

Rapid Estimation of Relative Protein–Ligand Binding Affinities Using a High-Throughput Version of MM-PBSA

Scott P. Brown* and Steven W. Muchmore

Abbott Laboratories, Global Pharmaceutical Research and Development, 100 Abbott Park Road,
Abbott Park, Illinois 60064-6115

Received January 30, 2007

By employing a modified protocol of the Molecular Mechanics with Poisson–Boltzmann Surface Area (MM-PBSA) methodology we substantially decrease the required computation time for calculating relative estimates of protein–ligand binding affinities. The modified method uses a generalized Born implicit solvation model during molecular dynamics to enhance conformational sampling as well as a very efficient Poisson–Boltzmann solver and a computational design based on a distributed-computing paradigm. This construction allows for reduction of the computational cost of the calculations by roughly 2 orders of magnitude compared to the traditional formulation of MM-PBSA. With this high-throughput version of MM-PBSA we show that one can produce efficient physics-based estimates of relative binding free energies with reasonable correlation to experimental data and a total computation time that is sufficiently low such that an industrially relevant throughput can be realized given currently accessible computing resources. We demonstrate this approach by performing a comparison of different MM-PBSA implementations on a set of 18 ligands for the protein target urokinase.

INTRODUCTION

The ability to calculate accurate protein–ligand binding affinities across sizable pharmaceutical compound libraries is highly desirable in a drug discovery setting, as it holds the potential to accelerate the process of finding and optimizing novel therapeutic compounds in a systematic way. Yet the methodologies that are capable of accurate estimates of binding free energies suffer from the classic tradeoff that, with the increased accuracy of a more powerful approach, there is an increasing computational cost associated with the calculation. The substantial computational barrier to obtaining accurate free energies has hindered the systematic application of the more powerful methodologies to pharmaceutical drug discovery.

The computational cost of computing binding affinities is centered on the need to search high-dimensional configuration space to find the thermal distribution of microscopic, accessible states in the system. This is complicated by the fact that the states are distributed across an energy surface dominated by energetic barriers, which is a consequence of the complicated interplay of the different types of interactions in the system, not the least of which is due to the presence of solvent molecules. Additionally, for a simulation in which solvent molecules are explicitly represented, the rate of exploration of configuration space by the solute is significantly retarded, and time scales for system equilibration are greatly exaggerated due to the damping influence of rapidly fluctuating solvent forces on the dynamics of solute degrees of freedom.

It is certainly the case that solvation can have a dramatic influence on the dynamics of embedded solutes, and thus it

is not surprising that some incorporation of the effects of solvation into molecular simulation should be pursued if one wishes to extract meaningful information. In aqueous systems there has been a longstanding effort to develop explicit molecular models of water, and, depending on the particular problem at hand, one may choose from an array of possibilities.^{1–8}

While the majority of development on explicit water models has been focused on incorporating adequate detail, alternative approaches that seek reduction in model complexity are also being explored in an attempt to elucidate the minimal aspects of water that are needed to reproduce its challenging thermodynamic spectrum.⁹ Other efforts bent on incorporating richer model complexity pursue ever more sophisticated approaches, such as accounting for the interactions of instantaneous dipoles and electronic polarizability.^{10–15}

An explicit representation of water certainly promises the most detailed glimpse into structural and dynamic properties of the solute from molecular simulation. Yet here we are primarily interested in investigating solute properties over long time scales, in which solvent molecules interacting with the solute will have presumably undergone many exchanges with bulk solvent. For this purpose it may not be necessary, or even desirable, to invest the computational time into processing explicit solvent degrees of freedom. The idea of capturing solvent effects in a mean-field way is attractive if one wishes to focus attention exclusively on solute dynamics.

There are a number of approaches to approximating the influence of solvation on solute energetics.^{16–18} The basic problem is the removal of explicit reference to solvent degrees of freedom from the expressions used to compute solute averages. One attempts to integrate over solvent coordinates to obtain a “potential of mean force” for solute–solvent interactions. The potential of mean force then allows

* Corresponding author phone: (847)937-1996; fax: (847)938-2478;
e-mail: scott.brown@abbott.com.

for the simulation of solute dynamics in the presence of an *implicit* solvent. The use of implicit solvation decreases the computational cost associated with the simulation of solute dynamics, allowing for access to conformational motion that may occur on inaccessible time scales in explicit-solvent simulation.

