JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

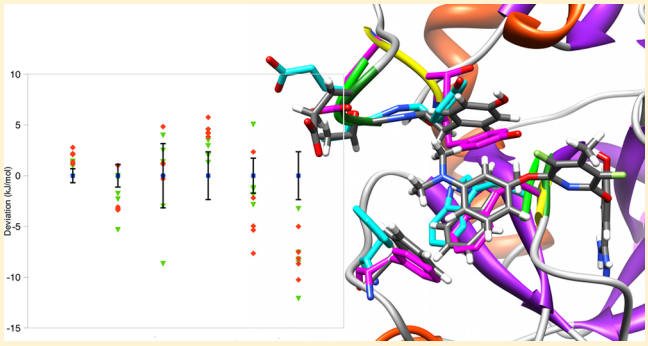
Are Homology Models Sufficiently Good for Free-Energy Simulations?

Samuel Genheden*

Division of Theoretical Chemistry, Department of Chemistry, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden

Supporting Information

ABSTRACT: In this paper, I evaluate the usefulness of protein homology models in rigorous free-energy simulations to determine ligand affinities. Two templates were used to create models of the factor Xa protein and one template was used for dihydrofolate reductase from *Plasmodium falciparum*. Then, the relative free energies for several pairs of ligands were estimated using thermodynamic integration with the homology models as starting point of the simulation. These binding affinities were compared to affinities obtained when using published crystal structures as starting point of the simulations. Encouragingly, the differences between the affinities obtained when starting from either homology models or crystal structure were not statistical significant for a majority of the considered pairs of ligands. Differences between 1 and 2 kJ/mol were observed for the dihydrofolate reductase ligands and differences between 0 and 8 kJ/mol were observed for the factor Xa ligands. The largest difference for factor Xa was caused by an erroneous modeling of a loop region close to two of the ligands, and it was only observed when using one of the templates. Therefore, it is advisible to always use more than one template when creating homology models if they should be used in free-energy simulations.



INTRODUCTION

Many biologically interesting processes can be studied using simulation techniques such as molecular dynamics and Metropolis Monte Carlo sampling.[1,2] Such processes are usually driven by a free-energy difference, and techniques of computing free-energy differences by simulating the system of interest have found widespread use in, for instance, drug design and enzymology.[3−5]

Most simulations of biological systems start with a structure of the macromolecule. The most accurate structures are obtained from experiments such as X-ray crystallography and NMR spectroscopy. Although the experimental database, the protein databank (PDB), grows each year, databases with sequence information grow at an even greater rate. In addition, all interesting macromolecules are not straightforwardly characterized by experimental techniques. To remedy the lack of experimental structures, theoretical techniques have been developed that are able to model protein structures.[6,7] The most widely used technique is homology or comparative modeling, in which an existing experimental structure is used as a template for the modeling of an unknown structure. Algorithms that can construct protein models without any known experimental structure are improving but are currently less reliable than homology modeling.

Naturally, it is important to evaluate theoretical techniques to model proteins. One celebrated approach is the CASP (critical assessment of techniques for protein structure prediction) challenge.[8] In this challenge, the contestants are asked to construct models for sequences that have recently been crystallized but not published. After submission, the constructed models are compared to the experimental structure and hence the theoretical techniques are evaluated in a blind-test fashion. Although the CASP challenge gives information on how well the models reproduce the experimental structure, the evaluation gives no clue to how the errors introduced in the models affect the final application of the models. One particularly interesting usage area for protein models is structure-based drug design.[9] In this area, several studies have been published on the use of protein models in virtual screening (docking).[10−14] The consensus is that the use of homology models can in some instances be as good as the use of experimental structures and that the use of models is always considerably better than random. Although virtual screening is useful for discarding a lot of bad molecules, the scoring functions used in docking are too inaccurate to be useful in later stages of drug design, e.g., lead optimization.[15] Here, more accurate simulation-based methods to estimate binding free energies should be used. Such methods can be divided into rigorous methods, e.g., alchemical perturbations methods,[4] that have a sound theoretical foundation and approximate methods that sacrifice the theoretical rigor for speed. Approximate

Received: July 26, 2012
Published: November 1, 2012

   dx.doi.org/10.1021/ci300349s | J. Chem. Inf. Model. 2012, 52, 3013−3021

methods such as MM/PBSA,[16] LIE,[17] and PDLD/s-LRA[18] are very popular but are typically less accurate.[19,20]

In this paper, I investigate if homology models can be used to obtain accurate binding free energies using alchemical perturbation techniques. To this end, I have used two templates to create homology models of the blood-clotting factor Xa and one template for *Plasmodium falciparum* dihydrofolate reductase, two pharmaceutically relevant proteins. For both proteins, there are crystal structures available in the protein databank.[21,22] Therefore, the relative free energies were computed for a series of inhibitors, starting both from the models and the experimental structures. The study reveals that homology models could be used to obtain accurate binding affinities, but it is advisible to check the results by creating models from several templates.

METHODS

Preparation of Systems. Two proteins were considered, human factor Xa (fXa) and *Plasmodium falciparum* dihydrofolate reductase (dhfr). Several crystal structures of both proteins are deposited in the protein databank; here I have chosen the 1fjs and 1j3i structures,[21,22] respectively, because they contain cocrystallized inhibitors that are similar to the ligands considered in this study. For fXa, the light chain was removed as it has been shown not to be essential for the estimation of accurate affinities.[23] For both proteins, all Glu and Asp residues were considered to be deprotonated, and all Arg, Lys as well as all isolated Cys residues were considered to be protonated. Eight Cys residues in fXa were involved in disulfide bridges and were deprotonated. The histidine residues in fXa were protonated as follow, after consideration of hydrogen-bond networks:[24] His145 and 199 were protonated on the NE2 atom, His83 and His91 were protonated on the ND1 atoms, and His57 was doubly protonated. All protein residues were described by the Amber99SB force field.[25]

The structural Ca^{2+} atom in fXa was kept and described with previously used parameters.[23] The NADPH cofactor of dhfr was kept as well and described with the general Amber force field[26] with charges obtained as for the ligands.

