

Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA

GIULIO RASTELLI, ALBERTO DEL RIO, GIANLUCA DEGLIESPOSTI, MIRIAM SGOBBA
Dipartimento di Scienze Farmaceutiche, Università di Modena e Reggio Emilia, Via Campi 183,
41100 Modena, Italy

Received 25 September 2008; Accepted 29 May 2009

DOI 10.1002/jcc.21372

Published online 30 June 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: In the drug discovery process, accurate methods of computing the affinity of small molecules with a biological target are strongly needed. This is particularly true for molecular docking and virtual screening methods, which use approximated scoring functions and struggle in estimating binding energies in correlation with experimental values. Among the various methods, MM-PBSA and MM-GBSA are emerging as useful and effective approaches. Although these methods are typically applied to large collections of equilibrated structures of protein-ligand complexes sampled during molecular dynamics in water, the possibility to reliably estimate ligand affinity using a single energy-minimized structure and implicit solvation models has not been explored in sufficient detail. Herein, we thoroughly investigate this hypothesis by comparing different methods for the generation of protein-ligand complexes and diverse methods for free energy prediction for their ability to correlate with experimental values. The methods were tested on a series of structurally diverse inhibitors of *Plasmodium falciparum* DHFR with known binding mode and measured affinities. The results showed that correlations between MM-PBSA or MM-GBSA binding free energies with experimental affinities were in most cases excellent. Importantly, we found that correlations obtained with the use of a single protein-ligand minimized structure and with implicit solvation models were similar to those obtained after averaging over multiple MD snapshots with explicit water molecules, with consequent save of computing time without loss of accuracy. When applied to a virtual screening experiment, such an approach proved to discriminate between true binders and decoy molecules and yielded significantly better enrichment curves.

© 2009 Wiley Periodicals, Inc. J Comput Chem 31: 797–810, 2010

Key words: binding free energy prediction; MM-PBSA; MM-GBSA; virtual screening; drug design

Introduction

Designing small molecules with desired biological activity has long been a desirable but challenging goal. In principle, a deep understanding of the biological structure and function of a target macromolecule, combined with appropriate computational tools able to model the interactions of potential ligands in the putative binding site, would allow the prediction of biologically active molecules before an experiment is made. In practice, the success of the computational prediction is often hampered by the complexity of biological macromolecules and by the partial ability of commonly used drug design tools to accurately predict active candidates. Therefore, advances in structure-based design approaches and techniques are essential for accelerating and improving the drug discovery and development process. One of the more popular approaches is molecular docking, which is able to screen and rank computationally a relatively large number of molecules into the active site of a target protein structure

in a reasonable amount of time.^{1–3} This is helpful to suggest binding modes and prioritize compounds for biological assays to discover new binders. Such docking methods evaluate the steric and electrostatic complementarities of ligands with the biological target in terms of energy scores derived from empirical equations. Although these methods have been significantly improved in the last years by including additional energy contributions and/or refining parameters in the scoring functions and by enhancing conformational sampling, we are still far from having docking tools that predict biologically active candidates in reasonable agreement with experimental data.^{2,4} Docking scores correlate only poorly with experimental affinities, as emphasized

Additional Supporting Information may be found in the online version of this article.

Correspondence to: G. Rastelli; e-mail: giulio.rastelli@unimore.it

by the many false positive hits in typical virtual screenings. As a matter of fact, it is becoming general opinion that docking results need to be postprocessed with more accurate methods before a selection of best candidates for biological testing is made. Above all, improvements should be made in scoring functions, for example, through better treatment of long-range electrostatics, desolvation of binding species, as well as entropic contributions. With various levels of accuracy and computational expense, computational approaches which can take into account such factors include free energy perturbation⁵ (FEP), thermodynamic integration⁶ (TI), linear response⁷ (LR), molecular mechanics Poisson-Boltzmann surface area (MM-PBSA), and molecular mechanics-generalized Born surface area (MM-GBSA).^{8,9} In particular, MM-PBSA and MM-GBSA combine molecular mechanics and continuum solvent models to estimate ligand binding affinities¹⁰ and are faster by several orders of magnitude than FEP or TI. However, generating and postprocessing an ensemble of representative protein-ligand snapshots from molecular dynamics (MD) trajectories using these methods still require computer-intensive calculations that are impractical for virtual screening applications, in which a large number of compounds need to be processed. If it were possible to reliably estimate ligand affinity with MM-PB(GB)SA using a single minimized protein-ligand structure instead of ensembles of snapshots derived from MD trajectories, this would undoubtedly constitute a valuable method to rapidly refine and rescore docking screening results. The validity of such an approach has been explored only in few studies.^{11–15} In particular, Kuhn et al validated the MM-PBSA approach on different biological systems by putting forward the idea of using single-minimized structure instead of MD trajectories.¹³ Beyond that study, so far, no systematic and extensive comparisons of the performance of different MD- and minimization-based methods for binding free energy calculation that take into account also the possibility to sample the structures with implicit solvent models has been carried out.

In this article, we thoroughly investigate this hypothesis by comparing the ability of different methods for the generation of protein-ligand complexes and different methods for the calculation of binding free energies to correlate with experimental values. Regarding structures, these were generated with 2 ns periodic boundary MD simulations in water as well as with four different energy-minimization methods in explicit or implicit water solvent. With reference to free energy calculations, these were performed with MM-PBSA and MM-GBSA methods implemented in Amber¹⁶ and Delphi.¹⁷

The performance of these methods was extensively tested on a series of twenty-two structurally diverse inhibitors of *Plasmodium falciparum* Dihydrofolate reductase (PfDHFR) with known binding mode and spanning seven orders of magnitude in biological activity.^{18–23} The results unequivocally show that the correlation between experimental and calculated binding free energies obtained from single energy-minimized structures is excellent and fully comparable to that obtained using significantly more time-consuming approaches like averaging over periodic boundary MD simulations in water. In addition, this methodology proved able to discriminate between true ligands and decoys very well. Therefore, it provides a fast and reliable way

to predict binding affinities of structurally unrelated molecules in correlation with experiments, and holds promise as a postprocessing method to more accurately score and rank hit compounds after docking screenings.

Computational Methods

The Ligands Data Set

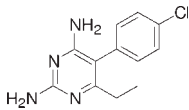
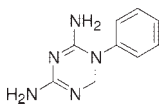
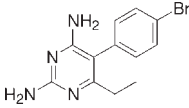
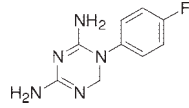
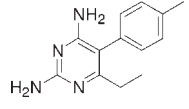
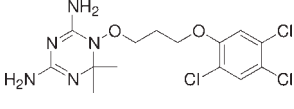
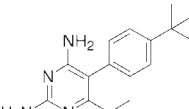
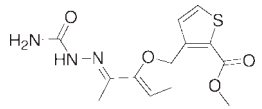
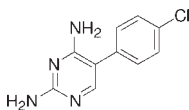
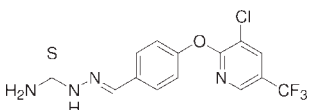
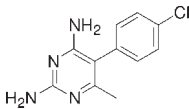
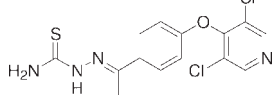
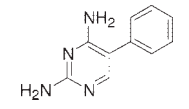
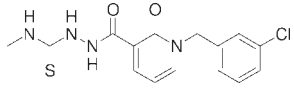
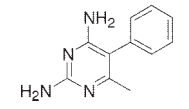
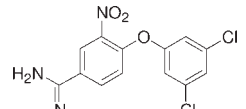
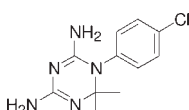
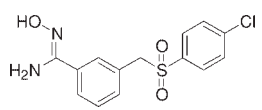
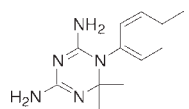
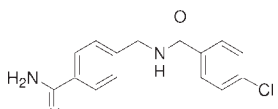
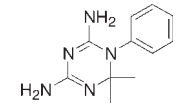
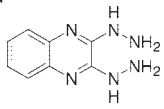
A set of twenty-two inhibitors of PfDHFR were selected for our study. The inhibitors were selected based on chemical diversity and significant variation in biological activity.^{18–23} All inhibition constants were determined in the same laboratory under the same experimental conditions, and K_i values spans seven orders of magnitude variation. The chemical diversity of the ligands data set is represented by different families of compounds such as pyrimidine-2,4-diamine compounds (Pyrimethamine analogs), [1,3,5]triazine-2,4-diamine analogs (Cycloguanil analogs), urea, thiourea, *N*-hydroxyamidine, and hydrazine inhibitors. Their structures are reported in Table 1 along with the measured K_i values. The pyrimethamine and cycloguanil derivatives **1a–1h** and **2a–2f** were treated as protonated at the nitrogen N1, as determined by NMR studies of DHFR-antifolate complexes^{24,25} and confirmed by crystal structures.¹⁸ All other inhibitors were neutral. The molecular mechanics (MM) parameters for these molecules (atom types and atomic charges) were assigned with the *antechamber* module of Amber 9.²⁶ In particular, the ligands were assigned generalized amber force field²⁷ (gaff) atom types and AM1-BCC²⁸ atomic charges calculated with the *divcon* program. Missing force-field parameters were assigned with the *parmcheck* utility.

The Protein-Ligand Complexes

The protein structure used in our investigation is the crystal structure of wild-type *Plasmodium falciparum* DHFR in complex with NADPH and the potent inhibitor WR99210¹⁸ (PDB code 1J3I), which is also included in our data set (compound **2f**). The structure was cut at residue Asn231, which corresponds to the DHFR domain of the bifunctional DHFR-TS structure. Of the dimer, unit B was chosen because it has fewer missing residues (Met1 and residues from Asp87 to Asn90). Met1 was built as in unit A. Residues from Asp87 to Asn90 were modeled with Modeller 8²⁹ by aligning the complete sequence of the protein with that of the X-ray structure 1J3I. Ten homology models were generated and the best model according to Prosa II³⁰ was saved; then, the coordinates of the four missing residues were inserted back into the original crystal structure in order to fill the gap.

