45]	28.9	2.80×10^{-15}	13.3	26.7	1.5	4.0	10.4	48.3

[a] See Table 1 for details.

In summary, although a number of groups have examined the use of homology models in virtual screening, there has been no broad and systematic assessment of the value of homology models in virtual screening and of the factors that determine the usefulness of a given model.

The present study, thus, seeks to address a number of issues in this area, employing a wide enough range of systems and models that generalizations can be made with reasonable confidence. Specific issues include the relevance of the template's similarity to the target and the use of the template itself instead of a homology model. In addition, on the basis of prior indications that the details of side-chain rotamers may strongly determine the usefulness of a structural model, even one based upon a highly similar template, we examine the possibility that allowing limited side-chain motion may improve virtual screening results based upon homology models.

METHODS

Target Proteins and Homology Models. Five different globular proteins known as drug targets were selected on the basis of the availability of proteins of similar sequence in the Protein Data Bank[11,12] (PDB) for use as structural templates and also on the basis of the availability of a selection of known ligands. The targets and the numbers of associated ligands are as follows: human carboxypeptidase A (CPA; 13 ligands), human coagulation factor Xa (FXa; 16 ligands), human peroxisome proliferator-activated receptor α (PPARα; 13 ligands), human cyclin dependence kinase 2 (CDK2; 22 ligands), and *Torpedo californica* acetylcholinesterase (AChE; 13 ligands). A crystal structure of each target was drawn from the PDB, and three to five templates used for model building also were found in the PDB via FASTA search.[13] The left-most columns of Tables 1−5 list the PDB codes of the targets (first line) and the templates (subsequent lines). Most of the structures, both targets and templates, were solved with a bound ligand; Tables 1−5 list in italics the PDB codes of those which were not. As previously described,[10] homology models were constructed via sequence alignment with FASTA, direct transfer of backbone and conserved side-chain torsion angles, construction of nonconserved side-chain conformations with the program SCWRL,[14] and repair of sequence gaps and insertions (which were generally remote from the binding site) with the program Quanta 2000 (Accelrys, Inc.). For all receptor structures, missing polar hydrogen atoms were added to the protein structures, and their positions were optimized in the absence of any ligand with a gradient energy minimization method until the energy change between steps was lower than 0.01 kcal/mol.

Several measures of the similarity of the template protein to the target were computed: sequence identity, sequence similarity, and E score from a FASTA search within PDB;

Table 4. Enrichment from Virtual Docking for Various Models of the Target Protein Cyclin Dependent Kinase 2[a]

target/template	PDB code	seq. ID (%)	FASTA E score	bind site seq. ID	bind site seq. simil.	overall RMSD	bind site RMSD	⟨rank⟩	⟨rank⟩ docking to template
CDK2	**1h1s**[46]	**100**		**100**	**100**	**0.0**	**0.0**	**22**	
CDK6	*1blx*[47]	48.2	2.68×10^{-50}	71.4	85.7	2.9	2.7	39	26
ERK2	3erk[48]	37.0	8.80×10^{-36}	42.9	85.7	1.5	2.7	30	23
PKA	1q8u[49]	31.2	6.50×10^{-14}	28.6	71.4	1.5	2.8	35	30
c-Abl	1iep[50]	28.6	1.00×10^{-13}	28.6	64.3	4.1	4.0	43	29

[a] See Table 1 for details.

Table 5. Enrichment from Virtual Docking for Various Models of the Target Protein Acetylcholinesterase[a]

target/template	PDB code	seq. ID (%)	FASTA E score	bind site seq. ID	bind site seq. simil.	overall RMSD	bind site RMSD	⟨rank⟩	⟨rank⟩ docking to template
AChE	**1e66**[51]	**100**		**100**	**100**	**0.0**	**0.0**	**16.7**	
BuChE	1p0m[52]	53.3	1.30×10^{-132}	80	90	1.4	0.2	19.0	27.8
carboxylesterase 1	1mx1[53]	36.2	1.80×10^{-57}	50	90	3.5	0.9	40.1	35.1
triacylglycerol lipase	1lps[54]	32.8	1.20×10^{-28}	60	80	5.8	2.0	48.0	31.0
cholesterol esterase	1cle[55]	32.8	1.40×10^{-26}	60	70	5.9	2.0	27.2	30.3
para-nitrobenzyl esterase	*1qe3*[56]	31.4	1.70×10^{-37}	50	70	6.2	2.7	25.4	27.7

[a] See Table 1 for details.

the root-mean-square deviation in Å (RMSD) for all α-carbon atoms; and the sequence identity, sequence similarity, and all-atom RMSD of only the binding-site residues. The binding-site residues were taken to be all residues within 5 Å from a reasonably large ligand bound in the target structure.

Known Ligands and Decoy Compounds. From 13 to 22 known ligands were included for each targeted protein. For AChE, FXa, and CDK2, initial ligand structures were drawn from crystal structures of complexes with the targeted receptor. For AChE, only ligands that bind to the deep part of the active site gorge were chosen. Some also form interactions with the peripheral anionic site,[15] but none bind exclusively to the peripheral anionic site. Too few ligands were found in the PDB for CPA and PPARα, so several additional ligands were drawn from the literature. The mean ± standard deviations of the molecular weights (Da) of each ligand set are as follows: 292 ± 126 for CPA; 473 ± 47 for FXa; 395 ± 104 for PPARα; 352 ± 55 for CDK2; and 287 ± 93 for AChE. The corresponding figures for their net charges are −1.38 ± 0.62, 0.38 ± 0.60, −0.92 ± 0.46, 0.09 ± 0.48, and 1.08 ± 0.49.

Initial conformations of the known binders were obtained as follows. For ligands with available crystal structures, hydrogen atoms were added to the crystal conformations and the energy was minimized with respect to only the hydrogen positions with 200 steps of the steepest descent algorithm with the program Quanta 2000. Ligands not available with a crystal structure were drawn on-screen with Quanta 2000, and their 3D conformations were optimized using 500 steps of a conjugate gradient. It should be emphasized that the starting torsion angles of the ligands do not influence the docking procedure because rotatable bonds are initially sampled randomly over a range of 0−360°. Also, for each target, at most one ligand was drawn from the crystal structure used in the present docking studies; the other ligands, thus, were cross-docked into the receptor structure used here. For all ligands, all amines and amidines were treated as cationic, all carboxyl groups were treated as anionic, and in-ring nitrogens of 4-aminoquinoline groups were protonated in the AChE inhibitors containing this moiety.

The "diversity set" compounds from the National Cancer Institute (NCI)[16] were employed as decoys presumed not to bind the protein targets. Compounds deemed problematic, such as those containing metal atoms, were removed, leaving 1844 compounds. The 3D conformations provided by the NCI were taken as initial 3D structures. Hydrogen atoms corresponding to protonation states appropriate to pH 7 were added using a version of Walters and Stahl's Babel program, locally modified by Robert Jorissen. The mean ± standard deviations of the molecular weight of the decoys are 304 ± 113 Da, and the corresponding figures for their net charges are 0.09 ± 1.11.

Docking Calculations. The program Vdock[17,18] was used to dock ligands and decoy compounds to the crystal structures and homology models of the targeted proteins. For each ligand−protein pair, 20 docked conformations were generated with 12 000 and 3000 trial conformations during the hunt and fine-tune phases, respectively. On the basis of test calculations (not shown), ligands were scored by their computed interaction with the receptor, neglecting the internal energy of the ligand and any consideration of the free ligand. As done previously,[18] the distance-dependent dielectric model with a coefficient of 4 was used to scale Coulombic interactions. In docking calculations where the protein was treated as rigid, serial docking[10] was used to accelerate calculations for the target structure and the homology models. However, serial docking was not used when side chains were treated as mobile during docking or for docking into the template structures.

The docking algorithm requires that the user specify the center and dimensions of a rectangular box (the "translational box") which defines the region over which the ligand will range during the search; a central atom of the ligand is confined to this box, while the rest of the ligand can extend outside it. Since our aim is to model a situation in which the structure of the target is not known, the translational boxes were centered on the ligand present in the template with the highest sequence identity. If there was no ligand in any of the templates, the center was chosen by hand. The precise position of the center of box is not critical because the size of the box was a rather generous 10 × 10 × 10 Å.

Table 6. Effect of Mobilizing Side Chains on Enrichment by Homology Models of Carboxypeptidase A[a]

template	PDB code	mobile side chains	⟨rank⟩ rigid	flexible	repacked
procarboxypeptidase A2	1aye	R145, Y248	17.3	13.9	16.5
procarboxypeptidase B	1kwm	H69, R127, R145	15.6	15.6	18.5
carboxypeptidase T	1obr	R145	8.5	10.6	7.0
procarboxypeptidase A	1jqg	R145, Y198, Y248	12.9	14.8	18.4

[a] template: Protein used as a template for the construction of the homology model. PDB code: Protein Data Bank ID of the template structure. rigid: Average rank of known binders when the structure is held rigid. flexible: Average rank when the specified side chains are treated as mobile during docking. repacked: Average rank of known binders when specified side chains are repacked in advance of docking.

Table 7. Effect of Mobilizing Side Chains on Enrichment by Homology Models of Factor Xa[a]

template	PDB code	mobile side chains	⟨rank⟩ rigid	flexible	repacked
factor IXa	1rfn	Y99, Q192	3.1	8.5	2.9
α-thrombin	1doj	T198	1.9	2.2	6.3
factor VIIa	1kli	Y99, F174	6.0	8.6	4.7
protein C	1aut		2.8	n/a	n/a

[a] See Table 6 for details. n/a = not applicable.