Seven fXa ligands and seven dhfr ligands were chosen and they are shown in Figure 1.[27,28] The ligands were built into the crystal structure by removing and adding a few atoms from the cocrystallized ligand as previously suggested.[29,30] The protonation states of the dhfr ligands were selected based on earlier studies.[28,30] All ligands were described with the general Amber force field.[26] The charges were obtained by first minimizing the ligands with the semiempirical AM1 method,[31] followed by calculation of the HF/6-31G* electrostatic potential of points sampled with the Merz−Kollman method.[32] Minimization and electrostatic potential calculations were performed with Gaussian 09.[33] Finally, charges were fitted to the ESP using the restrained ESP method.[34]

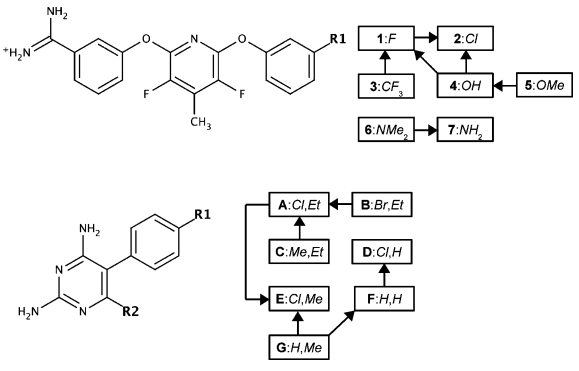


Figure 1. Ligands and transformations considered in this study. The upper panel shows the fXa ligands, and the lower panel shows the dhfr ligands. A template ligand is shown to the left and the sites changed during the transformations are denoted by R1 or R2. To the right, the ligands considered are shown in squares with their label in bold and the atoms at R1 or R2 in italics. The transformations are shown with arrows.

All protein−ligand complexes considered in this study were immersed in a pre-equilibrated truncated octahedral box of TIP3P water molecules.[35] The box extended at least 10 Å from the solute.

Homology Modeling. Two templates for fXa were selected from a previous study: factor IXa (pdb code: 1rfn[36]) with 44% sequence identity and protein C (pdb code: 1aut[37]) with 35% sequence identity.

Suitable template proteins for dhfr were found by a FASTA search[38] of the protein databank. Apart from structures from *P. falciparum*, such a search identified crystal structures from *P. vivax* (64% sequence identity), from *Babesia bovis* (35% sequence identity), and from *Cryptosporidium hominis* (30% sequence identity). The results of the FASTA search are summarized in Table 1. Finally, 2bl9, 3kjr, and 2oip were selected as templates from *P. vivax*, *B. bovis*, and *C. hominis*, respectively.[39−41]

The template structure was aligned with the sequence of the crystal structure of either fXa (only heavy chain) or dhfr using Modeler v9.10[42] using default gap penalties and substitution matrix. Thereafter, the *automodel* class in Modeler generated 100 models using default parameters. The models were assessed with DOPE (discrete optimized protein energy) score,[43] and the ten best models were used to compute free energies.

To obtain the final protein−ligand complex, the generated models were superposed onto the crystal structure by fitting the backbone heavy atoms using ProFit.[44] Thereafter, the coordinates of the ligand (as well as the Ca^{2+} ion for fXa and the NADPH molecule for dhfr) were copied to the model. This procedure eliminates the error source of inaccurate docking protocols.

Free-Energy Calculation. The relative free energies of binding for the pairs of ligands described in Figure 1 were computed using alchemical perturbations. Using a thermodynamic cycle, the relative free energy can be written as[3]

$$\Delta G_{\mathrm{bind}}(\mathrm{L1}) - \Delta G_{\mathrm{bind}}(\mathrm{L0}) = \Delta G_{\mathrm{bound}}(\mathrm{L0} \rightarrow \mathrm{L1}) - \Delta G_{\mathrm{free}}(\mathrm{L0} \rightarrow \mathrm{L1}) \quad (1)$$

where ΔG_{bind} is the binding free energy of a ligand (L0 or L1) and $\Delta G(\mathrm{L0}\rightarrow\mathrm{L1})$ is the free energy of alchemically transforming L0 to L1, either when the ligands are bound to the protein or when they are free in solution. The transformation free energies were estimated using thermodynamic integration[45]

$$\Delta G(\mathrm{L0} \rightarrow \mathrm{L1}) = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \quad (2)$$

where $V = (1 - \lambda)V_0 + \lambda V_1$, V_0, and V_1 are the potentials of the systems with L0 and L1, respectively, and λ is a coupling parameter. Simulations at specific λ-values were performed (as

Table 1. Statistics of FASTA Search and Homology Models[a]

fXa Models							
no. hits	PDB ID	FASTA identity	e-value	alignment identity	backbone RMSD	site backbone RMSD	site heavy atom RMSD
factor IXa template							
4	IRFN	43.8	6.0×10^{-44}	44.4	1.4 ± 0.0	0.8 ± 0.01	1.2 ± 0.02
factor C template							
2	1AUT	35.9	4.9×10^{-36}	37.2	2.6 ± 0.0	1.0 ± 0.02	1.4 ± 0.02
dhfr models							
no. hits	PDB ID	FASTA identity	e-value	alignment identity	backbone RMSD	site backbone RMSD	site heavy atom RMSD
Plasmodiumg vivax template							
4	2BL9	63.7	2.8×10^{-38}	68.01	3.8 ± 0.2	1.0 ± 0.01	1.2 ± 0.01
Babesia bovis template							
8	3KJR	34.9	5.7×10^{-16}	23.4	24.3 ± 1.0		
Cryptosporidium hominis template							
20	2OIP	30.4	5.7×10^{-15}	25.6	29 ± 1.1		

[a]The columns are from left to right: the number of FASTA search hits of each template, the PDB code of the chosen template, the reported sequence identity from the FASTA search in percentage, the FASTA e-value, the reported sequence identity from Modeler in percentage, the root mean squared deviation (RMSD) of all heavy backbone atoms (C, CA, O, and N) when comparing the homology model to the crystal structure in Ångström, the RMSD of only the backbone atoms of residues within 5 Å of ligand, the RMSD of all heavy (non-hydrogen) atoms of residues within 5 Å of the ligand. For the two last templates of dhfr, the site rms was not evaluated because the overall backbone rms was too high to be useful. The active site residues were determined from the crystal structure and therefore the list of residues is identical for each model.