The WR99210 ligand was removed from the structure, and the inhibitor structures were built and manually docked into the PfDHFR crystal structures. The pyrimidine-2,4-diamine ligands **1a–1h** and the [1,3,5]triazine-2,4-diamine ligands **2a–2f** were positioned according to the established binding modes of pyrimethamine, cycloguanil, and WR99210 as determined from crystal structure and modeling.^{18,21} In these structures, the protonated N1 and the N6-amino groups form bidentate hydrogen bonds with the conserved Asp54 residue of the active site. The initial orientations of the remaining inhibitors (**3a–3d**, **4a–4c**,

Table 1. Chemical Structures of the *Pf*DHFR Inhibitors Used for Free Energy Analyses and Their Experimental Inhibition Constants (K_i nM).

Compound	Structure	K_i (nM)	Compound	Structure	K_i (nM)
1a (Pyr)		0.6	2d		329
1b		0.3	2e		270
1c		0.4	2f (WR99210)		0.011
1d		0.6	3a		9700
1e		2.4	3b		5200
1f		2	3c		900
1g		41.9	3d		11,300
1h		10.9	4a		2400
2a (Cyc)		1.5	4b		102,000
2b		1.8	4c		77,200
2c		20	5a		3600

and **5a**) were taken from our previous work.²² All water molecules in the crystal structure were removed except for two conserved waters embedded into the protein (W1249 and W1250 in the original 1J3I crystal structure) and close to the important residue Asp54. These two water molecules were maintained for the

generation of structures with MM and MD in explicit and implicit solvent (see below), but removed for free energy calculations.

Hydrogens were added to the complexes using the internal coordinates of the Amber all-atom data base. All Lys and Arg residues were positively charged and Glu and Asp residues

negatively charged. The parameters of the NADPH cofactor were taken from our previous simulations.²¹

All calculations in this study were performed with Amber 9²⁶ suite of programs and the ff03³¹ force field for the protein and the gaff²⁷ force field for the ligands.

Structures Generated with Periodic Boundary MD Simulations in Water

Each protein-ligand complex was solvated in an octahedral box of TIP3P³² water molecules extending 5 Å outside the protein on all sides, resulting in more than 8200 waters added per complex. The electrostatics were treated with the particle-mesh Ewald^{33,34} method with a grid size of 72³ Å, a fourth-order B-spline interpolation, and a tolerance of 10⁻⁵. The *sander* module of Amber 9 was used for these simulations. The simulations used a residue-based cutoff of 8 Å, a time step of 2 fs and a constraint of bond lengths involving hydrogen atoms using the SHAKE³⁵ algorithm. The nonbonded pair list was updated every 50 fs. The solvated complexes were minimized with 2000 steps of conjugate gradient minimization and equilibrated with MD at 300 K as follows: firstly, 20 ps MD at constant volume with 1 kcal mol⁻¹ Å⁻² restraint on the protein, NADPH and the ligand were performed in order to equilibrate the temperature around 300 K without undesirable drifts of the structures; secondly, MD was continued under constant pressure conditions (1 atm) for 50 ps with 1 kcal mol⁻¹ Å⁻² restraint on the same atoms, followed by additional 20 ps MD with reduced (0.5 kcal mol⁻¹ Å⁻²) restraints. Then, the complexes were equilibrated for 2 ns without restraints. After equilibration, to assess free energy convergence, four consecutive production runs the length of 500 ps each were performed and coordinates were extracted every 5 ps, resulting in a hundred snapshots collected for each run and a total of four hundred snapshots collected for each compound. The coordinates were also averaged every 10 ps for visual inspection to make sure that the orientation of the inhibitor and the conformation of the protein did not change significantly during MD. Root mean squared deviation (RMSD) analyses were performed using the *ptraj* module of Amber 9. Mass-weighted RMSD values of protein-ligand complexes sampled by MD have been calculated relatively to the initial crystal structure.

Structures Generated with Molecular Mechanics Energy Minimization

In addition to the protein-ligand complexes generated with periodic boundary MD simulations in water, four other sets of structures were obtained with less time-consuming MM energy-minimization methods, namely (i) minimization in explicit water solvent with a octahedral box of water molecules, (ii) minimization in implicit solvent model with GBSA, (iii) minimization with a distance-dependent dielectric function, and (iv) minimization with a protocol consisting in distance-dependent dielectric minimization followed by molecular dynamics on the ligand and final reminimization.

Energy minimizations of the protein-ligand complexes were performed using the *sander* module of Amber 9. Regarding simulations in explicit solvent, each complex prepared as described

before was solvated in an octahedral box of TIP3P³² water molecules extending 5 Å outside the protein on all sides, resulting in more than 8200 waters added. The solvated complexes were minimized with 2000 steps of conjugate-gradient minimization without restraints, using the same conditions and parameters reported in the previous paragraph for periodic boundary simulations. 2000 steps of minimization proved sufficient to obtain stable free energy results in explicit and implicit water models because longer minimization tests resulted in similar binding free energies. Simulations in implicit solvent were performed with GBSA. At this purpose, 2000 steps of conjugate-gradient minimization were performed with the Tsui and Case parameters (igb = 1, viz GB^{HCT} model in the original paper^{36,37}), the surface area was computed and included in the solvation term, and a cutoff of 18 Å for nonbonded interactions was used. The third set of structures was obtained by 2000 steps of minimization with a distance-dependent dielectric constant $\epsilon = 4r$, and a cutoff of 12 Å. Finally, a multistep protocol consisting in 2000 steps of minimization on the entire complex, followed by 100 ps molecular dynamics in which the ligand alone was allowed to move, plus a final reminimization of the complex (2000 steps), was set up using distance-dependent dielectric constant conditions. When compared with single energy minimization, this protocol adds a molecular dynamics stage for the refinement of the orientation and conformation of the ligand which was specifically devised for relieving possibly incorrect assignments of ligand poses resulting from automated docking tools used in virtual screenings. MD on the ligand was performed at 300 K, with SHAKE³⁵ turned on for bonds involving hydrogens, allowing a time-step of 2.0 fs.

After each energy minimization, visual inspection of the complexes was performed to make sure that the protein and the ligand remained close to conformation observed in the crystal structures.

MM-PBSA and MM-GBSA Calculations

The values of the free energy of binding (ΔG_{bind}) of each inhibitor were calculated according to the equation

$$\Delta G_{\text{bind}} = G_{\text{com}} - G_{\text{rec}} - G_{\text{lig}} \quad (1)$$

where com, rec, and lig stand for complex, receptor, and ligand, respectively. The free energy of each of these was estimated as a sum of the four terms

$$G = \langle E_{\text{MM}} \rangle + \langle G_{\text{psolv}} \rangle + \langle G_{\text{npolv}} \rangle - T \langle S \rangle \quad (2)$$

where E_{MM} is the molecular mechanics energy of the molecule expressed as the sum of the internal energy of the molecule plus the electrostatics and van der Waals interactions, G_{psolv} is the polar contribution to the solvation energy of the molecule, G_{npolv} is the nonpolar solvation energy, T is the absolute temperature, and S is the entropy of the molecule.

The snapshots for MM-PBSA and MM-GBSA analyses were taken every 5 ps of each of the four 500 ps MD production runs, resulting in a total of four hundred snapshots per

compound analyzed. The energies were obtained using the *mm_pbsa* module of Amber 9. The internal, electrostatics and van der Waals energies were calculated with the *sander* module, with no cutoff for non-bonded interactions. The polar solvation free energies (G_{psolv}) were calculated by solving the Poisson Boltzmann (PB) equation with both Delphi¹⁷ and the Amber PBSA¹⁶ module, and by the generalized Born (GB) approach^{37,38} implemented in Amber 9. In Delphi, the grid spacing was set to 0.5 Å, with dielectric constants of 1 and 80 for the interior and exterior of the molecule, respectively. A cubic lattice with the largest dimension 80% filled by the longest dimension of the molecule was used. The dielectric boundary was defined using a 1.4 Å probe water on the atomic surface. The PB equation was solved using 1000 linear steps of finite difference. The Parse³⁹ set of radii was used for atoms of the molecule plus radii of 1.75 and 1.948 Å for F and Cl, respectively, which were taken from the Amber ff03 force field. Delphi calculations were performed using the same atomic charges used to minimize the complexes, that is, the standard Amber charges (ff03)³¹ for the protein and the AM1-BCC²⁸ charges for the cofactor and the ligands. In Amber PBSA, we used the Poisson-Boltzmann solver implemented in Amber 9 with the parameters already described for Delphi, with atomic cavity radii and atomic charges (ff03) taken from the topology files. Finally, energy estimations with GBSA were made with the Tsui and Case³⁶ parameters (igb = 1) and ff03 charges, with the same values of dielectric constants used for PB. All calculations used a solute internal dielectric constant equal to 1. However, to test the dependence of free energies on this parameter, calculations were also performed using increasing values of ϵ (2, 4, 10, and 25). These tests were carried out solely using the energy-minimized structures of the protein-ligand complexes in water, and the results were reported as Supporting Information.