Table 8. Effect of Mobilizing Side Chains on Enrichment by Homology Models of PPARα[a]

template	PDB code	mobile side chains	⟨rank⟩ rigid	flexible	repacked
PPARγ	2prg		9.5	n/a	n/a
RORα	1n83	Y334	23.6	10.6	46.5
LXRβ	1pq9		10.4	n/a	n/a

[a] See Table 6 for details. n/a = not applicable.

Table 9. Effect of Mobilizing Side Chains on Enrichment by Homology Models of CDK2[a]

template	PDB code	mobile side chains	⟨rank⟩ rigid	flexible	repacked
CDK6	1blx	K88, N132, D145	39	34	38
ERK2	3erk	F80, K89, N132	30	33	34
PKA	1q8u	K88, N132	35	27	30
c-Abl	1iep		43	n/a	n/a

[a] See Table 6 for details. n/a = not applicable.

We wished to determine whether mobilizing side chains before or during docking could improve the results of docking into homology models. Mobilizing a large number of side chains during docking hinders convergence, so it is preferable to select one or a few key side chains to be treated as mobile. However, it was not clear a priori how to identify which side chains are most important to mobilize. Here, all homology models of the target were superimposed and inspected. For each model, any binding-site side chain deemed to differ significantly from that present in the majority of the models was treated as mobile during docking. This approach typically led to mobilization of one to three side chains in a given model (see Tables 6−10). In cases where the sequence identity was so low that the homology model in question had numerous obvious errors, only those side chains that projected directly into the binding site, and thus could clash with ligands, were mobilized. Once a set of side chains to be mobilized had been identified, it was not modified on the basis of the results of subsequent docking experiments. Two methods of mobilizing side chains were tried. In the first (termed "flexible"), side-chain dihedrals were treated as rotatable during the docking process, as previously described.[18] In the second (termed "repacked"), the mobile side chains were repacked by Vdock in advance of the docking calculations, by running Vdock with no ligand but with the side-chain dihedrals treated as flexible. The resulting structures were then held fixed while the known binders and decoy compounds were docked.

Table 10. Effect of Mobilizing Side Chains on Enrichment by Homology Models of AChE[a]

template	PDB code	mobile side chains	⟨rank⟩ rigid	flexible	repacked
BuChE	1p0m	W84	19.0	16.9	17.8
carboxylesterase 1	1mx1	W84, F331	40.1	34.4	38.4
triacylglycerol lipase	1lps	Y121, F331	48.0	38.5	34.3
cholesterol esterase	1cle		27.2	n/a	n/a
para-nitrobenzyl esterase	1qe3		25.4	n/a	n/a

[a] See Table 6 for details. n/a = not applicable.

Evaluation of Virtual Screening Results. The docking results are evaluated by ranking all the docked compounds according to their best docked energies and determining the degree to which the known binders are concentrated near the top of the ranked list. Such results are frequently summarized as an enrichment factor EF(X), defined as the ratio of the number of known ligands lying within the top X% of the database to the number of known ligands expected in the top X% if the order of the compounds were randomized. However, such enrichment factors depend on an arbitrarily selected value of X and do not depend on the quality of the ranking results for the ligands falling within the top X% and the bottom 100 − X%. For example, EF(10) does not differentiate between rankings of 1, 2, and 50 versus 48, 49, and 50 in a ranked list of 1000 compounds. The average rank of the binders[10] is sensitive to the rankings of all compounds and, thus, arguably represents a more informative and less arbitrary measure of enrichment. The average rank, as a percentage, is calculated as

$$\langle r \rangle = \frac{100\%}{N_{\text{bind}} N_{\text{tot}}} \sum_{i=1}^{N_{\text{bind}}} r_i$$

where r_i is the rank of ligand i and N_{bind} and N_{tot} are the number of known ligands and the number of known ligands and decoys, respectively. Note that a random selection of compounds from the mixture of binders and decoys should yield an average rank of 50%. Symmetry-corrected geometric RMSD values for docked ligand poses were computed after global superposition of the receptor model with the corresponding crystal structure.

RESULTS

This section begins by benchmarking the screening capabilities of Vdock using crystal structures of protein targets and comparing them with published results for other docking programs. We then study what happens to screening results when crystal structures are replaced by homology

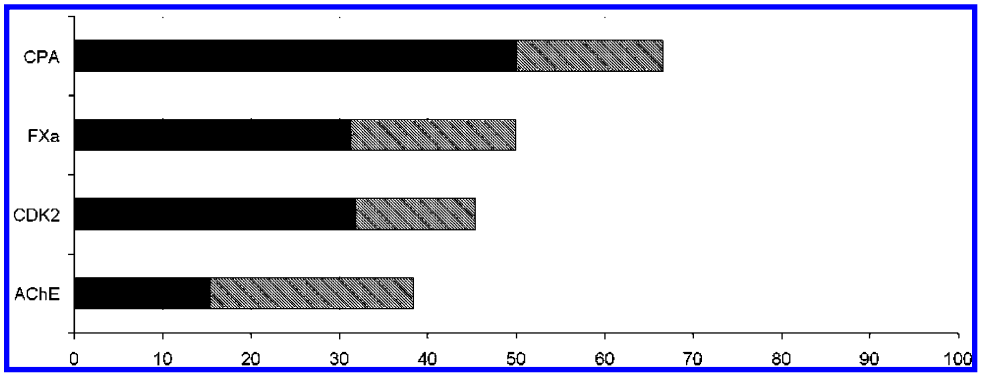

Figure 1. Accuracy of docking to crystal structures of receptors. Bars show the percentage of ligand−target complexes with available crystal structures for which Vdock's top-scoring pose has a symmetry-corrected RMSD < 2 Å (black) and < 4 Å (black + gray).

models built from templates of varied sequence identities to the target or by the templates themselves. The consequences of mobilizing selected side chains in the binding sites are then examined. A final section provides a more detailed analysis of results for the various protein targets.

Baseline Performance of Vdock in Virtual Screening. It is of interest to compare the performance of Vdock for the target proteins studied here with the performance of other current docking algorithms because similar performance would tend to support the generality of the results for homology models. Suitable prior results were found for FXa, AChE, and CDK2, though not for PPARα or CPA. Baxter et al. used the program PRO_LEADS[19] and a crystal structure of Factor Xa to screen a database of decoys mixed with known FXa binders. They found that the top-ranked 5% and 10% of the compounds were enriched by factors of 8.6 and 14.7, respectively. Vdock yields similar results for this definition of the enrichment factor, 8.1 and 15.0, respectively. Acetylcholinesterase is a complex target to study because it possesses many crystallographic water molecules and a large binding site with distinct subsites.[20] McGovern and Shoichet, using a version of DOCK3.5 (see, e.g., ref 21), found 25% of the known binders in the top-ranked 6.3% of a database of binders and decoys;[6] here, the top 25% were found in the top-ranked 2.2%, indicating a higher degree of enrichment. Halgren et al. used the program Glide[22] to screen for ligands of CDK2. Their results depend somewhat upon which crystal structure was used and more strongly upon the scoring function. For Glide Score 1.8, their modified enrichment factor EF′(70%) ranged from 2.1 to 3.6, depending upon the target structure; for Glide Score 2.0, the range was 3.8−3.9, and for Glide Score 2.5, the range was 5.4−6.8. For comparison, Vdock yields an EF′(70%) of 3.6, which is in the middle of the Glide Score results. Vdock also provides good enrichment for these modified enrichment factors for PPARα [EF′(2%) = 27 and EF′(70%) = 14] and for CPA [EF′(2%) = 27 and EF′(70%) = 16]. Like Glide, Vdock requires roughly 2−3 min on a current commodity CPU to dock each compound.

It can be shown that the enrichments reported here are not a trivial consequence of differences between the molecular weight distributions of the known binders and the decoys. This is done by ranking the compounds according to their molecular weight as if it were a scoring function, determining the average rank of the known binders, and comparing the results with those provided by Vdock. If the Vdock results were merely reporting on molecular weight, then scoring compounds by molecular weight would yield equally high enrichment. Although molecular weight does yield marked enrichment in some cases, the average ranks are still 1.5- to 6-fold worse than the Vdock scores: 59% versus 10% for CPA, 7.6% versus 2.6% for FXa, 26% versus 12% for PPARα, 32% versus 22% for CDK2, and 54% versus 17% for AChE. These figures are based upon an assumption that high molecular weight constitutes a better score. Reversing the ranking to favor low molecular weight replaces each average range $\langle r \rangle$ by $100 - \langle r \rangle$, yielding slightly improved results for CPA (41% instead of 59%) and AChE (46% instead of 54%) but much worse results for the other systems.

Finally, Figure 1 examines the geometric accuracy of the top-scored ligand poses for cases in which a crystal structure of the complex is available, plotting the percentage of top-scoring ligand poses with heavy atom RMSDs < 2 Å and < 4 Å. PPARα is omitted, however, because our data set includes only two crystal ligand poses and, therefore, cannot provide meaningful statistics. The percentages range between 15% and 50% for <2 Å and between 38% and 67% for <4 Å. These values are roughly commensurate with the "First Pose" docking results reported for some of the more successful docking methods in a recent comparative study of other programs (see Figure 2 of ref 23). However, an exact comparison of methods is not possible because of differences between the systems studied and the ligand sets employed.