In this work we employ a version of the approximate end-point free-energy method developed by Case and Kollman,¹⁹ which has been termed Molecular Mechanics with Poisson–Boltzmann Surface Area (MM-PBSA). MM-PBSA has recently been applied in a variety of contexts, to address a range of problems.^{20–31} In the original formulation of MM-PBSA,³² solute dynamics are explored in the presence of explicit water molecules; however, given current computing resources, this is too costly for use in the systematic analysis of large compound libraries. In light of this, we incorporate procedural modifications to the traditional MM-PBSA methodology introduced by Rizzo et al.³³ in an attempt to reduce the required computation time while still providing meaningful results. Our implementation of MM-PBSA employs the recent implicit-solvent model of Onufriev et al.,³⁴ compared to Rizzo et al. use of the older method of Hawkins et al.³⁵ Other differences between the implementations are seen with the construction of the dynamics as well as in the method used to obtain solutions to the Poisson–Boltzmann (PB) equation. The use of implicit solvation to generate the dynamics trajectories allows at least an order of magnitude reduction in the number of time steps compared to what is traditionally used. This results in an overall computational cost that is 2 orders of magnitude lower for the high-throughput version, which, when combined with the fact that the calculation is well suited to deployment onto an enterprise grid, makes it potentially a very relevant tool in industrial drug design.

We demonstrate that the high-throughput version of MM-PBSA is capable of producing predicted relative binding free energies with substantial correlation to experimental data. For comparison, we also show that the increased efficiency of the protocol achieves similar correlation to experiment as is obtained with the traditional implementation of MM-PBSA. This validation is significant as we wish to construct a robust physics-based scoring methodology that can be systematically applied in pharmaceutical drug discovery.

COMPUTATIONAL METHODS

Description of the Test Set. In previous work,³⁶ we examined relative free-energy predictions on a literature set of data for seven biotin analogs interacting with the protein avidin.²⁴ Here we present additional work investigating the calculation of binding affinities for a set of 18 ligands with the protein target urokinase. Urokinase is a serine protease and has been implicated in a number of tumor-related activities.^{37,38} The 18 compounds we investigate are part of a collection of molecules developed in our urokinase inhibitor program. We obtain the starting structures for the urokinase complexes from in-house X-ray crystallography.³⁹ In addition, all binding affinities for the small-molecule ligands are obtained from in-house enzyme-based assays.⁴⁰

Shown in Table 1 are the molecular structures of these compounds (designated by the numbers **1–18**) as well as their molecular weight, binding affinities, and the resolution

at which the crystal structure was determined. The urokinase compounds have a molecular weight range of 144–479 Dalton and potencies that span 4 orders of magnitude, with inhibition constants in the range 0.046–100 μ M. Note that in this data set there is a correlation of the potencies to compound molecular weight. We initially performed the analysis on a set that did not exhibit this type of correlation; however, portions of this set were subsequently revealed to not be suitable for release into the public domain, which serves to illustrate the reality of the difficulties involved in attempting to release proprietary data into the public sector.

In Figure 1 we show a cartoon depiction of the crystallographic coordinates for urokinase in complex with compound **12**. The binding mode shown is typical of the compounds investigated in this study, in which the ligand interacts with ASP193 and SER194 in the deepest part of the binding pocket. This interaction provides a shared anchor for all of the compounds and induces a common orientation to the binding modes in the pocket.

With this set of 18 urokinase ligands, we compare two different MM-PBSA constructions, termed traditional and high-throughput, which each use a different form of molecular dynamics (MD) routine to generate trajectories of the protein–ligand complex. For traditional MM-PBSA we perform the MD in the presence of explicit waters, whereas for the high-throughput method the MD is performed using an implicit solvent. Our motivation for performing this comparison is a desire to explore computational technologies for performing rapid, physics-based scoring of small-molecule drugs with sufficient throughput to impact industrial pharmaceutical research. The high-throughput version of MM-PBSA presents an efficient combination of methods and possesses enough of a foundation in physics to facilitate penetrable analysis of results.

MM-PBSA General Formulation. The framework of the general procedure is broken down into two separate steps: first, MD simulation is performed on the protein–ligand complex to generate a trajectory, from which a set of structures is obtained that represents the putative “thermal ensemble;” and second, analysis of gas-phase and solvation-phase energetics is performed on each member of the ensemble. The final binding free energy is then obtained by combining the various energetic averages over the ensemble.