The nonpolar solvation term (G_{npsolv}) was calculated from the solvent-accessible surface area (SASA)⁴⁰ using the equation

$$G_{\text{npsolv}} = \gamma \text{SASA} + b \quad (3)$$

where SASA was determined with the Molsurf⁴⁰ method using a probe radius of 1.4 Å. Parameters were $\gamma = 0.00542 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $b = 0.92 \text{ kcal mol}^{-1}$ to be used in combination with Delphi polar solvation energies and Parse radii, and $\gamma = 0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $b = 0 \text{ kcal mol}^{-1}$ to be used with Amber PB or GB polar solvation energies. The values of these parameters are well documented in literature.^{10–13}

Finally, the change in solute entropy during ligand association was estimated by a normal mode analysis⁴¹ of the vibration frequencies, calculated with the *nmode* module of Amber. Entropies were calculated using the entire protein-ligand complexes. Before normal mode calculations, the complexes were energy-minimized with a distance-dependent dielectric constant $\epsilon = 4r$ using a maximum of 40,000 steps and a target root-mean-square gradient of $10^{-4} \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. Tests performed with tighter convergence criteria (60,000 steps and a target rms of $10^{-5} \text{ kcal mol}^{-1} \text{ \AA}^{-1}$) made on a subset of compounds showed that $T\Delta S$ values did not differ for more than 0.2 kcal mol^{-1} , that is, energy minimization with the chosen settings was sufficient to

guarantee achievement of a true minimum (data not shown). Because of the high computational cost, we selected 20 regularly spaced snapshots along the 2 ns production trajectory for entropy calculations. Such a number of snapshots is in line with previous investigations.^{9,10,12,42} The entropic contribution to binding was calculated only for the 2 ns periodic boundary MD simulations in water.

Virtual Screening Protocol

The performance of MM-PBSA and MM-GBSA were further evaluated by setting up a virtual screening experiment. To assess the ability to discriminate between true ligands and decoys, we used the DHFR benchmarking data set developed by Shoichet and Irwin,⁴³ which has been downloaded from the directory of useful decoys (DUD) at blaster.docking.org/dud. The DHFR data set is composed of 201 known ligands and 7150 decoys with similar physical properties (e.g., molecular weight, hydrogen bond donors and acceptors, log P , etc.) but with chemical and topological differences, so that they are unlikely to be binders. The data set was docked into the DHFR crystal structure with AutoDock 4.⁴⁴ The Lamarckian genetic algorithm (LGA) was applied to model the interaction of the molecules with the protein. The docking area was defined using the AutoDock module ADT. The grid site was constrained to a 22.5 Å cubic space centered on the inhibitor WR99210, using a grid point spacing of 0.375 Å. For each molecule, 10 runs were carried out with 150 individuals in the first population and 2.5 million energy evaluations. The resulting protein-ligand and protein-decoy complexes were energy-minimized with Amber and further processed with MM-PBSA and MM-GBSA using our recently described automated workflow called BEAR (Binding Estimation After Refinement).⁴⁵ BEAR automatically and iteratively prepares ligand and receptor input files, assign Amber parameters and performs a multistep procedure consisting in energy minimization, molecular dynamics simulation in which only the ligand is allowed to move, and a final reminimization of the complex. Then, the single structure of each refined complex is used to estimate the free energy of binding through MM-PBSA and MM-GBSA algorithms, with parameter settings already described in this work. The lowest-energy orientation of the largest cluster found by AutoDock was used as input for the BEAR refinement and rescoring. After BEAR, the enrichment curves according to the AutoDock, MM-PBSA, and MM-GBSA scoring functions were calculated and compared.

Results and Discussion

In this investigation, we have explored different simulation protocols for their ability in predicting free energies of binding of a set of structurally diverse *Plasmodium falciparum* DHFR inhibitors with known binding mode and experimental activities (Table 1).^{18–23} Different methods for predicting binding free energies (Delphi, Amber PBSA, and Amber GBSA) as well as diverse methods for generating the structures of the protein-ligand complexes (periodic boundary MD simulations in water or less time-consuming MM energy minimizations in explicit

water, implicit water with GBSA, and distance-dependent dielectric constant $\epsilon = 4r$) have been explored. The accuracy and model dependency of these methods has been assessed and compared. Binding free energies of each inhibitor have been computed both with (ΔG_{bind}) and without ($\Delta G'_{\text{bind}}$) the inclusion of the entropic term:

$$\Delta G'_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} \quad (4)$$

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S_{\text{solute}} \quad (5)$$

where ΔE_{MM} is the molecular mechanics contribution expressed as the sum of the internal, electrostatic, and van der Waals contributions to binding *in vacuo*, ΔG_{solv} is the solvation free energy contribution to binding expressed as the sum of polar and nonpolar solvation free energies ($\Delta G_{\text{solv}} = \Delta G_{\text{psolv}} + \Delta G_{\text{npsolv}}$, respectively), and $T\Delta S_{\text{solute}}$ is the contribution of solute entropy to binding.

Free Energies of Binding Using Structures Generated with Periodic Boundary MD Simulations in Water

Binding free energy predictions were firstly performed on an ensemble of structures of protein-ligand complexes generated with 2 ns periodic boundary MD simulations in water. The average free energy results are given in Table 2. To assess convergence of the free energy results, the 2 ns production run was also divided into four consecutive 500 ps MD trajectories and free energies were calculated separately for each subset. Comparison of mean values and standard deviations of free energies obtained from each subset (Tables 1S–3S available as Supporting Information) suggests that free energies have reasonably converged. It is worth noting that, while Kuhn and other researchers pointed out that the use of longer MD trajectories may introduce additional uncertainties to the free energy estimation,¹³ we observe a consistency of the results among the four consecutive steps of 500 ps. Even if 2 ns MD simulations are rather short for proteins, this behavior can be explained on account of the well defined binding mode of the ligands in the DHFR binding pocket that imply few important structural rearrangements during molecular dynamics. Another reason is the choice of initial structures that, as described in the computational method section, are chosen in order to reproduce the crystallographic conformation that notably describe particularly favorable ligand-receptor interactions. As a matter of fact, the averaged root mean squared deviations (RMSD) between the structures sampled in the 2 ns MD in water and the X-ray structure (Table 4S in Supporting Information) show that the systems undergo dynamic changes as compared to the crystal structure. The RMSD values of the backbone atoms of the entire protein (overall fluctuations) and of the residues at a distance ≤ 8 Å from the inhibitor (binding site fluctuations) are ~ 2 and ~ 1 Å, respectively. Therefore, although MD indeed resulted in structural fluctuations of the complexes, which occur also in proximity of the binding site, these did not impair the overall binding mode and the most representative structures remained generally consistent with the crystallographic ones. Moreover, visualiza-

tion of the structures of the protein-ligand complexes along the trajectory confirmed that the ligands themselves did not undergo important conformational changes during MD. These findings are also consistent with the rather low standard deviations of $\Delta G'_{\text{bind}}$ values reported in Table 2. To further probe, the dependency of the free energy results on the simulation length, for two compounds (**1a** and the more flexible **2f** with six rotatable bonds) we also performed longer (10 ns) MD simulations. In comparison with the 2 ns results, the $\Delta G'_{\text{bind}}$ values obtained with the 10 ns simulations (included in Table 2 as footnotes) differed for only 1.6 (MM-PBSA) and 1.0 (MM-GBSA) kcal mol⁻¹ for compound **1a** and 2.6 (MM-PBSA) and 2.2 (MM-GBSA) kcal mol⁻¹ for compound **2f**.

As expected, the gas-phase interaction energies (Table 2, ΔE_{MM}) are scarcely correlated with the experimental free energies of binding ($r^2 = 0.56$) since the inclusion of additional terms such as desolvation energies are crucial for predicting binding affinities in correlation with experiments. The nonpolar solvation free energies calculated with Molsurf (Table 2) show that ΔG_{npsolv} values are always negative, that is, favorable to binding. Besides polar interactions established with the key residue Asp54 and backbone oxygen atoms of Ile14 and Ile164 of PfDHFR, hydrophobic interactions with Ala16, Phe58, Met55, and Leu46 also contribute favorably to binding affinity.^{18,21}

Polar solvation free energies (ΔG_{psolv}) have been calculated with Delphi, Amber PB and GB and the results are included in Table 2. According to all three methods, ΔG_{psolv} values are positive, that is, the molecules must pay a desolvation penalty upon protein binding. However, differences in absolute values are somewhat remarkable among the three methods, and a trend Delphi > Amber PB > Amber GB can be clearly identified (Table 2).

The estimated free energies of binding $\Delta G'_{\text{bind}}$ without the entropic contribution are reported in Table 2. Figures 1A–1C shows the regression plots of the experimental vs. computed ΔG of binding according to Delphi, Amber PBSA, and GBSA. The computed $\Delta G'_{\text{bind}}$ and the experimental ΔG_{expt} values were significantly correlated (Delphi $r^2 = 0.75$, $s = 1.4$, $F = 61$; Amber PBSA $r^2 = 0.93$, $s = 0.77$, $F = 248$; Amber GBSA $r^2 = 0.93$, $s = 0.72$, $F = 286$, Figs. 1A–1C). The use of two $\Delta G'_{\text{bind}}$ values calculated from the 10 ns instead of the 2 ns MD simulation (compounds **1a** and **2f**) had a negligible effect on the quality of the correlation. Therefore, despite the relatively large approximations inherent to the methodology, our converged free energy estimates reproduced the experimental trend very well. In reproducing binding free energies, Amber PBSA and GBSA performed significantly better than Delphi, which is often considered one of the most accurate methods for estimating polar solvations.¹² Given that Delphi is also significantly more time-consuming than Amber PBSA or GBSA, the results obtained on our training set highlight that the latter approaches are faster and even more accurate.