discussed in the text below) and the integral in eq 2 was evaluated using the trapezoid method. Values of the integrand at $\lambda = 0$ and 1 were obtained by linear extrapolation. In all simulations, soft-core Coulomb and Lennard-Jones potentials as implemented in Amber11 were used and charges and van der Waals parameters were perturbed simultaneously using a dual-topology.[46,47]

Simulations. All molecular dynamics simulations were run using the sander module of Amber 11.[48] A time step of 2 fs was used together with SHAKE constraints[49] on bonds involving hydrogen atoms. The cutoff for the van der Waals interactions was 8 Å, and the long-range correction was estimated using a continuum approach. Electrostatic interactions were treated using particle-mesh Ewald summation[50] with a real-space cutoff of 8 Å. The nonbonded pair list was updated every 50 fs. Temperature was kept constant at 300 K using Langevin dynamics,[51] and pressure was kept constant at 1 atm using a weak-coupling isotropic algorithm.[52]

For the simulations starting from the crystal structure, five independent simulations were initiated by solvating the system in different water boxes and assigning different starting velocities at each λ-value. Each of the ten generated homology models were solvated in different water boxes and assigned different starting velocities. Each simulation started with 100 steps of steepest descent minimization using a harmonic restraint with a force of constant of 418 kJ/(mol Å^2) on all atoms, except water molecules and hydrogen atoms, followed by a 20 ps simulation in the NPT ensemble with the same restraints, and a 500 ps simulation in the same ensemble but without any restraints. Finally, a 1 ns production run was performed in the NPT ensemble and free energy data were saved every 10 ps, i.e., 100 snapshots at each intermediate state were saved. This has previously been shown to be sufficient.[19]

Error Estimates. The uncertainty at each λ-value was obtained by calculating the standard deviation of the mean over the five or ten independent simulations. The total uncertainty was then obtained by error propagation.

The performance of the free-energy estimates was quantified by the mean unsigned error (MUE) compared to experimental data, the correlation coefficient (r^2), and Kendall's τ (calculated only for the pairs explicitly simulated). The uncertainty of these quality metrics were obtained by a simple parametric bootstrap simulation as has been described previously,[53] taking into account the uncertainty both in the estimations and the experimental data.

RESULTS AND DISCUSSION

I have estimated the free energy of binding for a set of fXa and dhfr ligands when the simulations were initiated both from a crystal structure and from ten homology models. Below, I discuss the results, first for fXa and then for dhfr.

Factor Xa. I computed the relative free energy of binding for seven ligands in a total of six transformations, shown in Figure 1. A different set of factor Xa (fXa) ligands has been studied previously with a similar protocol used herein.[19] Therefore, I used five λ-values (0.05, 0.3, 0.5, 0.7, 0.95) for the computation of ΔG_{bound} and ΔG_{free}. Even three λ-values have been shown to give excellent results for fXa,[19] but since that was for another set of ligands, I here use two more λ-values. First, I computed the free energies starting from a crystal structure and the results are shown in Table 2 together with the experimental results. It is clear that the computed energies reproduce the experimental energies excellently; the mean

Table 2. Free Energies for fXa Ligands in kJ/mol

	ΔG_{bound}[a]	ΔG_{free}	$\Delta\Delta G_{bind}$ (calc)	$\Delta\Delta G_{bind}$ (exp)
1 → 2	205.7 ± 0.3	207.6 ± 0.1	−1.9 ± 0.4	−1.6
3 → 1	−63.2 ± 0.6	−67.6 ± 0.1	4.4 ± 0.6	1.7
4 → 1	−222.9 ± 1.5	−221.6 ± 0.5	−1.4 ± 1.6	−1.1
4 → 2	−17.5 ± 1.1	−13.3 ± 0.5	−4.3 ± 1.2	−2.7
5 → 4	122.7 ± 0.8	115.4 ± 0.3	7.3 ± 0.9	3.3
6 → 7	297.4 ± 1.1	285.4 ± 0.5	12.0 ± 1.2	7.5
MUE[b]			2.2 ± 0.4	
r^2			0.98 ± 0.03	
τ			1.00 ± 0.13	

[a]Obtained by starting the simulations from the crystal structure. [b]Mean unsigned error (MUE), r^2, and τ calculated relative to the experimental results.

dx.doi.org/10.1021/ci300349s | J. Chem. Inf. Model. 2012, 52, 3013−3021

unsigned error (MUE) is 2 kJ/mol, the correlation coefficient, r^2, is 0.98, and Kendall's τ shows that all of the computed energies have the correct sign. The maximum deviation is 4 kJ/mol for the **5** → **4** and **6** → **7** mutations. The uncertainties of the computed energies are also good, around or below 1 kJ/mol for all of the transformations, with the exception of **4** → **1** that has an uncertainty of 2 kJ/mol.

Having established that accurate free energies can be estimated for fXa when starting from a crystal structure, 10 homology models were generated by Modeler with factor IXa as the template (100 were created but only the 10 best were used). Statistics of the generated models are shown in Table 1. The backbone root-mean-square deviation (RMSD) of all generated residues is 1.4 Å when compared to the crystal structure of fXa. However, considering only the residues within 5 Å of the ligands, the RMSD is only 0.8 Å for the backbone and 1.2 Å for all heavy (non-hydrogen) atoms. Heavy atoms RMSD of 5 and 6 Å are found for Glu98 and Thr99, residues that are only within 5 Å of ligands **6** and **7**. This indicate the Modeler were able to create models that overall were similar to the crystal structure, with some large deviations for a few binding-site residues.