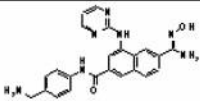
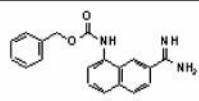
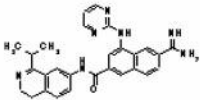
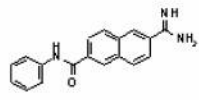
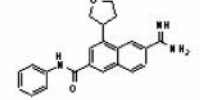
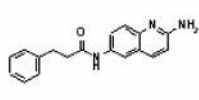
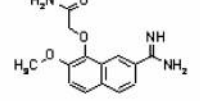
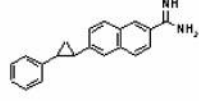
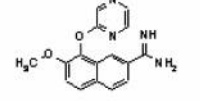
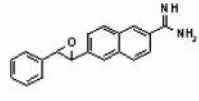
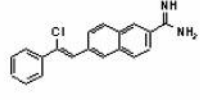
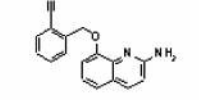
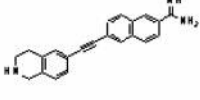
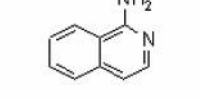
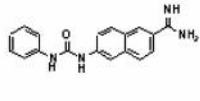
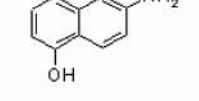
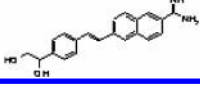
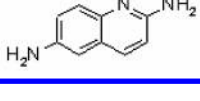
The following approximate form is assumed for the net change of free energy accompanying the association in solution of unbound protein with free ligand to form the bound protein–ligand complex

$$\langle \Delta G_{\text{bind}}^{\text{MM-PBSA}} \rangle = \langle \Delta E_{\text{MM}} \rangle + \langle \Delta G_{\text{PBSA}} \rangle - T \langle \Delta S_{\text{solute}} \rangle \quad (1)$$

where the above brackets $\langle \dots \rangle$ indicate the ensemble average of the enclosed quantity, ΔE_{MM} is the change in molecular mechanics energy upon binding, ΔG_{PBSA} is the change in solvation free energy upon binding, T is the absolute temperature of the system, and ΔS_{solute} is the change in internal entropy of the solutes upon binding. Note that the approximations inherent in this decomposition of the binding free energy are not necessarily valid,⁴¹ but for the current purpose we find the approximation to be reasonable and certainly convenient.

The quantity E_{MM} is obtained from the molecular mechanics (MM) force field used in the MD simulation, which is

Table 1. List of the Compounds Used in the Calculation of Urokinase Binding Affinities^a

Compound	MW (Dalton)	K _i (μM)	Resolution (Å)	Compound	MW (Dalton)	K _i (μM)	Resolution (Å)
	429.5	0.046	1.9		320.4	6.15	1.8
	478.6	0.070	3.3		290.3	7.86	2.0
	360.4	0.324	2.3		291.4	10.03	2.3
	274.3	0.624	2.0		287.4	13.15	2.1
	295.3	1.008	1.9		289.4	18.28	1.7
	307.8	1.770	2.3		275.3	29.85	2.0
	324.4	2.142	2		144.2	49.37	2.7
	305.4	2.294	1.6		159.2	50.21	2.3
	333.4	3.134	2.3		159.2	99.89	2.3

^a Each compound is designated by a number 1–18, ordered by decreasing affinity for urokinase. Shown for each compound are molecular weight, experimentally measured inhibition constant, and the resolution at which the structure was determined by in-house X-ray crystallography. Reported *K_i* values are obtained from IC₅₀s determined through in-house enzyme-based assays.

evaluated (after completion of the production MD runs) in the gas phase with no distance cutoff on nonbonded interactions.