Entropic Contribution

The entropic contribution to binding ($T\Delta S$) was estimated by a normal mode analysis of the vibration frequencies performed after energy-minimization of twenty equally spaced snapshots

Table 2. Binding Free Energy Components (kcal mol⁻¹) and Standard Deviations Calculated with MM-PBSA (Delphi and Amber PBSA) and MM-GBSA (Amber GBSA) Using Structures Generated with 2 ns MD Simulations in Water

N	ΔG_{exp}^a	ΔE_{MM}^b	ΔG_{npolv}				ΔG_{psolv} GB ^e	ΔG_{psolv} PB ^e	ΔG_{psolv} Delphi ^e	ΔG_{psolv} Delphi ^f	ΔG_{psolv} PBSA ^f	ΔG_{psolv} GBSA ^f	$\Delta G'_{\text{bind}}$ Delphi ^g	$\Delta G'_{\text{bind}}$ PBSA ^g	$\Delta G'_{\text{bind}}$ GBSA ^g	$T\Delta S^h$
			Molsurf ^c	Molsurf ^c	Delphi ^d	Molsurf ^c										
1a	-12.7	-70.6 ± 3.4	-4.6 ± 0.1	-4.4 ± 0.1	53.3 ± 3.4	34.5 ± 3.1	20.7 ± 2.9	49.0 ± 3.4	29.9 ± 3.1	16.1 ± 2.9	-21.6 ± 2.1	-40.7 ± 1.6 ⁱ	-54.5 ± 1.3 ⁱ	-20.6 ± 6.8		
1b	-13.1	-70.9 ± 3.5	-4.7 ± 0.1	-4.4 ± 0.1	53.1 ± 3.0	34.2 ± 2.7	19.5 ± 2.9	48.7 ± 3.1	29.5 ± 2.7	14.9 ± 2.9	-22.2 ± 1.7	-41.3 ± 1.9	-56.0 ± 1.3	-21.6 ± 10.0		
1c	-12.9	-70.6 ± 3.3	-4.5 ± 0.1	-4.3 ± 0.1	54.9 ± 3.1	35.5 ± 2.8	21.2 ± 2.9	50.6 ± 3.1	31.0 ± 2.8	16.8 ± 2.9	-20.0 ± 1.8	-39.6 ± 1.7	-53.8 ± 1.3	-16.6 ± 11.3		
1d	-12.7	-68.2 ± 3.3	-5.2 ± 0.1	-4.8 ± 0.1	49.1 ± 3.2	29.5 ± 2.9	15.2 ± 2.7	44.3 ± 3.2	24.4 ± 2.9	10.1 ± 2.8	-23.8 ± 1.7	-43.8 ± 1.8	-58.1 ± 1.3	-21.6 ± 9.3		
1e	-11.8	-72.7 ± 2.8	-4.0 ± 0.1	-3.9 ± 0.1	59.6 ± 2.7	38.3 ± 2.3	27.9 ± 2.4	55.7 ± 2.7	34.3 ± 2.3	23.9 ± 2.4	-17.0 ± 1.6	-38.4 ± 1.5	-48.8 ± 1.2	-13.1 ± 11.0		
1f	-11.9	-64.4 ± 3.1	-4.3 ± 0.1	-4.2 ± 0.1	50.1 ± 3.0	31.5 ± 2.7	16.2 ± 2.6	46.0 ± 3.0	27.2 ± 2.7	11.9 ± 2.7	-18.5 ± 1.7	-37.2 ± 1.7	-52.5 ± 1.2	-19.4 ± 4.9		
1g	-10.1	-74.8 ± 3.6	-3.9 ± 0.1	-3.8 ± 0.1	63.4 ± 2.9	43.3 ± 2.5	32.2 ± 2.5	59.5 ± 2.9	39.5 ± 2.5	28.3 ± 2.5	-15.3 ± 1.8	-35.4 ± 1.6	-46.5 ± 1.1	-20.3 ± 9.3		
1h	-10.9	-68.5 ± 3.1	-4.0 ± 0.1	-3.9 ± 0.1	56.3 ± 3.1	37.6 ± 2.7	24.4 ± 2.6	52.4 ± 3.1	33.6 ± 2.7	20.4 ± 2.6	-16.1 ± 1.7	-34.8 ± 1.5	-48.1 ± 1.1	-17.8 ± 9.1		
2a	-12.1	-74.3 ± 3.6	-4.4 ± 0.1	-4.3 ± 0.1	59.3 ± 3.7	39.7 ± 3.2	25.6 ± 3.2	55.0 ± 3.7	35.4 ± 3.3	21.1 ± 3.2	-19.2 ± 1.9	-39.0 ± 1.8	-53.2 ± 1.3	-22.0 ± 6.8		
2b	-12.0	-67.6 ± 3.0	-4.5 ± 0.1	-4.3 ± 0.1	52.1 ± 3.1	33.3 ± 2.9	18.7 ± 2.5	47.8 ± 3.1	28.8 ± 2.7	14.2 ± 2.5	-19.9 ± 1.8	-38.8 ± 1.7	-53.5 ± 1.2	-21.9 ± 6.0		
2c	-10.6	-72.4 ± 3.6	-4.2 ± 0.1	-4.1 ± 0.1	58.7 ± 3.3	41.0 ± 2.9	27.5 ± 3.0	54.7 ± 3.3	36.8 ± 2.9	23.4 ± 3.0	-17.8 ± 1.6	-35.6 ± 1.7	-49.1 ± 1.3	-19.2 ± 8.7		
2d	-8.9	-58.7 ± 4.7	-3.8 ± 0.1	-3.8 ± 0.1	48.2 ± 3.9	29.9 ± 3.3	18.0 ± 3.3	44.4 ± 3.9	26.1 ± 3.3	14.3 ± 3.3	-14.3 ± 1.6	-32.6 ± 1.4	-44.4 ± 1.2	-13.7 ± 9.0		
2e	-9.0	-67.2 ± 3.8	-3.9 ± 0.1	-3.8 ± 0.1	55.9 ± 3.6	36.0 ± 3.1	23.9 ± 3.1	52.1 ± 3.6	32.1 ± 3.1	20.0 ± 3.1	-15.0 ± 1.8	-35.0 ± 1.9	-47.1 ± 1.5	-16.7 ± 6.8		
2f	-15.0	-90.5 ± 3.3	-6.1 ± 0.1	-5.5 ± 0.1	67.5 ± 3.1	47.7 ± 2.7	29.8 ± 2.7	62.0 ± 3.1	41.6 ± 2.7	23.8 ± 2.7	-28.5 ± 1.8	-49.0 ± 2.0 ^j	-66.8 ± 1.5 ^j	-22.7 ± 7.6		
3a	-6.9	-72.6 ± 1.9	-5.2 ± 0.1	-4.9 ± 0.1	65.8 ± 2.1	54.4 ± 1.8	40.7 ± 1.3	60.9 ± 2.1	49.2 ± 1.7	35.5 ± 1.3	-11.6 ± 1.6	-23.4 ± 2.0	-37.1 ± 1.3	-20.4 ± 5.2		
3b	-7.3	-71.3 ± 2.0	-5.5 ± 0.1	-5.1 ± 0.1	61.4 ± 2.1	50.2 ± 1.7	38.9 ± 1.1	56.4 ± 2.0	44.7 ± 1.6	33.4 ± 1.2	-14.9 ± 1.8	-26.6 ± 1.7	-37.9 ± 1.4	-21.6 ± 5.2		
3c	-8.3	-61.7 ± 1.6	-5.6 ± 0.1	-5.1 ± 0.1	49.7 ± 2.1	46.2 ± 1.7	33.3 ± 1.2	44.6 ± 2.1	40.6 ± 1.7	27.7 ± 1.2	-17.2 ± 2.0	-21.2 ± 1.8	-34.0 ± 1.3	-19.4 ± 9.0		
3d	-6.8	-56.8 ± 1.7	-5.4 ± 0.1	-5.0 ± 0.1	48.5 ± 1.6	40.8 ± 1.6	28.1 ± 1.1	43.6 ± 1.6	35.4 ± 1.6	22.7 ± 1.1	-13.3 ± 1.5	-21.5 ± 1.4	-34.1 ± 1.2	-23.7 ± 7.3		
4a	-7.7	-53.4 ± 1.4	-5.5 ± 0.1	-5.0 ± 0.1	40.0 ± 1.6	33.2 ± 1.7	22.7 ± 0.8	35.0 ± 1.6	27.8 ± 1.7	17.3 ± 0.8	-18.4 ± 1.6	-25.6 ± 1.8	-36.1 ± 1.1	-17.1 ± 9.0		
4b	-5.5	-46.1 ± 1.8	-5.3 ± 0.1	-4.9 ± 0.1	41.9 ± 1.7	38.2 ± 1.9	29.5 ± 1.1	37.0 ± 1.7	33.0 ± 1.9	24.3 ± 1.1	-9.1 ± 1.4	-13.1 ± 2.0	-21.8 ± 1.1	-19.9 ± 10.4		
4c	-5.6	-44.5 ± 1.7	-5.2 ± 0.1	-4.9 ± 0.1	35.6 ± 1.7	31.6 ± 1.7	24.2 ± 1.1	30.7 ± 1.7	26.4 ± 1.7	18.9 ± 1.1	-13.8 ± 1.4	-18.1 ± 1.4	-25.6 ± 1.1	-25.0 ± 9.2		
5a	-7.5	-34.7 ± 3.7	-3.7 ± 0.1	-3.7 ± 0.1	29.4 ± 3.3	14.2 ± 2.9	4.4 ± 3.0	25.7 ± 3.3	10.6 ± 2.9	0.7 ± 3.0	-9.1 ± 1.7	-24.2 ± 1.8	-34.0 ± 1.5	-16.8 ± 5.3		

^aExperimental free energies of binding (kcal mol⁻¹) according to $\Delta G = -RT\ln(1/K_i)$.

^bMM gas-phase interaction energies (kcal mol⁻¹).