For each of the ten homology models, I computed the relative free energy for the six pairs of ligands. The results are summarized in Table 3, and a scatterplot of all the results are presented in Figure 2a. Considering the average over all ten models, the computed free energy is significantly different compared to the crystal-structure estimate for four of the six transformations. The transformations **4** → **1** and **5** → **4** show deviations of 0.3 and 2.0 kJ/mol, compared to the crystal-structure estimate, deviations that are not statistically significant. For the other transformations, the deviation ranges from 2.2 to 7.9 kJ/mol. The mean absolute deviation (MAD) is 2.9 kJ/mol. Interestingly, the relative binding estimates become closer to experiments for four of the six transformations, although the correlation deteriorates (r^2 = 0.73). Looking at the individual estimates from the models in Figure 2, it is first clear that there is no correlation between the DOPE score of the homology model and its deviation to the crystal-structure estimate, i.e., there is no indication that the five top-ranked models are grouped separately from the other models. For transformations **1** → **2** and **4** → **2**, all of the estimates based on homology models are more positive than the crystal-structure estimate, and for the transformation **6** → **7**, all of the homology-based estimates are more negative. For the other transformations, there seems to be a more random spread of the estimates. The standard deviation over all independent simulations (see Table 3) is similar to the uncertainties observed for the crystal-structure estimates. On a side note, it could be argued that the two calculations of standard deviations reflect different uncertainties because the uncertainties of the homology-model based results also reflect variability in the protein structure used to start the simulations. On the other side, the standard deviation is the uncertainty of an ensemble of five or ten free energy estimates, and all of these ensembles are equivalent from a thermodynamic perspective. The only difference being that the simulations was started from different locations in phase space. The effect of starting from different locations in phase space has been tested within the framework of MM/GBSA.[54]

Next, ten homology models for fXa were created by using protein C as the template to evaluate if the estimation of free energies is dependent on the template structure used. Statistics

Table 3. Free Energies for fXa Ligands When Starting from Homology Models in kJ/mol

	fIXa		protein C	
template	ΔG_{bound}	Δ^a	ΔG_{bound}	Δ^a
1 → **2**	207.1 ± 0.3	1.4	206.5 ± 0.3	0.8
3 → **1**	−65.3 ± 0.6	−2.2	−65.8 ± 0.7	−2.6
4 → **1**	−222.6 ± 1.2	0.3	−222.4 ± 1.0	0.5
4 → **2**	−13.9 ± 0.4	3.7	−15.7 ± 1.1	1.8
5 → **4**	120.7 ± 1.2	−2.0	121.2 ± 1.8	−1.5
6 → **7**	289.5 ± 0.8	−7.9	295.38 ± 0.9	−2.0
MAD		2.9	1.5	

[a]Deviation compared to starting from the crystal structure.

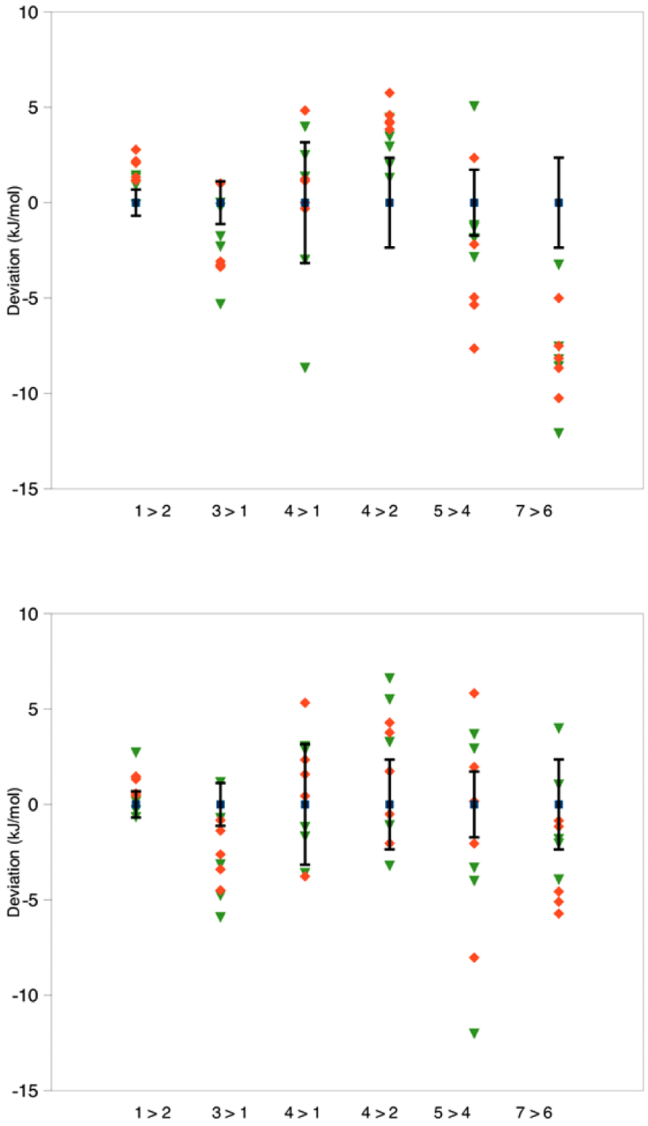


Figure 2. Free energies for the fXa ligands starting from either the crystal structure or the homology models. The free energies are plotted as a deviations compared to the crystal structure estimate. The error bars illustrate a 95% confidence interval of the crystal structure estimate. The squares show the estimates from the five first generated homology models (i.e., the five models with highest DOPE score), and the triangles show the estimates from the five last generated homology models. There is one column of data for each of the six considered transformations: (A) factor IXa template; (B) protein C template.

dx.doi.org/10.1021/ci300349s | J. Chem. Inf. Model. 2012, 52, 3013−3021

of the homology model are shown in Table 1. Because protein C has a lower sequence identity with fXa than flXa, it is not surprising to see that the overall backbone RMSD is slightly larger when using this template, 2.6 Å. Also the site RMSD is significantly higher, but still reasonably small 1.0 and 1.4, for backbone and all heavy atoms of the active site, respectively. The largest individual deviation, 2 Å, is found for Glu98, indicating that this template was in fact better than flXa. This has been noted previously as well.[13]

The computed free energies when using these homology models are summarized in Table 3 and plotted in Figure 2b. Considering that the generated models were closer to the crystal structure than when using flXa as the template, it is not so surprising that I obtain significantly better results with these models. Only for a single transformation, **3** → **1**, the use of homology models results in a significant difference when compared to the crystal-structure estimate. The MAD of the homology-based estimates is 1.5 kJ/mol when comparing to the crystal-structure estimates. Looking at the individual estimates in Figure 2b, it is again clear that there is no difference between the results initiated from the five first homology models and the other ones. Also, the standard deviation is similar to the crystal-structure estimates (compare Tables 2 and 3). By comparing the estimates obtained from the two sets of homology models, only the transformation **6** → **7** shows a statistically significant difference, and the MAD is 1.6 kJ/mol. This indicates that it is a mere coincidence that the protein C template resulted in estimates closer to the crystal-structure estimates for all transformations except the **6** → **7**.