The next term on the right-hand side of eq 1, ΔG_{PBSA} , is comprised of two separate parts

$$\Delta G_{\text{PBSA}} = \Delta E_{\text{PB}} + \Delta G_{\text{SA}} \quad (3)$$

E_{PB} is obtained by solution of the PB equation, which evaluates the work of forming a dielectric boundary around a charged solute embedded in a bulk dielectric medium. The PB equation provides the most rigorous framework for representing solvent as a uniform dielectric continuum and has been shown to provide robust predictions of electrostatic contributions to solvation free energies of small molecules as well as biological macromolecules.^{42–44} As PB solutions are in general fairly computationally demanding, the field has seen a number of efforts directed at developing approximate PB methods such as generalized Born (GB) approaches;^{45–48} one of which we use in this work to generate

MD trajectories (see below). There are well-known, documented limitations to GB methodologies,^{49,50} and it is thus of interest to pursue some form of implementation of PB forces into the dynamics calculations; however, here we use PB only for postproduction analysis and obtain solutions in separate calculations for the ligand, protein, and bound protein–ligand complex, after which the final solvation free-energy values are assembled by invoking a conceptual thermodynamic cycle for association in solution.^{51,52}

The last term on the right-hand side of eq 3, ΔG_{SA} , represents the contribution to the solvation free energy due to changes in the water-accessible nonpolar surface area of the solutes upon binding, which is given by the empirical relationship

$$\Delta G_{\text{SA}} = \gamma \cdot \Delta A \quad (4)$$

where ΔA is the change in accessible molecular surface area upon binding, and γ is the “microscopic” surface free energy for formation of a cavity in water.⁵³ The form of eq 4 derives

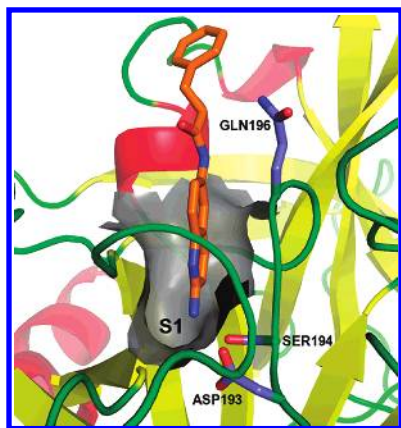


Figure 1. Schematic representation of the urokinase active site in complex with compound **12**, as determined by in-house X-ray crystallography. Shown at the lower end of the binding pocket are residues ASP193 and SER194, which are part of the S1 pocket of the active site of urokinase. The S1 binding pocket is a large pore, and all the compounds in this study are anchored in this cavity. Interactions with the residues in S1 present a common binding site and result in similar binding modes for all of the small-molecule ligands.

from empirical fit to experimental transfer free energies,^{54,55} and the precise value of γ depends upon the method used to determine the solvent-accessible surface area of the solute.^{42,56} Values for ΔG_{SA} are typically small relative to other energetic contributions, and thus their contribution to the overall binding affinity is relatively minor. Due to the fact that eq 4 is obtained from fit to experimental data it inherently includes contributions from solvent reorganizational entropy.

Solute Entropy. The final term on the right-hand side of eq 1, ΔS_{solute} , is the change in the internal solute entropy on formation of the protein–ligand complex. With our construction of the high-throughput MM-PBSA calculations, estimates of solute entropy would require an additional step in the calculation, which can be highly computationally expensive to perform using methods such as normal-mode analysis.⁵⁷ Alternatively, one could turn to more computationally accessible methods, such as those based on empirical conjecture of the average entropic cost of constraining rotation around a given torsional degree of freedom.⁵⁸ In our experience, neither approach produces quality estimates of solute entropy and instead tend to add a significant component of random scatter to the results of the calculation. In this work we neglect all changes in solute entropy for both the traditional and high-throughput methods. This omission is presumed to be a reasonable approximation due to the fact that we are interested only in a relative rank-ordering of binding free energies of a set of ligands for a particular protein target. Of course, that is not to say it is reasonable in general. Certainly it is unreasonable for the case where absolute comparisons of free energy are desired (e.g., across protein targets) and also perhaps for situations in which one expects a non-negligible range of perturbations on binding modes and pocket geometries to occur across the particular set of compounds at hand.

General Molecular Dynamics Protocol. To obtain meaningful energetic averages, we require an ensemble of structures to represent the thermal distribution of states. This ensemble of conformations is generated by running MD on the protein–ligand complex using SANDER from the AMBER (v8.0) simulation package.⁵⁹ We perform the MD

simulations using both explicit (traditional MM-PBSA) and implicit (high-throughput MM-PBSA) representations of water and compare results in order to investigate differences between the two methods.