^cNonpolar solvation energies (kcal mol⁻¹) according to $G_{\text{npolv}} = \gamma \text{SASA} + b$, where $\gamma = 0.0072$ kcal mol⁻¹ Å⁻² and $b = 0$ kcal mol⁻¹ for use in Amber PBSA and GBSA.

^dNonpolar solvation energies (kcal mol⁻¹) with $\gamma = 0.00542$ kcal mol⁻¹ Å⁻² and $b = 0.92$ kcal mol⁻¹ for use in Delphi.

^ePolar solvation free energies according to Delphi, Amber PB and GB.

^fTotal solvation free energies ($\Delta G_{\text{npolv}} + \Delta G_{\text{psolv}}$) according to Delphi, Amber PB and GB.

^gEstimated free energies of binding $\Delta G'_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{psolv}}$ according to Delphi, Amber PBSA, and GBSA. The entropic contribution $T\Delta S$ was not included.

^hEntropic contributions $T\Delta S$ ($T = 300$ K) evaluated from normal mode analyses.

ⁱ $\Delta G'_{\text{bind}}$ values obtained with the longer (10 ns) MD simulation are -39.1 and -53.5 kcal mol⁻¹ for MM-PBSA and MM-GBSA, respectively.

^j $\Delta G'_{\text{bind}}$ values obtained with the longer (10 ns) MD simulation are -51.6 and -69.0 kcal mol⁻¹ for MM-PBSA and MM-GBSA, respectively.

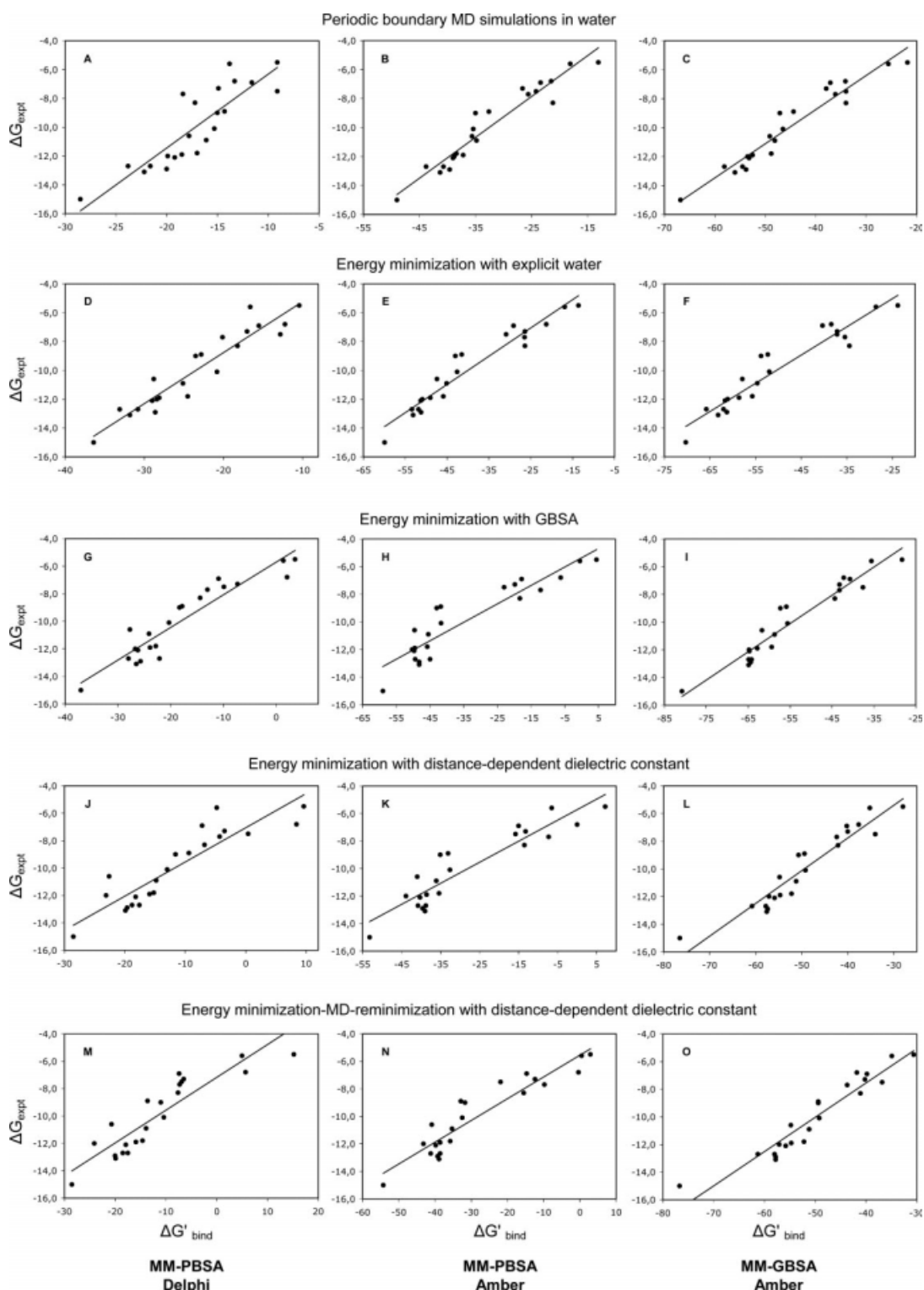


Figure 1. Correlation between experimental and estimated free energies of binding (kcal mol^{-1}). Free energies were obtained after 2 ns periodic boundary MD simulations in water (A–C), minimization with explicit water (D–F), minimization with GBSA (G–I), minimization with distance-dependent dielectric constant $\epsilon = 4r$ (J–L), and a multistep minimization-MD-reminimization protocol with distance-dependent dielectric constant (M–O). Solvation free energies were evaluated with Delphi (plots A, D, G, J, M), Amber PBSA (plots B, E, H, K, N), and Amber GBSA (plots C, F, I, L, O).

taken along the 2 ns MD production described above, and the results are included in Table 2. $T\Delta S$ values are dominated by the translational and rotational contributions (~ -23 kcal mol $^{-1}$), whereas their variation is caused by the vibrational contribution. Moreover, because $T\Delta S$ values are significantly negative, their inclusion into $\Delta G'_{\text{bind}}$ to account for entropic effects on binding ($\Delta G_{\text{bind}} = \Delta G'_{\text{bind}} - T\Delta S$) has a high impact on the computation of absolute free energies. With the $T\Delta S$ term, $\Delta G'_{\text{bind}}$ values become closer to the experimental ΔG_{expt} values when compared with $\Delta G'_{\text{bind}}$, but differences between computed and experimental values remain generally high (on average 12 kcal mol $^{-1}$ for Delphi, 7 kcal mol $^{-1}$ for Amber PBSA and 16 kcal mol $^{-1}$ for Amber GBSA) and, therefore, none of the three methods can be regarded as able to quantitatively reproduce absolute binding free energies under the chosen simulation conditions. In addition, we found that the inclusion of the entropic term significantly reduced the correlation with the experimental affinities, with reduction of r^2 values compared with the plots shown in Figures 1A–1C and detection of at least two or three outliers per plot (data not shown). We may therefore conclude that the entropic contribution, as estimated in the reported conditions, does not help in achieving quantitative predictions of absolute binding free energies and it also worsens the agreement with the ranking of binding affinities observed experimentally, which was already very good. Furthermore, $T\Delta S$ is by far the contribution with the largest standard deviations (~ 8 kcal mol $^{-1}$ as opposed to ~ 3 kcal mol $^{-1}$ for ΔE_{MM} , ~ 2 kcal mol $^{-1}$ for ΔG_{solv} and less than 2 kcal mol $^{-1}$ for $\Delta G'_{\text{bind}}$, Table 2), highlighting that the inclusion of such term in ΔG also significantly increases uncertainty. Indeed, it is important to note that such behavior is likely to be an effect of not sampling adequately the configuration space of the protein in entropy calculations. However, owing to the huge computational resources required by normal mode analysis simulations, a deeper sampling of the configuration space would be definitely impractical to postprocess docking screening results. Therefore, we can conclude that the inclusion of inaccurate entropy estimates in ΔG may harm more than completely neglecting this term. Moreover, uncertainties may come also from parts of the protein that are well beyond the binding site of the ligand, and proteins exhibiting much more flexibility than pfDHFR may exhibit even bigger discrepancies. Obviously, these conclusions apply to our training set and simulation conditions. In other applications, it has been shown that enthalpy/entropy compensation⁴⁶ is important to understand and reliably predict binding energies, and that the omission of entropies leads to an overestimate of binding affinities and degrades the correlation with experimental binding energies.^{47,48} Thus, incorporation of solute entropy in binding affinity calculations remains an important but daunting challenge for which improvements are needed in future research.⁴⁹