Because it seems problematic to estimate binding free energies for **6** → **7**, I looked closer at this transformation. This transformation involves the two largest ligand in the test set, and it therefore seem reasonable to look at residues close to the transformation site. Figure 3 shows residues within 5 Å of the

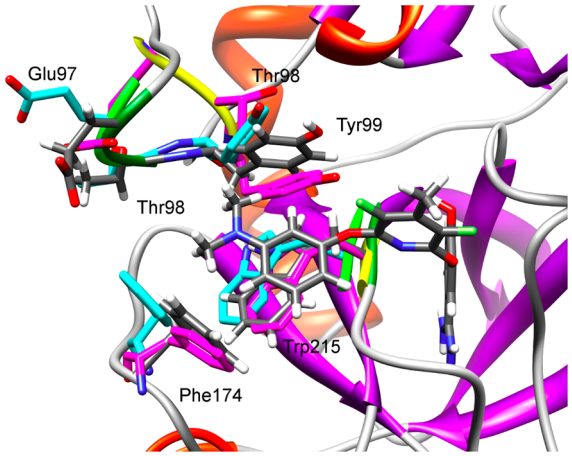


Figure 3. Residues around the site of transformation for the fXa ligands **6** and **7**. Only ligand **6** is shown together with five residues (Glu97, Thr98, Tyr99, Phe174, and Trp 215). The loop containing residues 97−99 is shown in in two different colors, one for the model based on the flXa template and one for the other structures.

nitrogen atom at the transformation site for the crystal structure and a representative model for each of the two templates. It is clear that the biggest problem with the flXa model is the loop containing residues Glu97, Thr98, and Tyr99. This loop has been modeled quite differently from the crystal structures when I used flXa as the template, whereas it has been modeled quite successfully when protein C was used as the template. It seems likely that Tyr99 that is fairly close to the nitro-group of **6** and **7** can affect the binding affinity. For the other ligands, there is more space available to accommodate both the ligand and the alternative conformation of Tyr99. This was confirmed by decomposing the total free energy on a residue basis.[55−57] Although it should be noted that this is an approximation, because it is only the total free energy that is well-determined, it will give qualitative information on the which residues contribute most to the free energy. The five residues that show the largest differences when comparing the crystal-structure estimate and the estimate based on the flXa template are exactly the five residues shown in Figure 3 (see Table 4).

Table 4. Contribution of Residues to ΔG_{bound} in kJ/mol for the 6 → 7 Transformation[a]

residue	crystal structure	flXa template	Δ
Tyr99	−1.6	−4.2	2.6
Thr98	−2.0	−0.2	−1.9
Phe174	1.4	0.1	1.3
Glu97	4.8	3.8	1.0
Trp215	1.9	1.2	0.7

[a]The five residues with the largest difference when comparing the crystal-structure estimate and the estimate based on flXa as a template. Contributions are shown for simulations starting from either the crystal structure or from homology models based on the flXa template, as well as the difference.

Interestingly, the contribution from four of the five residues is more favorable in the homology-model based estimate. The differences of these five residues amounts to a total of 4 kJ/mol, i.e., about half of the observed difference (cf. Table 3). The rest of the difference comes from the ligands.

I performed a 5 ns simulation of **6** for five of the homology models based on the flXa template in order to equilibrate the models. However, this did not correct the poorly modeled loop region but instead pushed the ligand to a slightly different binding mode. I also attempted to estimate free energies, starting from this equilibrated structure, and it resulted in an estimate of 290.8 ± 1.7 kJ/mol, i.e., not significantly different from the original estimate (cf. Table 3). This analysis shows that small changes to a few residues can have a considerable effect on the estimated free energies. To remedy this, more advanced modeling techniques could be attempted, e.g., modeling based on several templates simultaneously or loop modeling, but without comparing to the crystal structure it is not possible to determine if such approaches improves the model.

Dehydrofolate Reductase. Free energies for the dhfr ligands computed with alchemical perturbations have not been published previously, so here it is important to first establish a good simulation protocol. Therefore, I choose to estimate the free energy for the **A** → **E** and **B** → **A** transformations using 11 λ-values (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 0.9, and 0.95) as representative transformations for the 2 ligand sites (c.f. Figure 1). The derivative of the total potential ($\partial V/\partial\lambda$) as a function of λ is shown in Figure 4. For both transformations, I obtained rather smooth curves both in the bound and free states, which is important when using TI. Starting with the **A** → **E** transformation, $\partial V/\partial\lambda$ shows a minimum at about λ = 0.5 and is rather linear up to the end-points. This indicates that it should be sufficient with three λ-values at 0.05, 0.5, and 0.95.

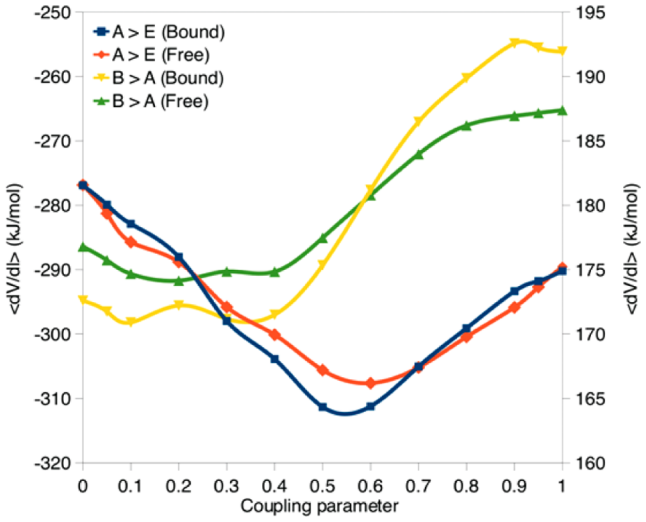


Figure 4. Derivative of potential as a function of coupling parameter (λ) for the transformations **A** → **E** (left axis) and **B** → **A** (right axis).