The protein force field is taken from ff03⁶⁰ (note that we also investigated using ff99⁶¹ and observed only minor qualitative differences in results). In the case of the small-molecule ligands we obtain partial charges using the AM1-BCC method.⁶² AM1-BCC is very efficient at calculating partial charges, which makes it ideal for use over large libraries of compounds. The force-field parameters for the small molecules are taken from the generalized force field of Wang et al.⁶³ We accomplish the assignments for both AM1-BCC partial charges and atomic force-field parameters by using ANTECHAMBER⁶⁴ (v1.23). This is a fully scripted process that is run as an automated procedure and thus is capable of processing large numbers of small molecules with little to no user intervention.

Note that in the original formulation of what we are calling traditional MM-PBSA it was necessary to perform an ab initio quantum-mechanics calculation for each ligand in order to obtain the small-molecule force-field (i.e., geometry) and partial-charge parameters. As our goal is to avoid labor-intensive system parametrization, we have chosen a force field for the small molecules for which rapid assignments of atom types can be performed and automated. A comparison has been performed between the Wang et al. small-molecule parameters and those obtained by performing quantum-mechanics calculation at the HF/6-31G* level for a small set of biotin analogs with the protein avidin.³⁶ The two force fields appear to produce comparable results based on the correlation to experimental data produced by the corresponding MM-PBSA calculations; however, it is significantly more difficult to automate the quantum calculations. In light of this we use identical protein and ligand force-field parameters and partial charges for all the MM-PBSA calculations performed herein.

We do not perform separate MD simulations for the isolated ligand and *apo* protein, and hence the MD trajectory of the protein–ligand complex is used as the source of conformations for the unbound and bound states. This is equivalent to assuming that the conformations explored in the protein–ligand complex are sufficiently similar to those conformations explored by the *apo* protein and isolated ligand in solution. While this assumption is not necessarily reasonable, the amount of noise added by doing separate runs increases the convergence time in the system such that the required computation time becomes prohibitively large;⁶⁵ however, it is true that for the system studied here one would only need to perform a single long MD run on the *apo* protein, and additional MD runs then only need be performed separately for each isolated ligand.

Explicit-Water Protocol. For the explicit-water MD we begin with our in-house crystallographic coordinates, and, while retaining any crystallographic waters that are present, we add a 15 Å layer of surrounding waters using TIP4P water model.⁴ This initial solvated system containing the protein–ligand complex is neutralized by the addition of an appropriate number of monovalent counterions. The system is then steepest-descent minimized for 500 steps, followed by 200 steps of conjugant-gradient minimization. This length of minimization is sufficient to remove any pathological clashes

between atom centers that are initially too close in proximity. The additional 200 steps of conjugant-gradient minimization may not be strictly needed, but it is added as a precautionary measure to avoid hydrogen-placement clashes that may be difficult to remove in the presence of explicit waters with the simple steepest-decent minimizer.

During the equilibration and production dynamics all hydrogen bonds are constrained using SHAKE,⁶⁶ which is a requirement when using a water model such as TIP4P. The numerical integrator used is the leapfrog Verlet algorithm.⁶⁷ Periodic boundary conditions are enforced with a time step of 2 fs, a distance cutoff of 10.0 Å on the nonbonded interactions with Particle Mesh Ewald⁶⁸ used for long-range electrostatic interactions. Starting from the minimized configuration, the system is equilibrated by gradually heating from 0 to 300 K over 20 ps under constant-volume conditions. The final structure from the constant-volume run is then passed on to a subsequent equilibration for 150 ps at constant pressure using an Andersen piston⁶⁹ that is maintained at the target temperature of 300 K using the Berendsen weak-coupling scheme.⁷⁰ The final configuration from the constant-pressure equilibration is then used as input for 500 ps of production MD. During the production runs we save 100 structural snapshots of the system at equally spaced intervals in time along the trajectory, which are subsequently used in the postproduction energetic analysis.

Implicit-Solvent Protocol. During the high-throughput MD runs all hydrogen bonds are constrained using SHAKE, with a time step of 2 fs, and a distance cutoff of 12.0 Å for the nonbonded interactions. To explore the effects of cutoff length on the simulations we looked at the impact of both longer and shorter cutoffs. While a shorter cutoff lowers the computational cost for numerical integration of the equations of motion, it may also have a negative impact on quality. It was apparent that 10.0 Å was too short to justify the speed up (we observed noticeable degradation in results), but a longer cutoff of 16.0 Å appeared to present only minimal changes in results. Hence, we found a value of 12.0 Å to be an acceptable balance between speed and accuracy. Finally, we use an internal (solute) dielectric of 1.0 with the external (bulk) dielectric set to 80.0.