Free Energies Using Single Energy-Minimized Structures

Although free energy predictions are usually made on large collections of equilibrated structures sampled during MD in water as performed above, we were interested in testing whether a single minimized structure was a reasonable approximation for rapidly estimating ligand binding energies. This approach is a

priori a simplification because free energies reflect all configurational states populated by the system at a given temperature. However, generating a single minimized structure of a protein-ligand complex and estimating the ΔG of binding from one structure is fast, and therefore suitable for high-throughput virtual screening studies. Moreover, previous studies show that the use of a single bound conformation may give promising results.¹³ To thoroughly test this hypothesis, we have performed binding free energy predictions using energy-minimized structures generated with four different methods. Table 3 reports the free energy results obtained using the various minimization and free energy methods. Again, none of the methods could reproduce the absolute ΔG_{expt} values quantitatively. However, correlations between computed $\Delta G'_{\text{bind}}$ and experimental ΔG_{expt} values were still very good (Figure 1). Importantly, with single energy-minimized structures, we achieved correlations with statistical quality in all similar to those obtained using the more time-consuming approach based on generation and analysis of MD ensemble of structures in water. Despite the approximation of retaining a single protein-ligand structure, we may explain this by considering that this conformation significantly contributes and thus dominates the Boltzmann-averaged potentials for the free energy estimation. This is particularly true when the bound conformation of the ligands corresponds to a particularly stable conformation of the free ligand. Therefore, in such situation the ligand is already arranged to achieve a good binding with the protein and the energetic penalties coming from the conformational strains and the entropic contribution play a minor role.⁴⁷ Figures 1D–1F shows the regression plots obtained with structures minimized in explicit water solvent (Delphi $r^2 = 0.90$, $s = 0.90$, $F = 172$; Amber PBSA $r^2 = 0.93$, $s = 0.75$, $F = 262$; Amber GBSA $r^2 = 0.91$, $s = 0.83$, $F = 207$). Structures generated after minimization with the implicit solvent model GBSA still yielded significant relationships between $\Delta G'_{\text{bind}}$ and ΔG_{expt} values, especially when free energies were evaluated with Amber GBSA (Delphi $r^2 = 0.88$, $s = 0.98$, $F = 144$; Amber PBSA $r^2 = 0.85$, $s = 1.09$, $F = 113$; Amber GBSA $r^2 = 0.93$, $s = 0.75$, $F = 258$, Figs. 1G–1I). When compared with energy-minimization in water, structures minimized with GBSA gave slightly lower regression coefficients when free energies of binding were evaluated with Delphi (0.88 vs. 0.90) or Amber PBSA (0.85 vs. 0.93), and these were only slightly better when evaluated with Amber GBSA (0.93 vs. 0.91). Notably, GBSA minimization was, on average, seven times slower than minimization in water (see next paragraph for a description of simulation times) and did not give better results. Free energy evaluations were then carried out on structures minimized with a distance-dependent dielectric constant $\epsilon = 4r$, a fast minimization protocol of protein-ligand complexes. The best correlation between $\Delta G'_{\text{bind}}$ and ΔG_{expt} values (Figs. 1J–1L) was observed with Amber GBSA ($r^2 = 0.91$, $s = 0.86$, $F = 191$), followed by Amber PBSA ($r^2 = 0.85$, $s = 1.08$, $F = 115$) and Delphi ($r^2 = 0.81$, $s = 1.21$, $F = 87$). Importantly, such relationships are similar to those obtained using explicit water or GBSA minimization, suggesting that free energy predictions using these structures still lead to significant correlation with experiment. Notably, minimization with distance-dependent dielectric constant is, on average, four times faster than minimization in

Table 3. Free Energies (kcal mol⁻¹) According to Delphi, Amber PBSA, and GBSA Calculations Performed on Structures Generated with Minimization in Explicit Water Solvent, GBSA, Distance-Dependent Dielectric Function, and Minimization-MD-re-minimization with Distance-Dependent Dielectric Function.

n°	ΔG_{expt}	ΔE_{MM}	$\Delta G_{\text{solv}}^{\text{Delphi}}$	$\Delta G_{\text{solv}}^{\text{PBSA}}$	$\Delta G_{\text{solv}}^{\text{GBSA}}$	$\Delta G_{\text{bind}}^{\text{Delphi}}$	$\Delta G_{\text{bind}}^{\text{PBSA}}$	$\Delta G_{\text{bind}}^{\text{GBSA}}$
Minimization in explicit water solvent								
1a	-12.7	-99.4	68.7	47.6	37.5	-30.8	-51.8	-62.0
1b	-13.1	-100.6	68.8	47.5	37.4	-31.8	-53.1	-63.2
1c	-12.9	-99.4	70.8	48.2	38.1	-28.6	-51.2	-61.3
1d	-12.7	-104.4	71.3	51.1	38.7	-33.1	-53.4	-65.8
1e	-11.8	-96.2	71.7	50.3	40.5	-24.5	-45.9	-55.7
1f	-11.9	-95.9	67.9	46.9	37.3	-28.1	-49.0	-58.6
1g	-10.1	-93.3	72.5	50.7	41.3	-20.8	-42.6	-52.0
1h	-10.9	-92.9	67.8	47.9	38.3	-25.1	-45.1	-54.6
2a	-12.1	-100.2	71.1	48.9	38.4	-29.0	-51.3	-61.7
2b	-12.0	-99.6	71.2	48.7	38.5	-28.4	-50.9	-61.1
2c	-10.6	-96.3	67.5	49.0	38.4	-28.8	-47.4	-57.9
2d	-8.9	-94.1	71.3	52.5	41.8	-22.8	-41.5	-52.3
2e	-9.0	-95.6	72.1	52.6	41.8	-23.5	-43.0	-53.8
2f	-15.0	-107.7	71.3	47.7	37.4	-36.4	-59.9	-70.3
3a	-6.9	-74.9	59.5	45.8	34.6	-15.5	-29.1	-40.3
3b	-7.3	-68.5	51.5	42.1	31.4	-17.0	-26.4	-37.1
3c	-8.3	-60.6	42.4	34.3	26.2	-18.2	-26.4	-34.4
3d	-6.8	-67.5	55.4	46.2	29.2	-12.2	-21.3	-38.4
4a	-7.7	-52.2	32.1	25.7	16.8	-20.1	-26.5	-35.4
4b	-5.5	-48.3	37.9	34.7	24.5	-10.4	-13.6	-23.8
4c	-5.6	-47.5	31.0	30.7	18.9	-16.6	-16.9	-28.6
5a	-7.5	-71.3	58.6	40.4	34.3	-12.8	-30.9	-37.1
Minimization with GBSA								
1a	-12.7	-87.0	59.0	37.7	22.2	-28.0	-49.4	-64.9
1b	-13.1	-86.9	60.4	38.7	22.0	-26.5	-48.2	-64.9
1c	-12.9	-88.0	62.3	39.8	23.5	-25.7	-48.2	-64.4
1d	-12.7	-84.8	62.7	39.8	20.6	-22.1	-45.0	-64.2
1e	-11.8	-86.3	63.5	40.5	26.9	-22.8	-45.8	-59.4
1f	-11.9	-86.8	62.9	37.2	24.0	-23.9	-49.5	-62.8
1g	-10.1	-82.7	62.4	41.0	27.0	-20.3	-41.7	-55.6
1h	-10.9	-82.8	58.7	37.3	24.1	-24.1	-45.5	-58.7
2a	-12.1	-88.4	62.2	38.7	23.7	-26.2	-49.7	-64.7
2b	-12.0	-90.0	63.3	39.6	25.2	-26.7	-50.3	-64.8
2c	-10.6	-85.7	58.0	36.1	24.0	-27.7	-49.6	-61.7
2d	-8.9	-83.6	65.8	41.8	27.7	-17.8	-41.8	-55.9
2e	-9.0	-84.8	66.5	41.8	27.5	-18.3	-43.0	-57.3
2f	-15.0	-105.2	68.3	46.3	24.4	-37.0	-59.0	-80.8
3a	-6.9	-73.1	62.2	55.4	32.5	-10.9	-17.8	-40.7
3b	-7.3	-76.1	68.8	56.3	32.9	-7.3	-19.8	-43.2
3c	-8.3	-69.9	55.5	51.5	25.6	-14.4	-18.3	-44.3
3d	-6.8	-63.0	65.0	56.8	20.8	2.1	-6.2	-42.2
4a	-7.7	-60.4	47.4	48.2	17.2	-13.0	-12.2	-43.2
4b	-5.5	-54.4	58.0	58.9	26.1	3.6	4.4	-28.3
4c	-5.6	-56.2	57.6	55.7	20.6	1.4	-0.5	-35.6
5a	-7.5	-45.8	35.9	22.9	8.2	-9.9	-23.0	-37.6
Minimization with distance-dependent dielectric constant $4r$								
1a	-12.7	-73.9	55.0	35.1	16.1	-18.8	-38.8	-57.8
1b	-13.1	-73.1	53.2	34.0	15.5	-19.9	-39.0	-57.6
1c	-12.9	-76.2	56.6	36.6	18.7	-19.6	-39.6	-57.4
1d	-12.7	-78.1	60.6	37.4	17.3	-17.6	-40.8	-60.8
1e	-11.8	-72.9	57.7	37.5	20.7	-15.2	-35.4	-52.2
1f	-11.9	-72.5	56.6	33.9	17.8	-15.9	-38.6	-54.7
1g	-10.1	-71.3	58.4	38.8	22.2	-13.0	-32.6	-49.2
1h	-10.9	-70.2	55.4	34.1	19.0	-14.8	-36.1	-51.2
2a	-12.1	-73.2	55.0	32.9	17.3	-18.2	-40.3	-55.9

(Continued)