The free energies are shown in Table 5, for 11, 5 (0.05, 0.3, 0.5, 0.7, and 0.95), and 3 λ-values. For the bound state, there is only a small difference when using either 3 or 5 λ-values compared to using 11. However, for the free state, there is a difference of 1.6 kJ/mol when comparing 11 and 3 λ-values. Therefore, it seems that five λ-values is preferable. The shape of the curve for the **B** → **A** transformation is a little bit more involved, but on average it is fairly flat from λ = 0 to 0.5 and fairly linear from λ = 0.5 to 1.0. Therefore, I attempted to estimate the free energies using five and three λ-values for this transformation as well. As can be seen in Table 5, this turns out to be an excellent approximation with differences less than 0.5 kJ/mol for both the free and bound state. However, taking into account the slightly larger deviation for the **A** → **E** transformation and to be on the safe side when estimating free energies for the other ligand pairs, I chose to use five λ-values in what follows.

Table 5. Free Energies for the A → E and B → A Transformations with a Varying Number of Integration Points in kJ/mol

no. λ-values	ΔG_{bound}	ΔG_{free}
	A → E	
11	−297.6 ± 0.3	−296.8 ± 0.2
5	−297.8 ± 0.4	−296.7 ± 0.4
3	−297.2 ± 0.7	−295.3 ± 0.5
	B → A	
11	179.3 ± 0.2	179.6 ± 0.1
5	179.3 ± 0.4	179.8 ± 0.1
3	179.0 ± 0.5	179.7 ± 0.1

The binding free energies for all six pair of ligands when starting from the crystal structure are shown in Table 6. The estimates are quite good, but slightly worse than those for fXa. The MUE is 3 kJ/mol and r^2 is 0.9, but τ shows that two of the transformations have the incorrect sign. However, for both of these transformations, **A** → **E** and **B** → **A**, the estimates are not significantly different from zero, hence it is not possible to significantly determine their signs. The uncertainty of the estimates are less than 1 kJ/mol, except for **G** → **F**. It should also be mentioned that all of the estimates are more negative than the experimental results, pointing to a systematic error. To further investigate this, a thermodynamic cycle was created by computing also the **D** → **B** transformation. The calculated free energy for this transformation was −1.2 ± 1.7 kJ/mol, compared to the experimental value of 5.2 kJ/mol. The total free energy of the cycle sums to 1.4 ± 2.5 kJ/mol, i.e., the cycle closes within the uncertainty of the calculations. Hence, it can be concluded that the calculations are consistent.

Next, I created 10 homology models using *P. vivax* as a template (100 were created but only the 10 best were used). Statistics of the generated models are shown in Table 1. Even though the sequence identity is much higher (68%) for this template than any of the templates considered for fXa, the overall backbone RMSD is rather high, 3.8 Å. However, considering only the residues within 5 Å of a representative ligand, the backbone RMSD and heavy atom RMSD is only 1.0 and 1.2 Å, respectively. Therefore, I expect that it should be possible to estimate accurate free energies using these models. The estimations in Table 7 are indeed very close to the crystal-

Table 7. Free Energies for dhfr Ligands When Starting from Homology Models in kJ/mol

	ΔG_{bound}	Δ^a
A → E	−297.0 ± 0.5	0.6
B → A	180.9 ± 0.3	1.6
C → A	−286.3 ± 0.4	−1.1
F → D	40.0 ± 0.5	−1.2
G → E	−51.5 ± 0.5	−0.5
G → F	−70.7 ± 1.2	−1.2
MAD		1.0

[a]Deviation compared to starting from the crystal structure.

Table 6. Free Energies for dhfr Ligands in kJ/mol

	ΔG_{bound} [a]	ΔG_{free}	$\Delta\Delta G_{bind}$ (calc)	$\Delta\Delta G_{bind}$ (exp)
A → E	−297.6 ± 0.4	−296.7 ± 0.3	−0.9 ± 0.5	3.0 ± 0.9
B → A	179.3 ± 0.5	179.8 ± 0.1	−0.5 ± 0.5	1.7 ± 0.8
C → A	−285.3 ± 0.2	−285.4 ± 0.3	0.1 ± 0.3	1.0 ± 0.8
F → D	41.2 ± 0.5	49.7 ± 0.2	−8.6 ± 0.5	−7.1 ± 0.4
G → E	−51.0 ± 0.7	−42.2 ± 0.4	−8.8 ± 0.8	−4.2 ± 0.4
G → F	−69.5 ± 1.3	−70.8 ± 0.3	1.3 ± 1.4	3.4 ± 0.4
MUE[b]			2.5 ± 0.4	
r^2			0.90 ± 0.06	
τ			0.33 ± 0.24	

[a]Obtained by starting the simulations from the crystal structure. [b]Mean unsigned error (MUE), r^2, and τ calculated relative to the experimental results.

dx.doi.org/10.1021/ci300349s | J. Chem. Inf. Model. 2012, 52, 3013−3021

structure estimates. For only a single transformation, **B** → **A**, there is a significant difference compared to the crystal structure-estimation. This can partly be attributed to the very precise estimate. The MAD of the homology-based estimates is 1 kJ/mol when comparing to the crystal-structure estimates. For the **A** → **F** and the **B** → **A** transformations, the homology model-based estimate is closer to experiments than the crystal-structure based estimate. In Figure 5, all the estimates are

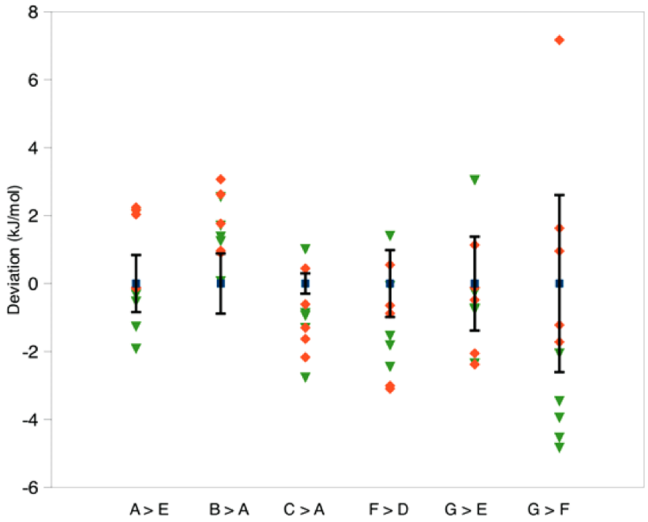


Figure 5. Free energies for the dhfr ligands starting from either the crystal structure or from the homology models. See Figure 2 for an explanation of the plot.

plotted and it is clear again that there is no grouping among models. Also, for most of the transformation, the model-based estimates seem to be statistically spread around the crystal-structure estimates. The uncertainty of the model-based estimates is similar to the uncertainty of the crystal-structure estimate.