Virtual Screening with Homology Models Based upon Varied Templates. Figures 2−6 show typical enrichment plots for docking known ligands and decoy compounds into a crystal structure of each target and into the various homology models, and Tables 1−5 provide further details regarding the structures and enrichment statistics. The first row of each table refers to the crystal structure of the target, and the subsequent rows refer to the homology models, which are listed in order of decreasing sequence identity of the template to the target. The results show considerable variation in the overall enrichments (average rank of known binders) from one target protein to another: FXa and its models yield excellent average ranks below 10%; CPA and PPARα and their models yield intermediate results, with average ranks in roughly the 10−20% range; and CDK2 and AChE yield the worst enrichment, with average rankings ranking from about 20% up to nearly 50%, depending upon the target structure used for the screening. In general, the crystal structures of the targets yield some of the best enrichment results. However, for each system, there is at

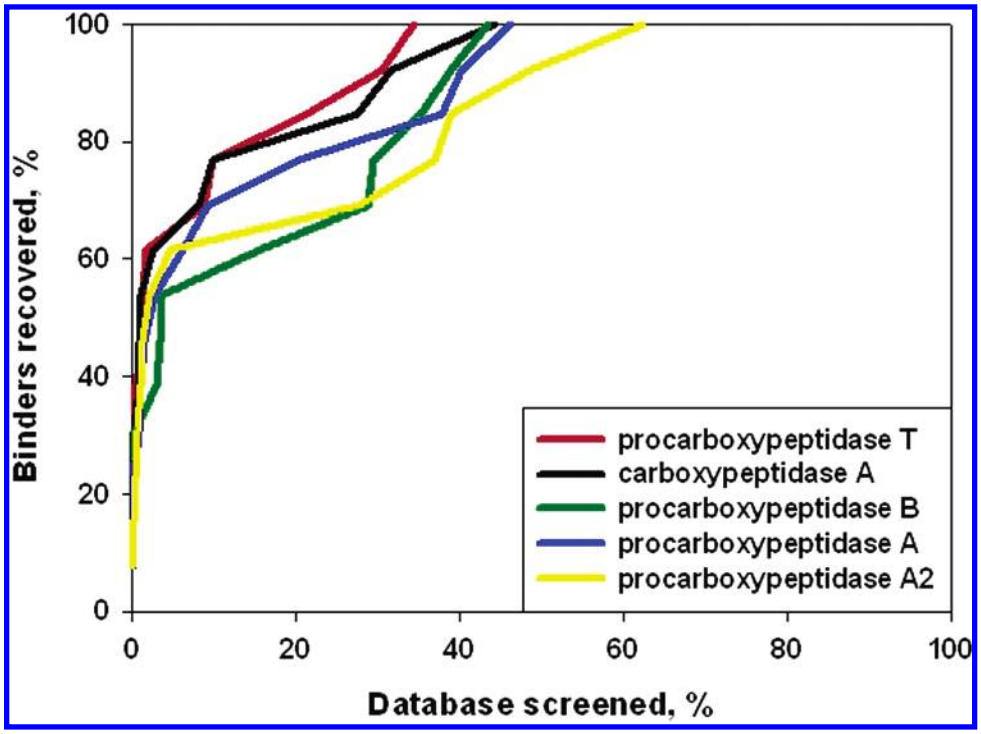

Figure 2. Recovery of carboxypeptidase A binders from decoy compounds by docking to the target structure and homology models based upon templates specified in the legend.

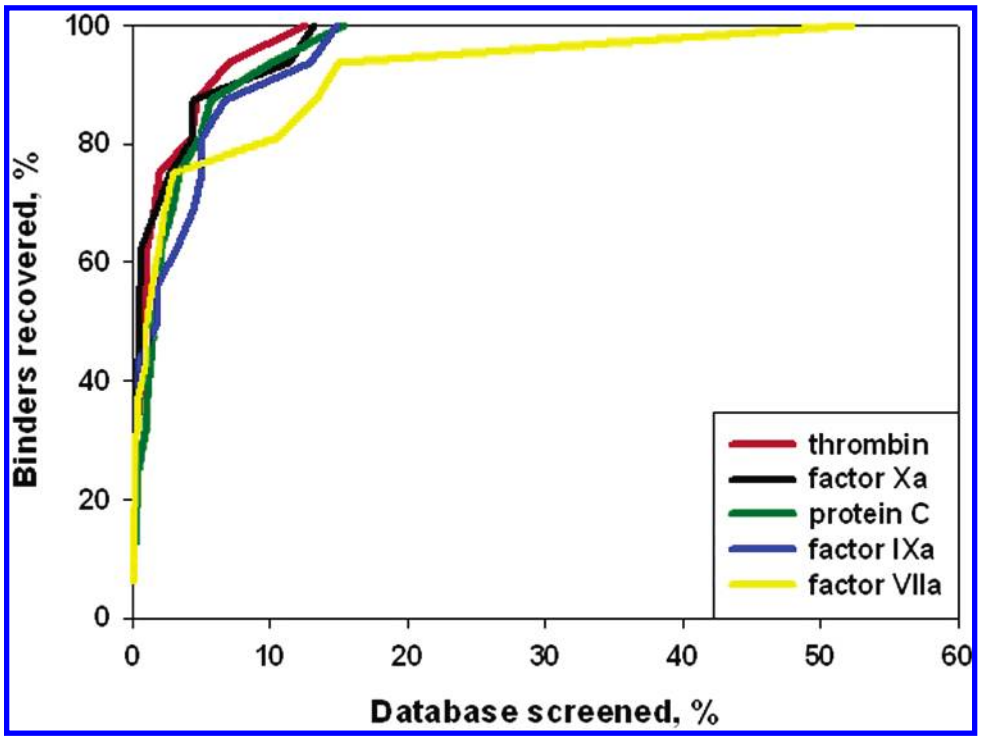

Figure 3. Recovery of factor Xa binders from decoy compounds by docking to the target structure and homology models based upon templates specified in the legend.

least one homology model which yields enrichment better than, or nearly as good as, that obtained from the crystal structure. Moreover, the homology model results are rather similar to those obtained from the target crystal structures for three of the systems, CPA, PPARα, and FXa. The homology models tend to perform considerably worse for CDK2 and AChE. Interestingly, these are also the systems for which the target structures themselves give the worst enrichment.

Figure 7a−f graphically represents the relationships between enrichment and various measures of the similarity of the template protein or the homology model itself to the actual target. The correlations observed between enrichment and the various measures of sequence similarity are remarkably inconsistent. (The figures show data points for the target crystal structures for reference, but these should not be considered when assessing correlations because they are not based upon homology models.) The most noticeable cor-

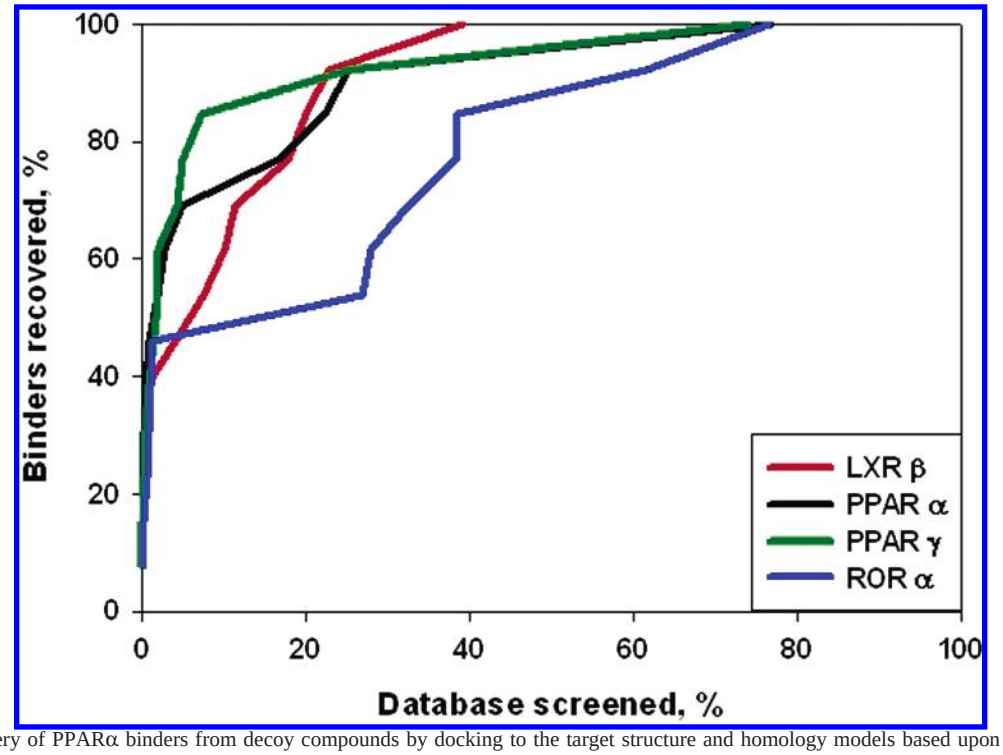

Figure 4. Recovery of PPARα binders from decoy compounds by docking to the target structure and homology models based upon templates specified in the legend.