Table 3. (Continued)

n°	ΔG_{expt}	ΔE_{MM}	$\Delta G_{\text{solv}}^{\text{Delphi}}$	$\Delta G_{\text{solv}}^{\text{PBSA}}$	$\Delta G_{\text{solv}}^{\text{GBSA}}$	$\Delta G_{\text{bind}}^{\text{Delphi}}$	$\Delta G_{\text{bind}}^{\text{PBSA}}$	$\Delta G_{\text{bind}}^{\text{GBSA}}$
2b	-12.0	-78.1	55.0	34.2	21.1	-23.1	-43.9	-57.1
2c	-10.6	-74.7	52.1	33.8	19.9	-22.6	-41.0	-54.8
2d	-8.9	-73.2	63.8	40.1	23.8	-9.4	-33.1	-49.4
2e	-9.0	-74.2	62.7	39.2	23.5	-11.6	-35.1	-50.7
2f	-15.0	-97.5	69.0	44.3	21.0	-28.5	-53.2	-76.5
3a	-6.9	-73.3	66.1	58.3	33.1	-7.2	-15.0	-40.2
3b	-7.3	-72.0	68.5	58.9	32.0	-3.5	-13.2	-40.0
3c	-8.3	-67.5	60.6	54.0	25.3	-6.8	-13.5	-42.1
3d	-6.8	-59.3	67.7	59.3	21.8	8.4	0.0	-37.6
4a	-7.7	-59.9	55.5	52.6	17.4	-4.3	-7.3	-42.4
4b	-5.5	-60.0	69.6	67.2	32.0	9.6	7.2	-28.0
4c	-5.6	-58.0	53.2	51.4	22.8	-4.8	-6.5	-35.2
5a	-7.5	-42.6	43.1	26.8	8.7	0.4	-15.8	-34.0
Minimization/MD/re-minimization with distance-dependent dielectric constant $4r$								
1a	-12.7	-75.2	56.8	36.6	17.3	-18.5	-38.6	-58.0
1b	-13.1	-74.5	54.7	35.6	16.7	-19.9	-38.9	-57.8
1c	-12.9	-78.2	58.2	38.9	20.4	-20.0	-39.3	-57.8
1d	-12.7	-80.6	63.1	39.5	19.3	-17.5	-41.2	-61.3
1e	-11.8	-73.4	58.8	37.6	21.2	-14.6	-35.8	-52.2
1f	-11.9	-73.4	57.5	34.8	18.7	-15.9	-38.6	-54.7
1g	-10.1	-72.0	61.6	39.5	22.8	-10.4	-32.5	-49.2
1h	-10.9	-70.8	56.9	35.5	19.6	-13.9	-35.3	-51.2
2a	-12.1	-73.5	55.6	33.7	17.7	-17.9	-39.8	-55.8
2b	-12.0	-78.8	54.7	35.6	21.7	-24.1	-43.2	-57.1
2c	-10.6	-75.2	54.6	34.3	20.5	-20.7	-40.9	-54.8
2d	-8.9	-72.8	59.2	39.9	23.3	-13.6	-32.9	-49.4
2e	-9.0	-71.0	60.1	39.4	21.7	-11.0	-31.7	-49.4
2f	-15.0	-98.2	69.8	43.9	21.5	-28.5	-54.3	-76.7
3a	-6.9	-73.3	65.9	58.6	33.5	-7.4	-14.7	-39.8
3b	-7.3	-72.3	65.8	59.9	32.1	-6.5	-12.4	-40.2
3c	-8.3	-65.6	58.0	50.0	24.5	-7.6	-15.5	-41.1
3d	-6.8	-59.6	65.3	59.3	17.9	5.7	-0.4	-41.8
4a	-7.7	-61.5	54.2	51.7	17.8	-7.3	-9.8	-43.7
4b	-5.5	-64.0	79.2	66.9	33.4	15.2	2.9	-30.6
4c	-5.6	-67.9	72.9	68.4	32.9	5.0	0.5	-34.9
5a	-7.5	-48.3	45.3	26.4	11.5	-6.9	-21.9	-36.8

water. For all methods used to generate structures, binding free energies estimated with Amber PBSA and GBSA resulted in better correlations to experiments if compared to the Delphi method. Moreover, although binding free energies estimated with GBSA proved to be the less sensible to the method used to generate structures, PBSA gave slightly lower r^2 values when structures were generated without explicit water.

Finally, we tested the effect of introducing a short molecular dynamics step for the refinement of the orientations of the ligand in the complex. Accordingly, complexes were minimized with distance-dependent dielectric function, MD was performed on the ligand alone, and the structures of the complexes were re-minimized after MD. The motivation behind this is that an intermediate MD step can help to relieve possible incorrect ligand conformations and/or orientations assigned by automated docking tools. In fact, while the inhibitors in our training set have known binding modes, typical applications in virtual screening

should deal with the fact that docking tools in some cases fail to predict "correct" solutions.⁴ Such an approach has been implemented in an automated workflow named BEAR (Binding Estimation After Refinement),⁴⁵ that automatically and iteratively prepares ligand, receptor, and complex topologies, performs structural refinement with Amber and evaluates binding free energies with MM-PBSA and MM-GBSA, using the settings described in this work. Free energies of binding estimated using these last structures were remarkably similar to those obtained with simple energy minimization (Table 3). Regression plots are given in Figures 1M–1O (Delphi $r^2 = 0.81$, $s = 1.23$, $F = 84$; Amber PBSA $r^2 = 0.87$, $s = 0.99$, $F = 140$; Amber GBSA $r^2 = 0.91$, $s = 0.84$, $F = 203$). This result was expected considering that our ligands already have correct initial orientations and conformations, but it is important to ascertain that the application of the intermediate MD refinement step does not alter free energy predictions. It is worth noting that in this work we

did not test the effect of starting MD calculations from ligand orientations different from the crystallographic ones. The assessment of the “correct” binding mode of a ligand in a protein pocket was out of the scope of this work, and further validations regarding conformational refinement of docking poses are in progress. For this reason our choice of using crystal structures as initial conformation constitute a reasonable and convenient way to exploit structures that are intrinsically among the most representative for assessing free energies of ligand-receptor complexes.

Some final considerations should be done about the robustness of the correlations that we obtained. As we have seen, the calculations presented in this study aim at correlating calculated binding free energies with experimental free energies obtained from inhibition constants. Obviously, predicting absolute or relative free energies of binding in quantitative agreement with experiment is a much more complicated task, but such effort becomes far less relevant when using the method as a scoring function, that is, when the problem is typically to rank ligands and predict better binders. The fact that ΔG values calculated with MM-PBSA may be far from experimental values on an absolute scale is a known issue that has already been reported in many other studies, in which free energies turned out to be usually overestimated. One way to bring free energies within the range of experimental data is to include entropic effects as already discussed in the paragraph describing the entropic contribution. Another possibility suggested in some applications is to scale down binding energies by increasing the internal dielectric constant.^{48,50–54} However, the validity and generality of such an approach is still matter of debate, because the dielectric constant of proteins depends in a critical way on its definition and on the specific computational models used, and the precise physical meaning of the dielectric constant in these continuum models is ambiguous and system-dependent.^{48,50,51,53} To better shed light on these issues, we have carried out further calculations with higher values of internal ϵ . The results reported in Table 5S (Supporting Information) show that increasing the value of ϵ from 1 to 2, 4, 10, or 25 not only did not improve, but sometimes worsened the agreement with experimental ΔG values. In addition to this, the application of higher dielectric constants resulted in a remarkable decrease of correlation between computed and experimental binding free energies that we attributed to the different electrostatic dependence of complexes whose ligands have different net charges. In fact, free energies tended to scale uniformly with respect to ϵ only for ligands with the same net charge, so that mixing positively charged and neutral ligand in the same dataset was detrimental for the correlation. This observation is important because, to our knowledge, the dependency of free energies on the internal dielectric constants with sets of ligands with different charges was not explored in previous studies. In light of these findings and given that the main purpose of the study was to validate a method for ranking and not for predicting absolute free energies, we may conclude that the optimal usage of the dielectric constant is for ϵ equal to 1.

Comparison of Simulation Times

While generation of a 2 ns MD trajectory in water required approximately 5 days per molecule on a single core of a 2.4

GHz AMD Opteron processor, single energy minimization in explicit water or implicit water with GBSA required only 12 and 77 min, respectively. Notably, energy minimization with distance-dependent dielectric constant $\epsilon = 4r$ required only 3 min, and minimizations plus intermediate MD refinement of the ligand took less than 8 min to complete. Regarding free energy calculations, Delphi took around 4 min per structure, whereas Amber PBSA and GBSA were significantly faster (1 and 0.2 min, respectively).

These numbers unequivocally show that simple energy minimizations with distance-dependent dielectric constant or explicit water coupled with Amber PBSA or GBSA free energy predictions are largely compatible with the typical size of virtual screening deployments. In fact, the possibility of running these calculations on large-scale computing facilities such as HPC clusters or grid platforms enable the processing of several thousands of compounds in a day.

Performance of MM-PBSA and MM-GBSA in a Virtual Screening Experiment

Virtual screening is an important computational approach for the identification of biologically active compounds. The screening is typically performed with molecular docking, which assesses the orientation of molecules in the target active site of a macromolecule and uses scoring functions to predict the strength of their association. Despite recent improvements, one major drawback of these methods is that scoring functions still struggle in estimating ligand binding energies in agreement with experiment, with the consequence that false positive and false negative hits largely populate the ranked list of compounds. In order to test the performance of our single-structure approach combined with the use of MM-PBSA and MM-GBSA scoring functions, we have carried out a virtual screening experiment using the DHFR benchmarking data set put forward by Shoichet and Irwin,⁴³ in their directory of useful decoys (DUD). The data set, composed of 201 known DHFR ligands and 7150 decoy molecules that resemble the physical properties of the ligands but are chemically and topologically distinct from them, was docked into the DHFR crystal structure using AutoDock. Then, the resulting protein-ligand and protein-decoy complexes were further processed using our methodology.