Starting from the crystal-structure coordinates of the protein-ligand complex, we strip all water molecules and nonpeptide atoms before performing a steepest-decent minimization for 500 steps. The final structure from minimization is passed on to an equilibration phase, in which we heat the system from 0 to 300 K over 6 ps using constant-temperature Langevin dynamics with a leapfrog integrator⁷¹ and a collision frequency of 2.5 ps⁻¹. The final structure from equilibration is then input as the initial structure into the MD production runs. We perform production MD runs for 13 ps and discard data from the first 3 ps. This is done to avoid capturing snapshots with significant correlation to the Langevin equilibration. Over the course of the final 10 ps we save to disk a series of 10 coordinate snapshots of the protein-ligand complex, taken at evenly spaced intervals in time along the trajectory.

Zap Analysis. The snapshots generated during MD are used to perform the postproduction solvation-energy analysis with the Zap module from OpenEye Scientific Software.⁷² Zap is a very efficient PB solver and can, in some cases, be faster than GB methodologies.⁴⁸ ZAP uses a representation

of the dielectric boundary that alleviates a number of grid-related problems present in other PB treatments.⁷³ The partial charges for the ZAP calculations are those initially used for the MD runs (see above), and the radii used are the default set implemented within ZAP. An internal (solute) dielectric of 1.0 is used, with an external (bulk) dielectric of 80.0. The PB calculations in ZAP are performed using a focused grid that is centered on the ligand with a grid spacing of 0.5 Å and a boundary spacing of 3.0 Å. This construction results in very rapid convergence of the PB solutions, yielding average timings on a 3.0 GHz Xeon computer of around 0.25 s per protein-ligand complex. For the ΔG_{SA} term in eq 4 we use a value of $\gamma = 0.005 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$.⁷⁴ All surface areas for the protein, ligand, and protein-ligand complex are obtained from the ZAP module during the PB calculations.

Comparison of Computational Cost. For the explicit-solvent construction, completion of the MD part of the calculation (the computational bottleneck) takes approximately 7 CPU days per complex running serially on a 3.0 GHz Xeon processor. In comparison, the implicit-solvent MD takes on the order of 100 CPU min to complete. While the wall-clock time for the explicit-solvent calculation could be significantly reduced through parallelization of the MD routine, in order to approach the desired scale of analysis it would be necessary to acquire high-end, specialized hardware, which typically requires large capital expenditure. In contrast, the design of the high-throughput MM-PBSA calculation is well suited to deployment onto the coarse-grain-parallel architecture of a distributed grid, which allows for very rapid, cost-efficient turnaround times. All of the high-throughput results reported herein were obtained from calculation on our enterprise grid, from which we harness otherwise idle CPU cycles on employee desktop personal computers. As described previously,³⁶ our enterprise grid is built upon the freely distributed package CONDOR from the computer science department at the University of Wisconsin, Madison.⁷⁵

RESULTS AND DISCUSSION

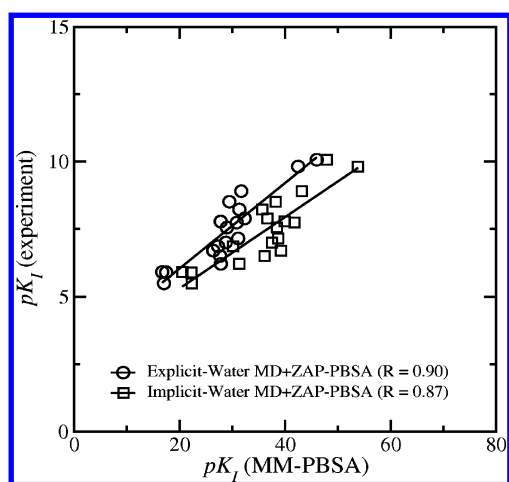
The main results of this work are shown in Table 2, in which we report the predicted binding affinities from the traditional and high-throughput MM-PBSA implementations for the 18 urokinase compounds. Also shown in Table 2 are the numeric rankings for the predictions. As the compound list was initially sorted based on decreasing binding affinity for urokinase, the numeric designation assigned to each compound represents its ideal rank (i.e., compound **1** is the tightest binder, and compound **18** the weakest). It is clear from inspection of the predicted rankings that both methods are distinctly able to differentiate the tightest from the weakest binders. This is seen by noting that it is more likely to find a higher ranking (lower number) at the top of Table 2 than it is to find a lower ranking (higher number) at the top, and vice versa, for both of the methods.