I also attempted to create homology models using *B. bovis* and *C. homini* as templates. However, Modeler were unable to create any reasonable models shown by the overall backbone RMSD > 20 Å. This can most likely be attributed to the low sequence identity. Finally, I also attempted to build protein models using *B. bovis* as the template with the I-TASSER web-service.[58] While the backbone RMSD was ∼12 Å and, therefore, better than the homology models of Modeler, I considered this to be an unusable model.

CONCLUSIONS

I have evaluated the usefulness of protein homology models in rigorous free-energy simulations by utilizing something that could be called unnecessary models. The models are unnecessary in the way that crystal structures do exist for the chosen proteins, but in the context of the current study they are very useful. The ligand binding affinities obtained by using the homology models can be compared to the affinities obtained when using the crystal structure and as such, this study gives an indication how forgiving the free-energy simulations are to the errors introduced by the homology modeling (assuming docking errors are small). This is an important question to ask when a crystal structure is missing, because then one has to resort to modeling. Naturally, if a crystal structure exists, the homology models are unnecessary (unless it is of interest to create an ensemble of different structures). Although size of the test is rather small, I believe that results give clear indication on the possibility and limitations of the use of homology models in free-energy simulations.

First, it is clear that for most of the transformations and test systems considered in this study, it does not matter whether you start from the crystal structure or a homology model because the differences in calculated affinity are not statistically significant. This is encouraging and shows that free-energy simulations are insensitive to moderate errors of 1−4 Å as characterized by the total backbone RMSD, and errors of 1−2 Å of binding site heavy atom RMSD. It also shows that if the sequence identity is fair, 44−70%, one could obtain accurate free-energy estimates. Naturally, it is difficult to draw any general conclusions regarding what level of sequence identity is necessary from this study.

Furthermore, I have shown that the estimates obtained from the homology models usually show a similar uncertainty as the crystal-structure estimates. This might come as a surprise because the protein structures were identical in the five independent simulations started from the crystal structure, whereas they were all different in the case of homology models. Hence, it seems that the homology models should sample a greater part of phase space, and it has previously been observed that MD simulations sample a greater part of phase space than an ensemble of crystal structures.[59] Furthermore, I have noted that the homology model-based estimates sometimes become closer to the experimental binding affinities. This indicates that the used crystal structure could be a suboptimal starting structure for some of the ligands and the uncertainties in the protein models lead to sampling different parts of phase space. These two observations suggest that if the homology models can be trusted, they can be used to sample a greater part of phase space and at the same time have an acceptable precision. This could perhaps be used as a complement to simpler structural perturbations, e.g., rotation of side-chain and hydroxyl groups, that have been shown to enhance sampling for some protein systems.[54]

This study has also showed that larger errors introduced at residues in the binding sites can have large impact on the calculated free energies. This was exemplified by the **6** → **7** transformation in fXa. Here did the erroneous modeling of only three residues affect the calculated relative binding affinity by 8 kJ/mol. Interestingly, this occurred when using fIXa as a template, which has a 44% sequence identity, but not when using protein C as a template, which only has 37% sequence identity. This shows that the success of homology modeling is a much more complex issue than simply selecting the template with the highest sequence identity. The question is how one should relate to this in a blind (prospective) scenario, i.e., when there is no crystal structure to compare to. My suggestion is to create several homology models from different templates and compute free energies for all of them. If the difference in the estimates from the different templates is significant, one has to consider the estimate to be highly uncertain. For the estimates where the difference is not significant, one can argue that the estimation is reliable. In conclusion, it is worth trying to compute accurate binding affinities with homology models when there is no experimental structure available.

ASSOCIATED CONTENT

Supporting Information

The force field of all ligands as well as the NADPH cofactor and PBD files of all homology models used in the free-energy

dx.doi.org/10.1021/ci300349s | J. Chem. Inf. Model. 2012, 52, 3013−3021

simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Samuel.Genheden@teokem.lu.se. Tel.: +46-46 2224915. Fax: +46-46 2228648.

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This investigation has been supported by grants from the Research school in pharmaceutical science. It has also been supported by computer resources of LUNARC at Lund University (project SNIC001-11-279) and HPC2N at Umeå University (project SNIC020-11-20).

REFERENCES

(1) Adcock, S. A.; McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106*, 1589−1615.

(2) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättlie, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular modeling: Goals, problems, perspectives. *Ang. Chem. Int. Ed.* **2006**, *45*, 4064−4092.

(3) Michel, J.; Essex, J. Prediction of protein−ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* **2010**, *24*, 639−658.

(4) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical free energy calculations: ready for prime time? *Ann. Rep. Comput. Chem.* **2007**, *3*, 41−59.

(5) Senn, H. M.; Thiel, W. Atomistic approaches in modern biology: from quantum chemistry to molecular simulations. *Top. Curr. Chem.* **2007**, *268*, 173−290.

(6) Floudas, C. A.; Fung, H. K.; McAllister, S. R.; Mönnigmann, M.; Rajgaria, R. Advances in protein structure prediction and de novo protein design: a review. *Chem. Eng. Sci.* **2006**, *61*, 966−988.

(7) Zhang, Y. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.* **2008**, *18*, 342−348.

(8) Moult, J.; Fidelis, K.; Kryshtafovych, A.; Tramonanto, A. Critical assessment of methods of protein structure prediction (CASP) − round IX. *Proteins* **2011**, *79*, 1−5.