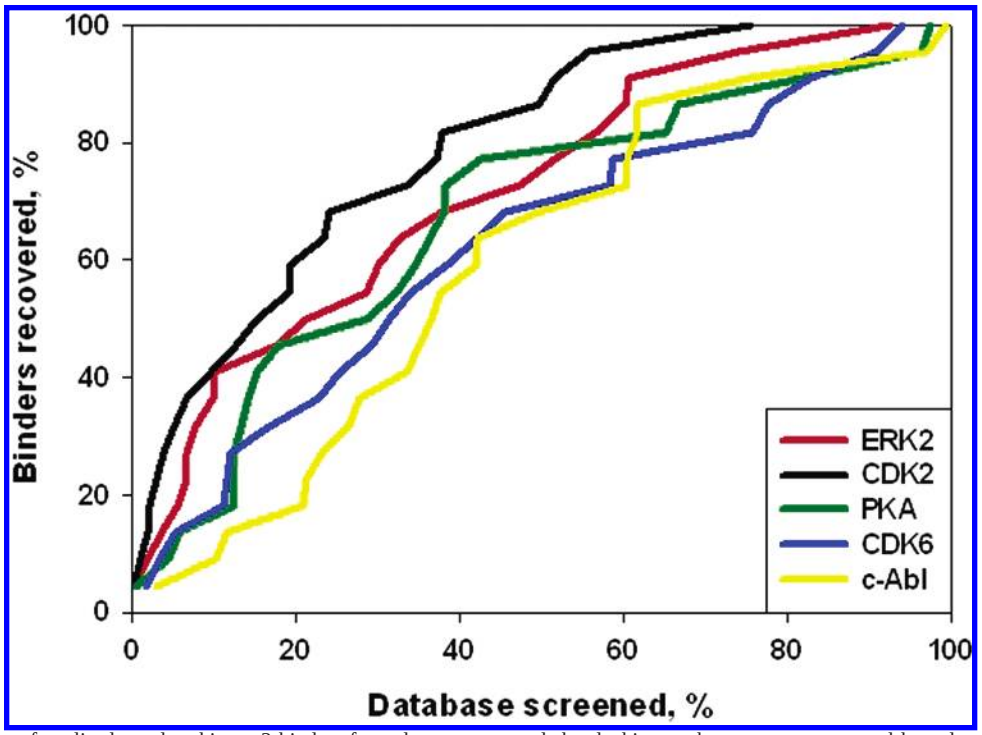

Figure 5. Recovery of cyclin dependent kinase 2 binders from decoy compounds by docking to the target structure and homology models based upon templates specified in the legend.

relations are observed for the two systems with the lowest overall enrichments, CDK2 and AChE, which both display some correlation with overall (Figure 7a) and binding-site sequence identity (Figure 7b). There is also some correlation between enrichment and the overall RMSD of the homology model from the target crystal structure, for each system and also when all the data are viewed together (Figure 7e). Thus, the average rank of the binders is typically below 20% when the RMSD is less than 1 Å, and it is always greater than 20% when the overall RMSD is greater than 2.5 Å. One can also discern a correlation between average rank and the RMSD of the active site of the homology model, especially if the AChE system is omitted from consideration. The lack of a stronger correlation between active-site RMSD and enrichment may have to do with the fact that a given RMSD can be achieved by small shifts of multiple atoms, or by a large shift of a small number of atoms, which is more likely to damage the ability of the structure to enrich for known ligands.

It is of interest to inquire whether templates solved with a bound ligand yield more useful homology models than those solved without any bound ligand, although the present

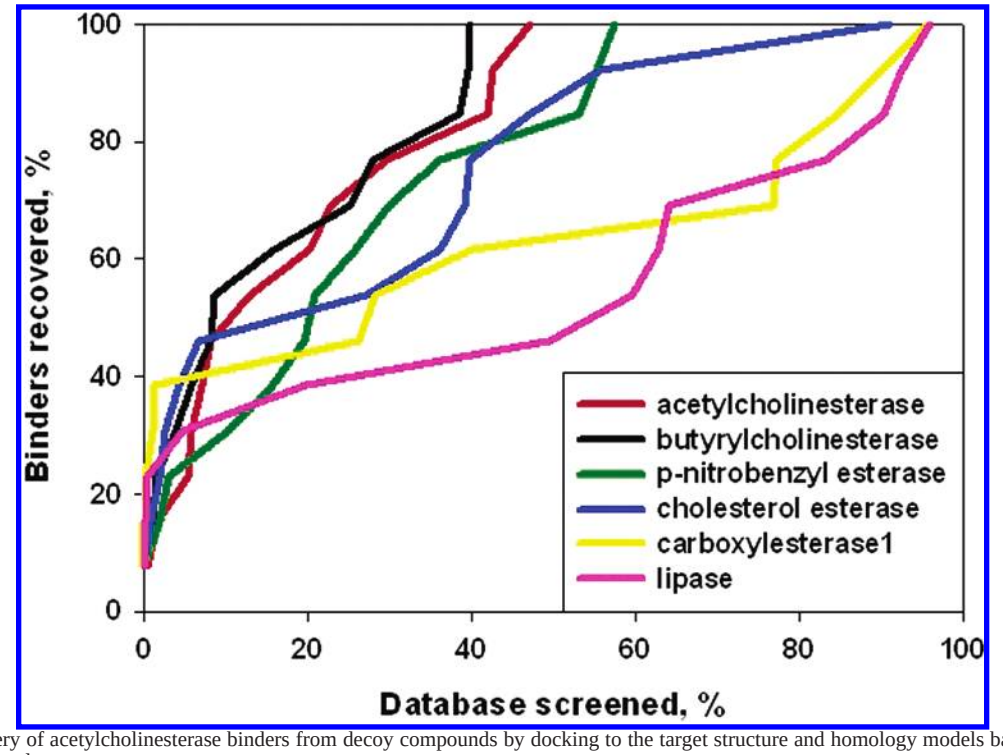

Figure 6. Recovery of acetylcholinesterase binders from decoy compounds by docking to the target structure and homology models based upon templates specified in the legend.

study is not designed to systematically address this question. In Tables 1−5, crystal structures (both template and target) solved *without* a bound ligand are indicated by a PDB code in italics. There is no clear trend favoring structures solved with or without a bound ligand. For example, the structure of para-nitrobenzylesterase was solved with no bound ligand, but the homology model of AChE built from it yields a better average rank (25%) than the model built from triacylglycerol lipase, which was solved with a bound ligand. On the other hand, the models built from factor IXa and factor VIIa, which were solved with a very small bound ligand, do underperform the other models of factor Xa, which were built from templates solved with larger ligands.

Accuracy of Docked Geometries in Homology Models. Figure 8 examines the geometric accuracy of the top-scored ligand poses for docking into homology models, for cases in which a crystal structure of the complex is available, by plotting the percentage of top-scoring ligand poses with heavy-atom RMSDs < 2 Å and < 4 Å. These may be compared with the results for docking into crystal structures, shown in Figure 1. Note that the RMSDs for docking into homology models can be artificially elevated by geometric shifts related to the superposition of the homology models on the crystal structures of the targets. Therefore, Figure 9 reports the same statistics but with looser tolerances, that is, heavy-atom RMSDs < 3 Å and < 5 Å. A comparison of Figure 8 with Figure 1 shows that some models of carboxypeptidase A perform about as well as the crystal structure. However, the homology models yield substantially less-accurate poses for the other systems.

Interestingly, even a homology model that yields good enrichment of known binders—that is, high average ranks—can provide poor structural RMSD values. This result was studied by inspecting the docked geometries of active compounds with poor RMSDs but high rank on the basis of energy. In some cases, the docking calculations recognized a key chemical feature of the actives but failed to generate a highly accurate geometry. For example, factor Xa models can recognize true actives as elongated compounds usually possessing a terminal cation, and acetylcholinesterase's closed active site imposes strong shape restrictions which are better satisfied by the actives than the inactives.

Docking into Templates Instead of Homology Models. It is expected a priori that docking into a homology model, rather than the template used to construct the model, will give greater enrichment. Tables 1−5 (final column) address this by showing the average ranks of the binders obtained by docking into the template proteins themselves, and Figure 10 compares these results with those obtained from the homology models. A clear correlation is evident between the enrichments obtained from the homology models and the corresponding templates, and overall, the templates perform about as well as the models. Indeed, the templates actually provide greater enrichment than the corresponding homology models for some systems. It is, perhaps, significant that comparatively weak enrichment (high average ranks; points in top-right of graph) is obtained in most of these cases. However, for systems where high enrichment is obtained, the homology models tend to provide better results than the templates (low average ranks; points in lower-left of graph).

Modeling the Flexibility of Receptor Side Chains in Homology Models. Figure 11 and Tables 6−10 report the consequences of modeling side chains in the binding site as flexible either before or during docking. Repacking the mobile side chains and then holding them fixed during docking produces mixed results with no clear trend toward improvement (Figure 11, top). However, treating side chains as flexible during docking tended to improve the results, though the effect is generally weak (Figure 11, bottom). The most consistent improvements are observed for the models of AChE and CDK2, while mobilizing side chains in the