It is obvious from a cursory inspection of the binding-affinity values in Table 2 that the predicted affinities are substantially, and systematically, larger than the experimental values. This is true for both the traditional and high-throughput approaches and is due, to a significant degree, to the systematic error introduced by the absence of an

Table 2. Comparison of Experimental (expt) Binding Affinities for 18 Urokinase Compounds with Predictions from Both Traditional MM-PBSA (trad) and the High-Throughput Version of MM-PBSA (HT)^a

	pK_i^{expt}	pK_i^{trad}	rank (trad)	pK_i^{HT}	rank (HT)
1	7.3	46	1	481	2
2	7.2	42	2	54	1
3	6.5	32	4	43	3
4	6.2	29	8	38	9
5	6.0	31	5	36	13
6	5.8	32	3	37	11
7	5.7	28	12	40	5
8	5.6	31	7	42	4
9	5.5	29	9	38	8
10	5.2	31	6	39	7
11	5.1	29	10	38	10
12	5.0	27	14	30	15
13	4.9	26	15	39	6
14	4.7	28	13	36	12
15	4.5	28	11	31	14
16	4.3	17	18	21	18
17	4.3	17	16	22	17
18	4.0	17	17	22	16

^a The predicted rank for each compound is shown for clarity. The number designation chosen for each compound (first column) indicates its ideal rank as the compounds are ordered by decreasing potency (see Table 1).

**Figure 2.** Predicted binding affinities versus experimentally measured binding affinities, expressed as $-\log(K_i)$, for 18 urokinase ligands: circles, traditional MM-PBSA; squares, high-throughput MM-PBSA. The lines show linear least-square fits to the data, with the corresponding correlation coefficients indicated in the legend.

estimation of solute entropy in the calculation. But, as we are primarily interested in performing rapid rank ordering of relative binding affinities, we find it to be computationally advantageous to neglect all entropic contributions from solute reorganization upon binding, which appears to be reasonable based on the predicted rankings in Table 2.

To attempt to quantify the correlation between prediction and experiment, we plot in Figure 2 the predicted affinities against experimental data. The predictions obtained for both the traditional and high-throughput versions of MM-PBSA show comparable correlation to the experimental data, with correlation coefficients of 0.90 for traditional MM-PBSA and 0.87 for the high-throughput version.

Yet the computational cost to process a single structure with the traditional formulation of MM-PBSA is about 2 orders of magnitude greater than for the high-throughput

method. Based on the observed correlation for this set of urokinase compounds, it seems unnecessary to perform a traditional-type MM-PBSA analysis if one is primarily interested in rapid, relative rank ordering of binding affinities for a particular protein target. It is worth emphasizing that, in order for a computational tool to impact industrial pharmaceutical research, it is crucial for results to be delivered in a timely manner.

To take a deeper look into the MM-PBSA calculations on urokinase, we examine the energetic breakdown of the predicted affinities. Shown in Table 3 are the computed averages for the different contributions to the energetic changes upon binding that are combined to form the final MM-PBSA binding-affinity estimates. It can be seen that the average changes in van der Waals energies (ΔE_{vdw}), and corresponding standard deviations, are fairly similar between methods. There does appear to be perhaps a slight systematic discrepancy, with the implicit-solvent runs producing slightly less negative values. In contrast, the net changes in electrostatic interactions (ΔE_{elec}) appear to exhibit a bias toward more negative values for the implicit-solvent runs and a greater average discrepancy between values, and standard deviations, for the two methods than is the case for the van der Waals values. This is presumably a direct consequence of the possibly substantially different free-energy surfaces for proteins in implicit-solvent models compared to those using explicit-water representations.⁷⁶ As electrostatic interactions act over a significantly longer range than do van der Waals forces, it is not surprising that the electrostatic interactions are more sensitive to total system configuration. Thus systematic differences in electrostatic energies in the systems perhaps serve to underline the truly different nature of the underlying free-energy surfaces upon which we are sampling, as both the implicit- and explicit-solvent methodologies evaluate the postproduction energies in the same manner: with no distance cutoff; all solvent atoms removed (if present); and identical force fields and partial charges for both protein and small-molecule ligand.