(9) Cavasotto, C. N.; Phatak, S. S. Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* **2009**, *14*, 676−683.

(10) Fan, H.; Irwin, J. J.; Webb, B. M.; Klebe, G.; Shoichet, B. K.; Sali, A. Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.* **2009**, *49*, 2512−2527.

(11) 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.

(12) 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.

(13) Kairys, V.; Fernandes, M. X.; Gilson, M. K. Screening drug-like compounds by docking to homology models: A systematic study. *J. Chem. Inf. Model.* **2006**, *46*, 365−379.

(14) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screen against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 365−379.

(15) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Ang. Chem. Int. Ed.* **2002**, *41*, 2644−2676.

(16) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(17) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* **1998**, *12*, 27−35.

(18) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE. PDLD-LRA, and PDLS/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 393−407.

(19) Genheden, S.; Nilsson, I.; Ryde, U. Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51*, 947−958.

(20) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host−guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput. Aided Mol. Des.* **2012**, *26*, 527−541.

(21) Adler, M.; Davey, D. D.; Phillips, G. B.; Kim, S. H.; Jancarik, J.; Rumenik, G.; Light, D. R. Whitlow, M. Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. *Biochem* **2000**, *39*, 12534−12542.

(22) Yuvaniyama, J.; Chitnumsub, P.; Kamchonwongpaisan, S.; Vanichtanankul, J.; Sirawaraporn, W.; Taylor, P.; Walkinshaw, M. D.; Yuthavong, Y. Insights into antifolate resistance from malarial DHFR-TS structures. *Nat. Struct. Biol.* **2003**, *10*, 357−365.

(23) Genheden, S.; Ryde, U. Improving efficiency of protein−ligand binding free-energy calculations by system truncation. *J. Chem. Theory Comput.* **2012**, *8*, 1449−1458.

(24) Kongsted, J.; Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comput. Aided Mol. Des.* **2009**, *23*, 63−71.

(25) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinform.* **2006**, *65*, 712−725.

(26) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(27) Phillips, G.; Davey, D. D.; Eagen, K. A.; Koovakkat, S. K.; Liang, A.; Ng, H. P.; Pinkerton, M.; Trinh, L. Whitlow, M.; Beatty, A. M.; Morrissey, M. M. Design, synthesis, and activity of 2,6-diphenoxypyridine-derived factor Xa inhibitors. *J. Med. Chem.* **1999**, *42*, 1749−1756.

(28) Tarnchompoo, B.; Sirichaiwat, C.; Phupong, W.; Intaraudom, C.; Sirawaraporn, W.; Kamchonwongpaisan, S.; Vanichtanakul, J.; Thebtaranonth, Y.; Yuthavong, Y. Development of 2,4-diaminopyrimidines as antimalarials based on inhibition of the S108N and C59R +S108N mutants of dihydrofolate reductase from pyrimethamine-resistant *Plasmodium falciparum*. *J. Med. Chem.* **2002**, *45*, 1244−1252.

(29) Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958−970.

(30) Rastelli, G.; del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and accurate predicitions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797−810.

(31) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(32) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H, P.; Cross, J,

B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G, A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Cliffordmgrt, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Gaussian, Inc.: Wallingford, CT, 2004.

(34) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(36) 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. *Struct. Fold. Des.* **1999**, *7*, 989−996.

(37) Mather, T.; Oganessyan, V.; Hof, P.; Huber, R. Foundling, S.; Esmon, C.; Bode, W. The 2.8 A crystal structure of Gla-domainless activated protein C. *EMBO J.* **1996**, *15*, 6822−6831.

(38) Lipman, D. J.; Pearson, W. R. Rapid and sensitive protein similarity searches. *Science* **1984**, *227*, 1435−1441.

(39) Kongsaeree, P.; Khongsuk, P.; Leartsakulpanich, U.; Chitnumsub, P.; Tarnchompoo, B.; Walkinshaw, M. D.; Yuthavong, Y. Crystal structure of dihydrofolate reductase from Plasmodium vivax: pyrimethamine displacement linked with mutation-induced resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13046−13051.

(40) Li, L.; Du, W.; Ismagilov, R. F. Multiparameter screening on SlipChip used for nanoliter protein crystallization combining free interface diffusion and microbatch methods. *J. Am. Chem. Soc.* **2010**, *132*, 112−119.

(41) Doan, L. T.; Martucci, W. E.; Vargo, M. A.; Atreya, C. E.; Anderson, K. S. Nonconserved residues Ala287 and Ser290 of the Cryptosporidium hominis thymidylate synthase domain facilitate its rapid rate of catalysis. *Biochemistry* **2007**, *46*, 8370−8391.

(42) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

(43) Shen, M. Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15*, 2507−2524.

(44) Martin, A. C. R. http://www.bioinf.org.uk/software/profit/ (accessed November 1, 2012).

(45) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.

(46) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.

(47) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J. Comput. Chem.* **2011**, *32*, 3253−3263.

(48) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(49) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(50) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(51) Wu., X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512−518.

(52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(53) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *31*, 837−846.

(54) Genheden, S.; Ryde, U. Comparison of different initialisation protocols to generate independent molecular dynamics simulations. *J. Comput. Chem.* **2012**, *32*, 187−195.

(55) Mark, A. E.; van Gunsteren, W. F. Decomposition of the free energy of a system in terms of specific interactions. *J. Mol. Biol.* **1994**, *240*, 167−176.

(56) Brady, G. P.; Szabo, A.; Sharp, K. A. On the decomposition of free energies. *J. Mol. Biol.* **1996**, *263*, 123−125.

(57) Bren, M; Florian, J.; Mavri, J.; Bren, U. Do all pieces make a whole? Thiele cumulants and the free energy decomposition. *Theor. Chem. Acc.* **2007**, *117*, 535−540.

(58) Ambrish, R.; Alper, K.; Yang, Z. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Prot.* **2010**, *5*, 725−738.

(59) Bren, U.; Oostenbrink, C. Cytochrome P450 3A4 Inhibition by Ketoconazole: Tackling the Problem of Ligand Cooperativity Using Molecular Dynamics Simulations and Free-Energy Calculations. *J. Chem. Inf. Model.* **2012**, *52*, 1573−1582.