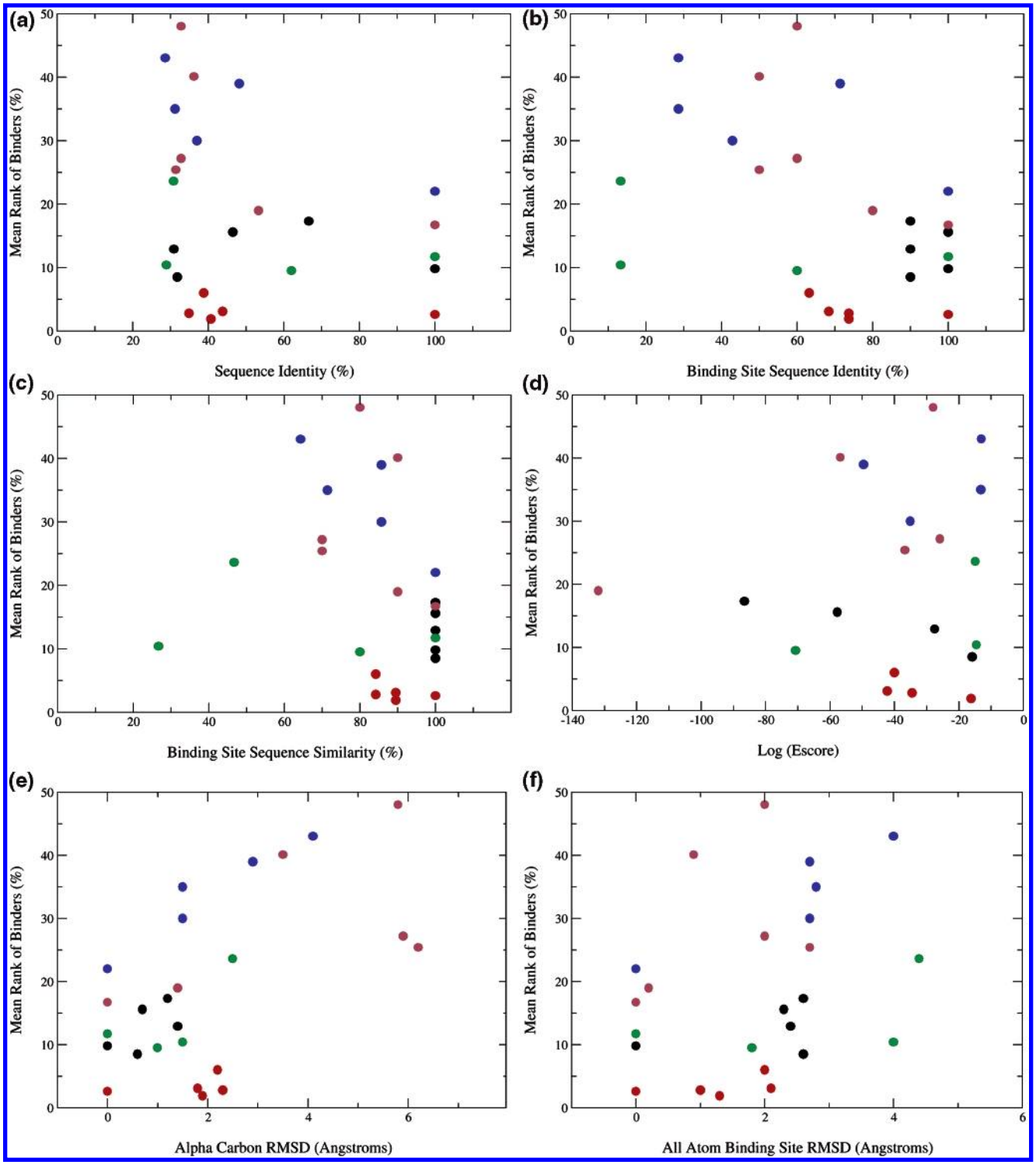

Figure 7. (a) Scatter plot of the mean rank of known binders obtained with homology models versus the template-to-target sequence identity. Results for the crystal structures of the targets are shown at a sequence identity of 100%. For all parts (a−f): black, CPA; red, FXa; green, PPARα; blue, CDK2; brown, AChE. (b) Scatter plot of the mean rank of known binders obtained with homology models versus the template-to-target sequence identity computed only for binding-site residues. Results for the crystal structures of the targets are shown at a sequence identity of 100%. (c) Scatter plot of the mean rank of known binders obtained with homology models versus the template-to-target sequence similarity for binding-site residues. Results for the crystal structures of the targets are shown at a sequence identity of 100%. (d) Scatter plot of the mean rank of known binders obtained with homology models versus the logarithm of the E score of the template sequence from a FASTA search in the Protein Data Bank with the target sequence. Results for the crystal structures of the targets are now shown. (e) Scatter plot of the mean rank of known binders obtained with homology models versus the root-mean-square deviation of the homology model from the target structure (α carbons only). Results for the crystal structures of the targets are shown at RMSD = 0. (f) Scatter plot of the mean rank of known binders obtained with homology models versus the root-mean-square deviation of the homology model from the target structure (all binding-site atoms). Results for the crystal structures of the targets are shown at RMSD = 0.

FXa models tends to worsen enrichment. The largest differences among the three approaches (rigid model, re-packed model, and flexible model) are observed for the model of PPARα based upon retinoid-related orphan receptor

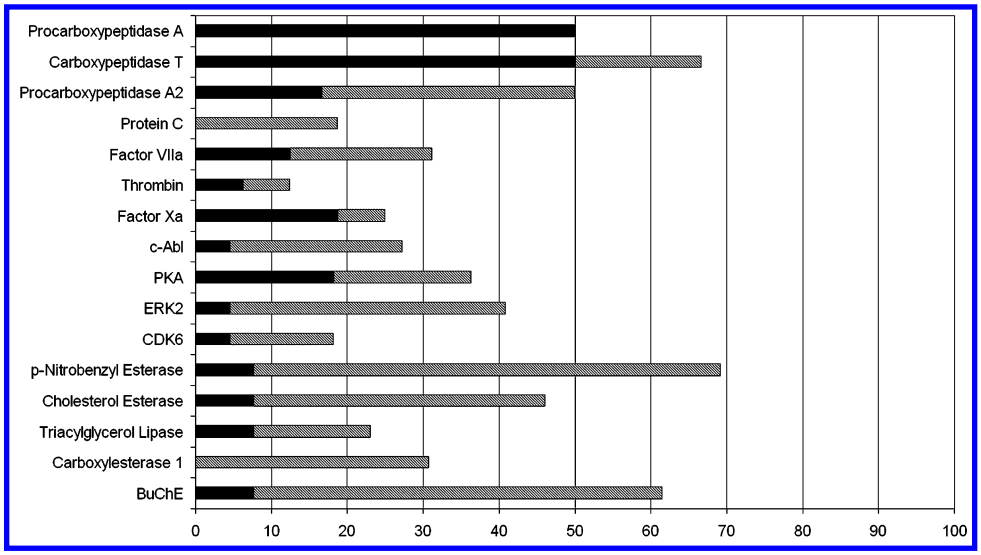

Figure 8. Accuracy of docking to homology models of receptors. Bars show the percentage of ligand−target complexes with available crystal structures for which Vdock's top-scoring pose has a symmetry-corrected RMSD < 2 Å (black) and < 4 Å (black + gray).

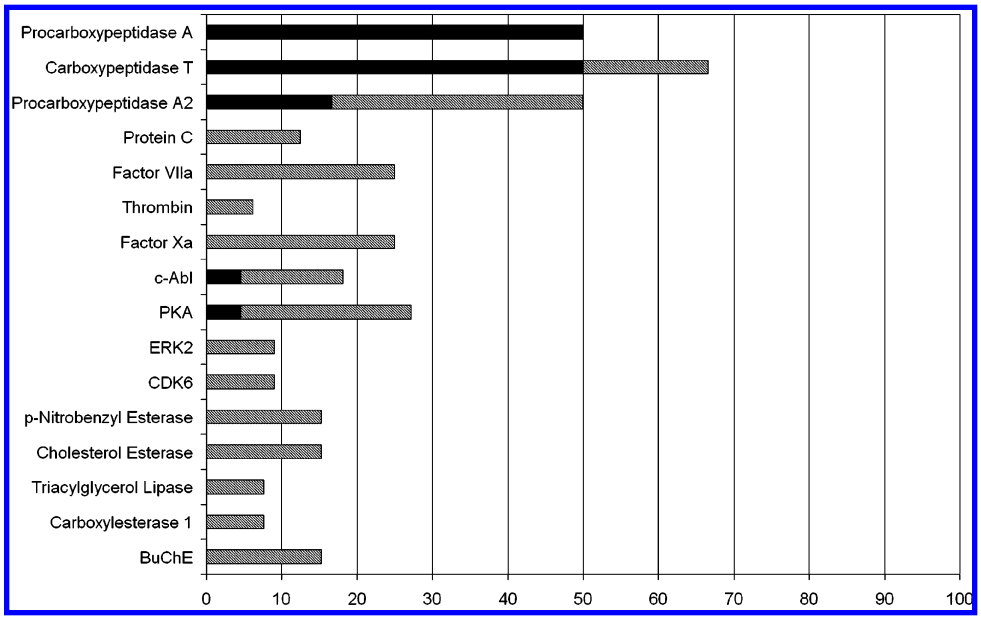

Figure 9. Same as Figure 8, but with criteria RMSD < 3 Å (black) and < 5 Å (black + gray).

α (RORα). Here, Tyr334 partly obstructs the binding site in the baseline model, leading to worse enrichment (23.6%) than for the other two models of PPARα (9.5% and 10.4%). Repacking this side chain in advance of docking proves to worsen its overlap with the binding site, but treating it as mobile during docking causes it to shift and allow the known binders to adopt lower-energy docked conformations.

Analysis of Individual Targets. *Carboxypeptidase A.* CPA, a zinc protease, is a digestive enzyme that hydrolyzes the C-terminal peptide bond of proteins and polypeptides. The crystal structure of CPA has been used as a model for the design of medically relevant zinc protease inhibitors, including angiotensin-converting enzyme inhibitors for the treatment of hypertension (see ref 24 and citations therein). Overall, the crystal structure and homology models of CPA yield substantial enrichment, with average ranks of the binders ranging between 8.5% and 17.3%. Interestingly, the best result is obtained from a homology model based upon carboxypeptidase T, whose sequence identity with CPA is only 32%. However, the crystal structure of the target is a close second with an average rank of 9.8%. It is worth noting that higher rankings might have been obtained across all of the models if the thiols present in some of the ligands had been treated as ionized, because the force field does not effectively model the thiol−zinc dative bond.

It appears that the performance of the various structures is determined in large part by the position of the guanidinium group of Arg145, because of its electrostatic interaction with the anionic inhibitors. As illustrated in Figure 12, the guanidinium moiety is placed most favorably in the carboxypeptidase T model, closely followed by the crystal structure of CPA itself. Mobilizing the side chain of the Arg145 in the other homology models does not improve the screening results, apparently because Arg145 tends to associate with the nearby side chain of Asp142, and thus does not become available to interact with the inhibitors.