The performance of the virtual screening was evaluated by its capacity to enrich the small number of known active compounds in the top ranks of the screen from among a much greater number of decoy molecules present in the database. The enrichment curves obtained with Autodock and MM-PBSA and MM-GBSA are shown in Figure 2, in which the higher the percentage of known ligands found at a given percentage of the ranked database, the better the enrichment performance of the virtual screen. When compared with Autodock, we found that MM-PBSA and MM-GBSA yielded strikingly better enrichments. Therefore, the application of our refinement and rescoring procedure resulted in a significant improvement of the ranking of known inhibitors. It is worth noting that most of the active compounds in the DHFR data set contain folate, pteridine, pyrimidine, and triazine scaffolds with known general binding mode (a protonated heterocycle interacting with D54, I164, and

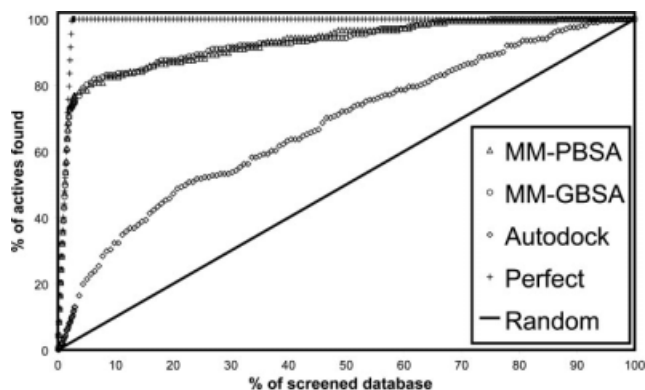


Figure 2. Enrichment plot showing the percentage of ranked database plotted against the percentage of active ligands found at any given percentage of ranked database. The plot compares the performance of the original AutoDock scores with those obtained with MM-PBSA and MM-GBSA rescoring.

A14 residues).^{18,21–23} Although AutoDock was able to predict the binding orientation of these compounds in agreement with the available crystal structures, we remark significant difficulties in ranking them amongst the best compounds. As some relatively small fraction of the best-ranked compounds are typically selected for biological assays, according to the original AutoDock scores alone the known active compounds would not have been selected for such purpose.

Conclusions

We have tested the performance of different theoretical approaches for their ability in correlating experimental and calculated binding free energies. The various methods were devised by combining different techniques for the generation of structures and different algorithms for the estimation of binding free energies, and were validated on a training set of twenty-two diverse *Plasmodium falciparum* DHFR inhibitors with known binding mode and measured affinities.

The results showed that the MM-PBSA and MM-GBSA methods achieve promising accuracy and rank the inhibitors in striking correlation with experimental binding affinities. Importantly, we found that correlations obtained with the use of a single protein-ligand minimized structure were in all similar to those obtained after averaging over multiple MD snapshots, with consequent save of computing time without loss of accuracy. This finding is particularly relevant in the context of virtual screenings, in which the large number of molecules typically processed precludes the use of time-consuming procedures. Therefore, our findings based on single minimized structures qualify this approach as an attractive opportunity for postprocessing molecular docking results and more accurately rank-ordering potential ligands before biological evaluation. Remarkably, such an approach can handle structurally dissimilar ligands and provides results at a fraction of the computational cost required to generate and analyze molecular dynamics trajectories

in explicit water molecules. Therefore, after further validation on different proteins and ligand datasets, such approach may have great interest for high throughput virtual screening. A benchmarking screening performed on DHFR showed that our methodology was indeed able to discriminate between true ligands and decoy molecules with similar physical properties. Notably, the enrichment curves obtained with our approach showed a marked improvement with respect to that obtained with a standard docking scoring function, resulting in better enrichment of known active compounds and significant improvement of the ranked lists.

Although the aim of this study was to shed light on the ability and usefulness of different algorithms (MM-PBSA and MM-GBSA) and different computational protocols (sampling of conformations and use of explicit/implicit solvent models) to rapidly assess the free energy of binding in protein-ligand systems such as pfDHFR, additional investigation on other biological targets that show, for instance, higher conformational flexibility are under way to further validate these techniques towards their use in the context of virtual screening campaigns. Finally, it is worth noting that while the present study focused on protein-ligand interactions to validate a fast and accurate method to predict thermodynamics of binding *in silico*, broader applications of this methodology can be envisaged for other types of host-guest intermolecular complexes, such as those involved in supramolecular assemblies of molecules, molecular recognition, chiral separation, and others.

Acknowledgments

We wish to thank Corinne Portioli and Marco Daniele Parenti for their assistance in performing the virtual screening experiment on DHFR.

References

1. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat Rev Drug Discov* 2004, 3, 935.
2. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. *J Med Chem* 2006, 49, 5851.
3. McInnes, C. *Curr Opin Chem Biol* 2007, 11, 494.
4. 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. *J Med Chem* 2006, 49, 5912.
5. Kollman, P. A. *Chem Rev* 1993, 93, 2395.
6. Lybrand, T.; Mccammon, J. A.; Wipff, G. *Proc Natl Acad Sci USA* 1986, 83, 833.
7. Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng* 1994, 7, 385.
8. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc Chem Res* 2000, 33, 889.
9. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J Am Chem Soc* 2001, 123, 5221.
10. Pearlman, D. A. *J Med Chem* 2005, 48, 7796.
11. Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J Med Chem* 2006, 49, 4805.
12. Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. *J Med Chem* 2006, 49, 6596.

13. Kuhn, B.; Gerber, P.; Schultz-Gasch, T.; Stahl, M. *J Med Chem* 2005, 48, 4040.
14. Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. *Bioorg Med Chem* 2007, 15, 7865.
15. Guimaraes, C. R. W.; Cardozo, M. *J Chem Inf Model* 2008, 48, 958.
16. Case, D. A.; Cheatham, T. E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J Comput Chem* 2005, 16, 1668.
17. Rocchia, W.; Alexov, E.; Honig, B. *J Phys Chem* 2001, 105, 6507.
18. Yuvaniyama, J.; Chitnumsub, P.; Kamchonwongpaisan, S.; Vanichtanankul, J.; Sirawaraporn, W.; Taylor, P.; Walkinshaw, M. D.; Yuthavong, Y. *Nat Struct Biol* 2003, 10, 357.
19. Yuthavong, Y.; Vilaivan, T.; Chareonsethakul, N.; Kamchonwongpaisan, S.; Sirawaraporn, W.; Quarrell, R.; Lowe, G. *J Med Chem* 2000, 43, 2738.
20. Tarnchompoo, B.; Sirichaiwat, C.; Phupong, W.; Intaraudom, C.; Sirawaraporn, W.; Kamchonwongpaisan, S.; Vanichtanankul, J.; Thebtaranonth, Y.; Yuthavong, Y. *J Med Chem* 2002, 45, 1244.
21. Rastelli, G.; Sirawaraporn, W.; Sompornpisut, P.; Vilavain, T.; Kamchonwongpaisan, S.; Quarrel, R.; Lowe, G.; Thebtaranonth, Y.; Yuthavong, Y. *Bioorg Med Chem* 2000, 8, 1117.
22. Rastelli, G.; Pacchioni, S.; Sirawaraporn, W.; Sirawaraporn, R.; Parenti, M. D.; Ferrari, A. M. *J Med Chem* 2003, 46, 2834.
23. Parenti, M. D.; Pacchioni, S.; Ferrari, A. M.; Rastelli, G. *J Med Chem* 2004, 47, 4258.
24. Cocco, L.; Roth, B.; Temple, C.; Montgomery, J. A.; London, R. E.; Blakley, R. L. *Arch Biochem Biophys* 1983, 226, 567.
25. Cocco, L.; Groff, J. P.; Temple, C.; Montgomery, J. A.; London, R. E.; Matwiyoff, N. A.; Blakley, R. L. *Biochemistry* 1981, 20, 3972.
26. Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.
27. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* 2004, 25, 1157.
28. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J Comput Chem* 2000, 21, 132.
29. Šali, A.; Blundell, T. L. *J Mol Biol* 1993, 234, 779.
30. Sippl, M. J. *Proteins* 1993, 17, 355.
31. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J Comput Chem* 2003, 21, 1999.
32. Åqvist, J. *J Phys Chem* 1990, 94, 8021.
33. Darden, T.; York, D.; Pedersen, L. *J Chem Phys* 1993, 98, 10089.
34. Petersen, H. G. *J Chem Phys* 1995, 103, 3668.
35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* 1977, 23, 327.
36. Tsui, V.; Case, D. A. *Biopolymer* 2001, 56, 275.
37. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. *J Comput Chem* 2004, 25, 265.
38. Feig, M.; Im, W.; Brooks, C. L., III. *J Chem Phys* 2004, 120, 903.
39. Sitkoff, D.; Sharp, K.; Honig, B. *J Phys Chem* 1994, 98, 1978.
40. Connolly, M. L. *J Appl Cryst* 1983, 16, 548.
41. Brooks, B. R.; Janezic, D.; Karplus, M. *J Comput Chem* 1995, 16, 1522.
42. Beà, I.; Jaime, C.; Kollman, P. A. *Theor Chem Acc* 2002, 108, 286.
43. Huang, N.; Shoichet, B. K.; Irwin, J. J. *J Med Chem* 2006, 49, 6789.
44. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *J Comput Chem* 2007, 28, 1145.
45. Rastelli, G.; Degliesposti, G.; Del Rio, A.; Sgobba, M. *Chem Biol Drug Des* 2009, 73, 283.
46. Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Amzel, L. M.; Freire, E. *Chem Biol Drug Des* 2007, 69, 413.
47. Gilson, M. K.; Zhou, H. X. *Annu Rev Biophys Biomol Struct* 2007, 36, 21.
48. Stoica, I.; Sadiq, S. K.; Coveney, P. V. *J Am Chem Soc* 2008, 130, 2639.
49. Foloppe, N.; Hubbard, R. *Curr Med Chem* 2006, 13, 3583.
50. Naim, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. *J Chem Inf Model* 2007, 47, 122.
51. Schutz, C. N.; Warshel, A. *Proteins* 2001, 44, 400.
52. Rankin, K. N.; Sulea, T.; Purisima, E. O. *J Comp Chem* 2003, 24, 954.
53. Fogolari, F.; Brigo, A.; Molinari, H. *Biophys J* 2003, 85, 159.
54. Wittayanarakul, K.; Aruksakunwong, O.; Saen-Oon, S.; Chantratita, W.; Parasuk, V.; Sompornpisut, P.; Hannongbua, S. *Biophys J* 2005, 88, 867.