Factor Xa. The serine protease FXa links the extrinsic and intrinsic pathways of blood coagulation, upon activation, converting prothrombin to thrombin, and is a target for the development of new anticoagulant medications.[25] Overall,

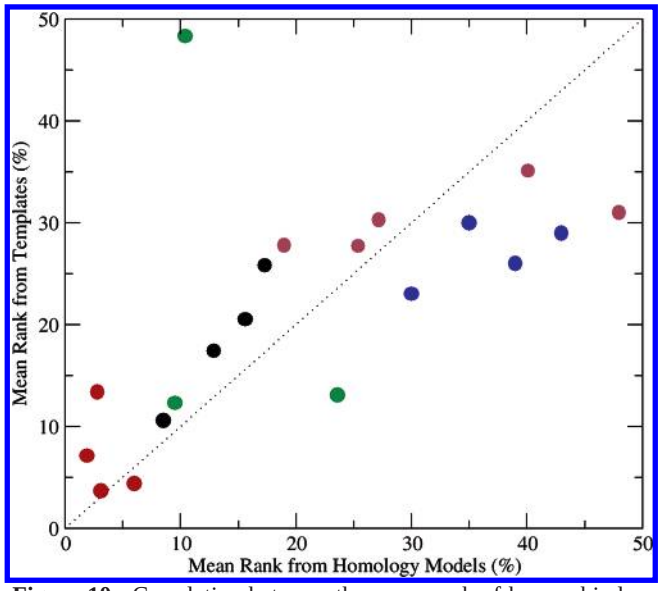

Figure 10. Correlation between the mean rank of known binders from docking to homology models versus that from docking to the corresponding template structures. Identical results would lie on the diagonal line. Black, CPA; red, FXa; green, PPARα; blue, CDK2; brown, AChE.

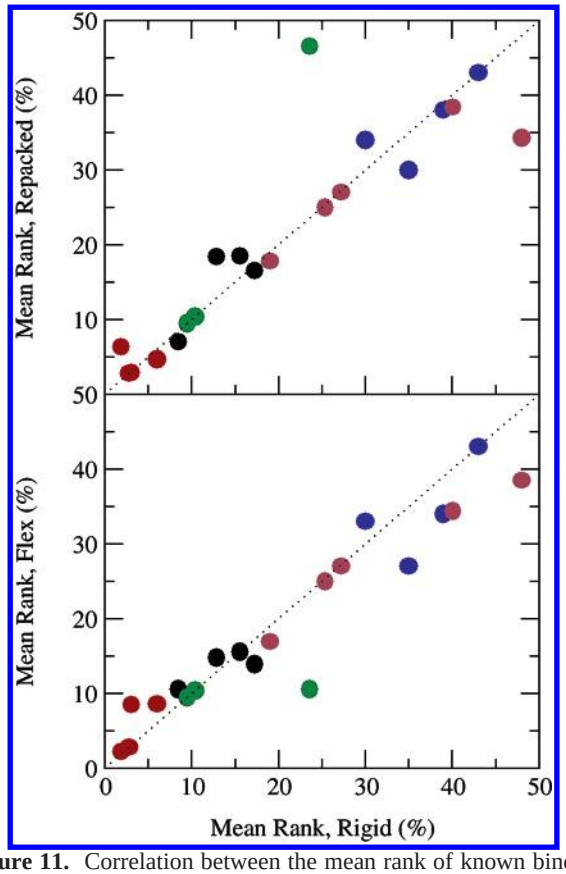

Figure 11. Correlation between the mean rank of known binders obtained by docking into a rigid receptor structure (horizontal axis) and that obtained by a receptor structure with key side chains treated as flexible (vertical axis). Top: side chains repacked before docking. Bottom: side chains mobile during docking. Dotted lines indicate identity between the flexible side-chain and rigid results. Black, CPA; red, FXa; green, PPARα; blue, CDK2; brown, AChE.

excellent enrichment is obtained by virtual screening with either the crystal structure or any of four homology models of this target; the average ranks of the known binders range between 1.9% and 6.0%, and only the model based upon

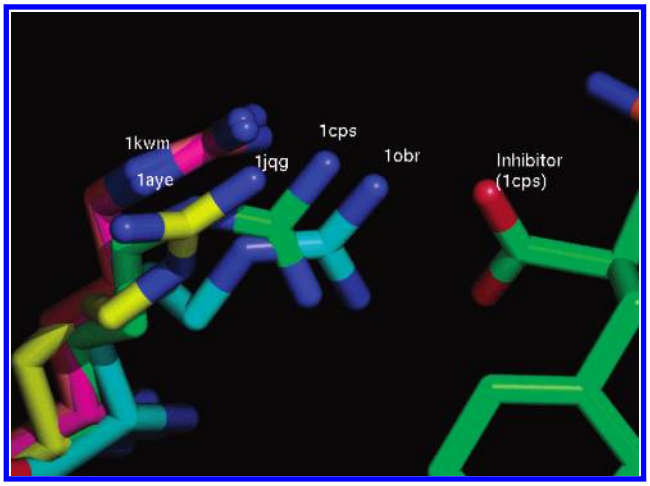

Figure 12. Positions of the conserved Arg145 for homology models and the crystal structure of carboxypeptidase A, along with the carboxylate group of the bound inhibitor in its crystal conformation.

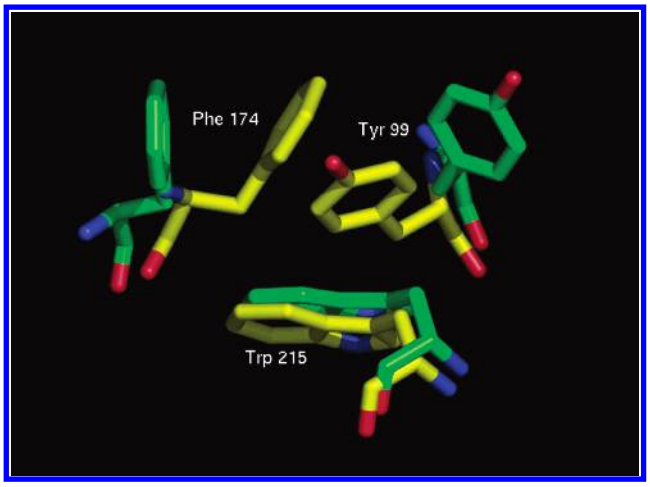

Figure 13. Position of residues Trp215, Phe174, and Tyr99 in the S4 subsite of the factor Xa target (green) and the homology model based on factor VIIa (yellow).

factor VIIa yields a mean rank greater than 3.1%. The favorable overall results are traceable, in part, to strongly favorable electrostatic interactions, according to the docking energy model, between the aspartate group in the S1 pocket of the target and the amidinium or guanidinium moieties of 11 out of 16 of the known binders. Thus, the average rank of these cationic compounds is 0.81% when docking to the target crystal structure, whereas the average rank is 6.4% for the five known binders that do not contact the S1 aspartyl with a cationic group. A more sophisticated solvation model might mitigate this problem.

As previously noted,[10] the comparatively poor enrichment from the model based upon factor VIIa seems traceable to the positions of the side chains of Phe174 and Tyr99, which effectively close the hydrophobic S4 subsite, as illustrated in Figure 13. (Conventional residue numbering is followed here, rather than the numbering found in 1g2l.) For certain homology models, good enrichment appears to be obtained somewhat fortuitously. For example, some known binders docked to the thrombin-based model of FXa formed stabilizing interactions with Arg150 or Arg222 in place of the Phe174 and Tyr99 in the S4 pocket.

Mobilizing side chains in the homology models of FXa tended to worsen enrichment. Even treating the misplaced Phe174 and Tyr99 side chains of the FVIIa model as mobile during docking led to worse enrichment. Interestingly, the ligands tended to displace the mobile side chains and, thus, take up markedly incorrect positions. This effect can be reproduced by mobilizing Tyr99 in the crystal structure of FXa, which generates a 2.3-fold worse enrichment of the known binders. On the other hand, repacking the side chains of the FVIIa-based model in advance of docking improved the enrichment somewhat, from an average rank of 6.0% to 4.7%. The reason, presumably, is that this procedure does not allow the ligands to compete with side chains for positions on the surface of the protein.

Peroxisome Proliferator-Activated Receptor α. PPARα is a member of the family of nuclear receptors and is the target of the fibrate class of drugs for lowering blood triglyceride and cholesterol levels.[26] It also appears to play a role in the regulation of hunger and food intake through association with the endogenous agonist oleylethanolamide.[27] As shown in Table 3, the average rank of the known binders is about 10% for the crystal structure of the target and two of the homology models and is 23.6% for the third homology model, which is based upon RORα. The top-ranked 20 compounds include six of the known binders when the crystal structure of the target is used for screening and four to five of the known binders when the homology models are used. This is despite significant variations in the binding modes for the docked structures: the binding site is roughly T-shaped,[28] with the stem of the T opening to the solvent, and the elongated ligands can dock into either of the two distal cavities.

The relatively poor enrichment obtained with the model from RORα proves to result from a blockade of the binding site by the misplaced side chain of Tyr334, and mobilization of this side chain during docking improves the average rank of the binders more than 2-fold, from 23.6% to 10.6%, similar to that obtained for the other structures. On the other hand, repacking Tyr334 in advance of docking worsens the average rank to 46.5% (see Table 8).

Cyclin-Dependent Kinase 2. Cyclin-dependent kinase 2 is part of the cellular mechanism controlling progression through the cell cycle and has been identified as a drug target for the treatment of cancer.[29] Overall, the enrichment of known actives from both the target structure and the homology models of CDK2 is only moderate, with average rankings of known binders ranging from 22% to 43%. Not surprisingly, the best results are obtained with the target crystal structure. The worst results are obtained from the homology model based upon c-Abl, which has multiple backbone errors and side-chain clashes. The next worst results are obtained from the model based upon CDK6. The problems here appear to result largely from a backbone twist relative to the CDK2 structure, which in turn appear attributable to specific interactions between the CDK6 template and the tumor suppressor protein $p19^{INK4d}$ bound to the CDK6 template.[30] The backbone twist rotates the carbonyl of Leu83 into the binding site, making it more difficult to form the correct hydrogen bonds with the known binders, and also rotates Phe82 deeper into the binding pocket, reducing its volume (Figure 14). The PKA-based homology model resembles the target except for a misalignment of residues 85−89, which causes the side chain of

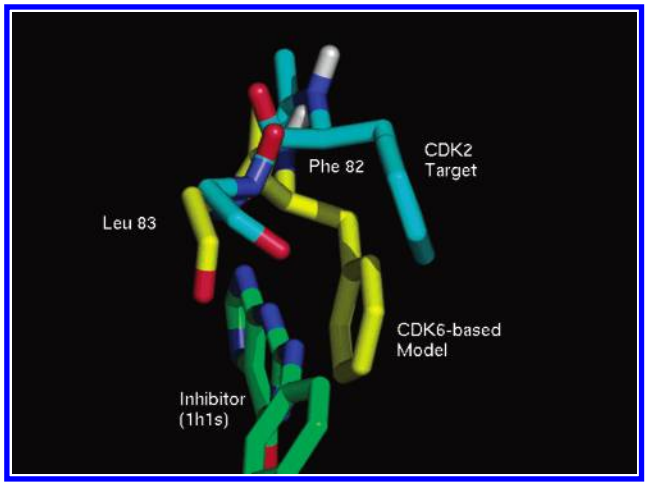

Figure 14. Residue Phe82 and the backbone of Leu83 in the target protein CDK2 and in the CDK6-based model.

Lys88 to enter the binding site. Mobilizing this side chain, along with Asn132, thus generates a significant improvement in enrichment, as shown in Table 9. Finally, in the model based upon ERK2, the side chains of Phe80 and Lys89 partly block the binding site and the important main-chain atoms of Glu81 and Leu83 are misplaced. However, the L2 loop is much more open compared to the target, so there is still room in the binding site for the larger CDK2 inhibitors.

Acetylcholinesterase. Acetylcholinesterase cleaves the neurotransmitter acetylcholine in the neuromuscular junction. Some inhibitors of this critical enzyme are potent neurotoxins used as insecticides and are incorporated into chemical warfare agents, while others are used as medications for the treatment of myasthenia gravis, glaucoma, and Alzheimer's disease. AChE was expected to be a challenging target for Vdock because some inhibitors appear to gain affinity from interactions that may not be fully accounted for by the CHARMM force field used here, notably $\pi-\pi$ and cation$-\pi$ interactions[31] and, for the poly(ethylene glycol)-based inhibitor in 1jjb, several C−H···O and C−H···π interactions. In addition, ligand binding is often mediated by several water molecules.[32] As shown in Table 5, the range of enrichment varies widely across the various structures, with average ranks ranging from 16.7% for the crystal structure of the target to 48% for the model based upon triacylglycerol lipase.

The homology model based upon butyrylcholinesterase yields results only slightly worse (mean rank 19%) than those from the crystal structure of the target (16.7%). The structure of the binding site is very similar to that of the target (0.2 Å RMSD); the chief difference is a rotation of the Phe330 side chain at the bottom of the gorge, which weakens the interactions of the known inhibitors. (Note that the binding-site RMSD does not include atoms in the peripheral anionic site.) Interestingly, Phe330 adopts different conformations in different crystal structures of AChE itself.[31,32] The next-best results are obtained with models based upon cholesterol esterase and para-nitrobenzyl esterase (mean ranks of about 26%), despite gross backbone errors which move the key residue Trp84 and its neighbors by over 10 Å relative to the correct structure. These errors are so large that the enrichments obtained with these models appear to be at least partly fortuitous. Large backbone shifts are also seen in the model based on triacylglycerol lipase, but the results are worse because the binding site is partly blocked by the side chains

of Tyr121 and Phe331 and the backbone of Gly117, Gly118, and Gly119. Mobilizing Tyr121 and Phe331, thus, improves the results somewhat (Table 10). Similarly, mobilizing the obstructing Trp84 and Phe331 for the carboxylesterase 1 based model improves the results by 6%.

DISCUSSION AND CONCLUSIONS

Docking calculations with homology models of a targeted protein can yield a substantial enrichment of known actives against a background of decoy compounds. Indeed, some homology models yield enrichment equal to or even greater than that obtained with the crystal structure of the target. Nonetheless, the data indicate that the crystal structure is, in general, a safer choice than a homology model. It is also worth noting that only one crystal structure was tested for each target in this study, whereas multiple homology models were tested. If multiple crystal structures had been tried, the best crystal structure might have significantly outperformed the best homology model.

Unexpectedly, the similarity of the template to the target, measured in various ways, is at best weakly predictive of the enrichment obtained with the resulting homology model. For example, a homology model based on a template with only 30% sequence identity to the target can yield enrichment as high as that obtained from the crystal structure of the target—or significantly lower. It is also unexpected that docking into the template structures themselves, rather than the homology models built from them, yields a similar overall enrichment of known binders. Interestingly, too, there is a good correlation between the enrichment obtained with a template and that obtained with a homology model based upon the template.

The lack of a smooth relationship between similarity and enrichment probably has multiple causes. First, whereas the true structure of the target protein varies when different ligands are bound, the conformation is kept fixed during docking and scoring to the crystal structure. This implies that the crystal structure itself must, to some degree, be incorrect for all of the bound compounds (or all but one, in cases where the target structure was solved with one of the known binders). Thus, similarity to the crystal structure of the target will not necessarily increase enrichment because the crystal structure itself is not exactly right for every ligand. Another likely factor is that the present energy model includes a hard Lennard-Jones term, so small changes in the shape of a binding site can produce large changes in energy. Thus, even if two protein structures are similar by the broad measures of molecular similarity used here, subtle differences in the structure of the binding site may still lead to significantly different compound rankings. Finally, the energy model used here is highly simplified, so enrichment would not be perfect even if the binding-site conformation could be perfectly modeled for each compound.

Two of these factors—dependence of the target structure upon the bound compound and high sensitivity to details of the binding-site conformation—could, in principle, be addressed by allowing for conformational flexibility. On the other hand, allowing greater conformational flexibility poses a greater computational burden and creates more opportunities for an imperfect energy model to generate misleading results. In the present study, a very limited treatment of protein mobility was tested. For homology models only, a side chain was mobilized if its modeled conformation deviated markedly from that in the other homology models of the same target. This procedure does tend to improve enrichment, but not much; the expected benefit of allowing a side chain to adapt to the presence of a ligand appears to be offset by the increased difficulty of identifying the correct global energy minimum. For example, mobilization of a bulky side chain can open a new binding subsite in the protein surface and, thus, allow a decoy ligand to bind better than a known binder. (Although this behavior leads to errors here, it could actually be useful as a basis for discovering new types of ligands for a targeted protein.) Note, too, that main-chain shifts are an important source of error in some homology models, and the present study makes no attempt to repair them.

The same physical and algorithmic considerations are likely to apply to the broad class of similar docking and scoring methods. Indeed, the present results are consistent with those obtained in earlier more limited studies with other methods (see the Introduction). In addition, the baseline performance of Vdock for the systems examined here is similar to that of other current methods, as shown in the Results section. Thus, although only further studies can be definitive, it is likely that the conclusions drawn here will prove to be rather general.

Multiple structural models are often available for a given drug target; options may include homology models, the corresponding protein templates, and experimentally determined structures of the target itself. In light of the present results, it may be difficult to decide which structure to use for virtual screening. In particular, the measures of template—target similarity tested here have not proven to be reliable predictors of how well a given homology model will perform when used for virtual screening. Given a priori information on a set of known ligands, perhaps the best approach is to test the various models with the same sort of docking and scoring exercise as used here and then proceed with the protein structure that provides maximal enrichment.

ACKNOWLEDGMENT

We thank Eugene Melamud and Dr. John Moult for their advice and assistance in constructing homology models, Dr. Robert Jorissen for providing the version of Babel used here, and the anonymous reviewers for their thoughtful comments. This publication was made possible by a subcontract from VeraChem LLC under NIH Grant GM62050 and by NIH Grant GM61300. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

REFERENCES AND NOTES

(1) Sabnis, Y.; Rosenthal, P. J.; Desai, P.; Avery, M. A. Homology modeling of falcipain-2: validation, de novo ligand design and synthesis of novel inhibitors. *J. Biomol. Struct. Dyn.* **2002**, *19*, 765—774.

(2) Takami, A.; Iwakubo, M.; Okada, Y.; Kawata, T.; Odai, H.; Takahashi, N.; Shindo, K.; Kimura, K.; Tagami, Y.; Miyake, M.; Fukushima, K.; Inagaki, M.; Amano, M.; Kaibuchi, K.; Iijima, H. Design and synthesis of Rho kinase inhibitors (I). *Bioorg. Med. Chem.* **2004**, *12*, 2115—2137.

(3) Schapira, M.; Raaka, B. M.; Das, S.; Fan, L.; Totrov, M.; Zhou, Z.; Wilson, S. R.; Abagyan, R.; Samuels, H. H. Discovery of diverse

thyroid hormone receptor antagonists by high-throughput docking. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7354−7359.
(4) Diller, D. J.; Li, R. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **2003**, *46*, 4638−4647.
(5) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **2003**, *50*, 5−25.
(6) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.
(7) Moult, J.; Fidelis, K.; Zemla, A.; Hubbard, T. Critical assessment of methods of protein structure prediction (CASP)-round V. *Proteins* **2003**, *53*, 334−339.
(8) DeWeese-Scott, C.; Moult, J. Molecular modeling of protein function regions. *Proteins* **2004**, *55*, 942−961.
(9) Oshiro, C.; Bradley, E. K.; Eksterowicz, J.; Evensen, E.; Lamb, M. L.; Lanctot, J. K.; Putta, S.; Stanton, R.; Grootenhuis, P. D. J. Performance of 3D-database molecular docking studies into homology models. *J. Med. Chem.* **2004**, *47*, 764−767.
(10) Fernandes, M. X.; Kairys, V.; Gilson, M. K. Comparing ligand interactions with multiple receptors via serial docking. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1961−1970.
(11) Bernstein, F. C.; Koetzle, T. F.; Williams, T. F.; Meyer, J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(12) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(13) Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2444−2448.
(14) Bower, M. J.; Dunbrack, R. L. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.* **1997**, *267*, 1268−1282.
(15) Greenblatt, H. M.; Dvir, H.; Silman, I.; Sussman, J. L. Acetylcholinesterase: a multifaceted target for structure-based drug design of anticholinesterase agents for the treatment of Alzheimer's disease. *J. Mol. Neurosci.* **2003**, *20*, 369−383.
(16) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S. M.; Zaharevitz, D. National-Cancer-Institute Drug Information-System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219−1224.
(17) David, L.; Luo, R.; Gilson, M. K. Ligand−receptor docking with the Mining Minima optimizer. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157−171.
(18) Kairys, V.; Gilson, M. K. Enhanced docking with the Mining Minima optimizer: acceleration and side-chain flexibility. *J. Comput. Chem.* **2002**, *23*, 1656−1670.
(19) Baxter, C. A.; Murray, C. W.; Waszkowycz, B.; Li, J.; Sykes, R. A.; Bone, R. G. A.; Perkins, T. D. J.; Wylie, W. New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 254−262.
(20) Jacobsson, M.; Lidén, P.; Stjernschantz, E.; Boström, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781−5789.
(21) Elaine, C. M.; Brian, K. S.; Irwin, D. K. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(22) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(23) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2005**, ASAP Article, DOI: 10.1021/jm050362n.
(24) Han, M. S.; Kim, D. H. Effect of zinc ion on the inhibition of carboxypeptidase A by imidazole-bearing substrate analogues. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1425−1427.
(25) Maignan, S.; Guilloteau, J.-P.; Pouzieux, S.; Choi-Sledeski, Y. M.; Becker, M. R.; Klein, S. I.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Crystal structures of human factor Xa complexed with potent inhibitors. *J. Med. Chem.* **2000**, *43*, 3226−3232.
(26) Aranda, A.; Pascual, A. Nuclear hormone receptors and gene expression. *Physiol. Rev.* **2001**, *81*, 1269−1304.
(27) Fu, J.; Gaetani, S.; Oveisi, F.; Lo Verme, J.; Serrano, A.; Rodriguez De Fonseca, F.; Rosengarth, A.; Leucke, H.; DiGiacomo, B.; Tarzia, G.; Piomelli, D. Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPARα. *Nature* **2005**, *425*, 90−93.
(28) Cronet, P.; Petersen, J. F. W.; Folmer, R.; Blomberg, N.; Sjöblom, K.; Karlsson, U.; Lindstedt, E.-L.; Bamberg, K. Structure of the PPARα and -γ ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. *Structure* **2001**, *9*, 699−706.
(29) Nabel, E. G. CDKs and CKIs: molecular targets for tissue remodelling. *Nat. Rev. Drug Discovery* **2002**, *1*, 587−598.
(30) Russo, A. A.; Tong, L.; Lee, J.-O.; Jeffrey, P. D.; Pavletich, N. P. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor $p16^{INK4a}$. *Nature* **1998**, *395*, 237−243.
(31) Kryger, G.; Silman, I.; Sussman, J. L. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. *Structure* **1999**, *7*, 297−307.
(32) Koellner, G.; Steiner, T.; Millard, C. B.; Silman, I.; Sussman, J. L. A neutral molecule in a cation-binding site: specific binding of a PEG-SH to acetylcholinesterase from Torpedo californica. *J. Mol. Biol.* **2002**, *320*, 721−725.
(33) Cappalonga, A.; Alexander, R. S.; Christianson, D. W. Structural comparison of sulfodiimine and sulfonamide inhibitors in their complexes with zinc enzymes. *J. Biol. Chem.* **1992**, *267*, 19192−19197.
(34) Coll, M.; Guasch, A.; Aviles, F. X.; Huber, R. Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J.* **1991**, *10*, 1−9.
(35) Barbosa Pereira, P. J.; Segura-Martín, S.; Oliva, B.; Ferrer-Orta, C.; Avilés, F. X.; Coll, M.; Gomis-Rüth, F. X.; Vendrell, J. Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J. Mol. Biol.* **2002**, *321*, 537−547.
(36) Teplyakov, A.; Polyakov, K.; Obmolova, G.; Strokopytov, B.; Kuranova, I.; Osterman, A.; Grishin, N.; Smulevitch, S.; Zagnito, O.; Galperina, O. Crystal structure of carboxypeptidase T from Thermoactinomyces vulgaris. *Eur. J. Biochem.* **1992**, *208*, 281−288.
(37) Estébanez-Perpiñá, E.; Bayés, A.; Vendrell, J.; Jongsma, M. A.; Bown, D. P.; Gatehouse, J. A.; Huber, R.; Bode, W.; Avilés, F. X.; Reverter, D. Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest Helicoverpa armigera. *J. Mol. Biol.* **2001**, *313*, 629−638.
(38) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J. M.; Wienen, W.; Priepke, H. W.; Kauffmann, I. K.; Ries, U. J.; Hauel, N. H. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure* **2001**, *9*, 29−37.
(39) Hopfner, K.-P.; Lang, A.; Karcher, A.; Sichler, K.; Kopetzki, E.; Brandstetter, H.; Huber, R.; Bode, W.; Engh, R. A. Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding. *Structure* **1999**, *7*, 989−996.
(40) Recacha, R.; Costanzo, M. J.; Maryanoff, B. E.; Carson, M.; DeLucas, L.; Chattopdhyay, D. Structure of human α-thrombin complexed with RWJ-51438 at 1.7 A: unusual perturbation of the 60A-60I insertion loop. *Acta Crystallogr., Sect. D.* **2000**, *56*, 1395−1400.
(41) Sichler, K.; Banner, D. W.; D'Arcy, A.; Hopfner, K.-P.; Huber, R.; Bode, W.; Kresse, G. B.; Kopetzki, E.; Brandstetter, H. Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J. Mol. Biol.* **2002**, *322*, 591−603.
(42) Mather, T.; Oganessyan, V.; Hof, P.; Huber, R.; Foundling, S.; Esmon, C.; Bode, W. The 2.8 Å crystal structure of Gla-domainless activated protein C. *EMBO J.* **1996**, *15*, 6822−6831.
(43) Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-γ. *Nature* **1998**, *395*, 137−143.
(44) Kallen, J. A.; Schlaeppi, J.-M.; Bitsch, F.; Geisse, S.; Geiser, M.; Delhon, I.; Fournier, B. X-ray structure of the hRORα LBD at 1.63 Å: structural and functional data that cholesterol or a cholesterol derivative is the natural ligand of RORα. *Structure* **2002**, *10*, 1697−1707.
(45) Färnegårdh, M.; Bonn, T.; Sun, S.; Ljunggren, J.; Ahola, H.; Wilhelmsson, A.; Gustafsson, J. A.; Carlquist, M. The three-dimensional structure of the liver X receptor β reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *J. Biol. Chem.* **2003**, *278*, 38821−38828.
(46) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745−749.
(47) Brotherton, D. H.; Dhanaraj, V.; Wick, S.; Brizuela, L.; Domaille, P. J.; Volyanik, E.; Xu, X.; Parisini, E.; Smith, B. O.; Archer, S. J.; Serrano, M.; Brenner, S. L.; Blundell, T. L.; Laue, E. D. Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. *Nature* **1998**, *395*, 244−250.

(48) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117−1128.
(49) Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. Protein kinase A in complex with Rho-kinase inhibitors Y-27632, Fasudil, and H-1152P: structural basis of selectivity. *Structure* **2003**, *11*, 1595−1607.
(50) Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **2002**, *62*, 4236−4243.
(51) Dvir, H.; Wong, D. M.; Harel, M.; Barril, X.; Orozco, M.; Luque, F. J.; Munoz-Torrero, D.; Camps, P.; Rosenberry, T. L.; Silman, I.; Sussman, J. L. 3D structure of Torpedo californica acetylcholinesterase complexed with huprine X at 2.1 A resolution: kinetic and molecular dynamic correlates. *Biochemistry* **2005**, *41*, 2970−2981.
(52) Nicolet, Y.; Lockridge, O.; Masson, P.; Fontecilla-Camps, J. C.; Nachon, F. Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products. *J. Biol. Chem.* **2003**, *278*, 41141−41147.
(53) Bencharit, S.; Morton, C. L.; Hyatt, J. L.; Kuhn, P.; Danks, M. K.; Potter, P. M.; Redinbo, M. R. Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition. *Chem. Biol.* **2003**, *10*, 341−349.
(54) Grochulski, P. G.; Cygler, M. C. A structural basis for the chiral preferences of lipases. *J. Am. Chem. Soc.* **1994**, *116*, 3180−3186.
(55) Ghosh, D.; Wawrzak, Z.; Pletnev, V. Z.; Li, N. Y.; Kaiser, R.; Pangborn, W.; Jornvall, H.; Erman, M.; Duax, W. L. Structure of uncomplexed and linoleate-bound Candida cylindracea cholesterol esterase. *Structure* **1995**, *3*, 279−288.
(56) Spiller, B.; Gershenson, A.; Arnold, F. H.; Stevens, R. C. A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12305−12310.

CI050238C