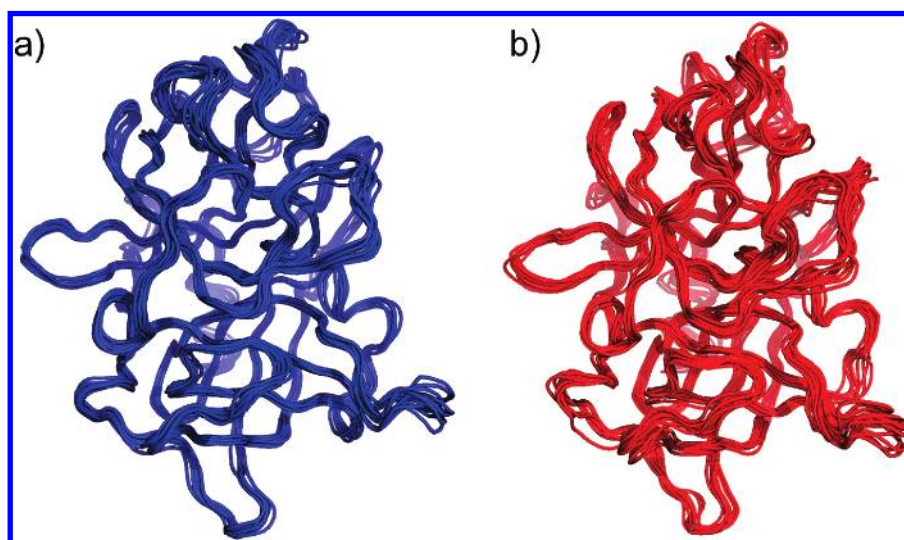
Differences can also be seen when comparing Table 3 values of the solvent-energy contribution to predicted affinities for the electrostatic component (ΔE_{PB}) as well as the nonpolar surface-area component (ΔG_{SA}). Some discrepancies are to be expected due to the different MD techniques used in generating the ensemble averages; however, there does not appear to be any systematic deviation. Although perhaps the implicit-solvent average ΔE_{PB} values are larger than the explicit-solvent values more often than not, but then the standard deviations of the implicit-solvent averages are larger more often than not.

The above discrepancies are a manifestation of sampling differences between explicit- and implicit-solvent dynamics. To investigate these differences in a qualitative manner we look at a typical example of the conformational motion over the production dynamics runs for both methods. Superimposed in Figure 3 are sets of backbone traces that span the production MD trajectories of urokinase in complex with compound 4, for both explicit and implicit solvation. To reduce the amount of information and help facilitate comparison we select 10 states equally spaced in time along the trajectories. While the differences are not dramatic (both backbone traces follow similar paths through space), careful inspection will reveal the greater variation in backbone

Table 3. Comparison of the Breakdown of the Energetic Contributions to MM-PBSA Predicted Binding Affinities for Traditional MM-PBSA (Explicit Water MM, OpenEye Zap) and the High-Throughput Version of MM-PBSA (Implicit Water MM, OpenEye Zap)^a

	traditional MM-PBSA								high-throughput MM-PBSA							
	explicit water MM				zap PBSA				implicit solvent MM				zap PBSA			
	$\langle\Delta E_{\text{vdW}}\rangle$	σ_{vdW}	$\langle\Delta E_{\text{elec}}\rangle$	σ_{elec}	$\langle\Delta E_{\text{PB}}\rangle$	σ_{PB}	$\langle\delta G_{\text{SA}}\rangle$	σ_{SA}	$\langle\Delta E_{\text{vdW}}\rangle$	σ_{vdW}	$\langle\Delta E_{\text{elec}}\rangle$	σ_{elec}	$\langle\Delta E_{\text{PB}}\rangle$	σ_{PB}	$\langle\delta G_{\text{SA}}\rangle$	σ_{SA}
1	-51	5	-147	20	138	18	-4.3	0.1	-42	6	-144	18	148	13	-3.8	0.1
2	-54	5	-90	7	90	6	-4.3	0.1	-48	5	-99	9	96	10	-4.5	0.2
3	-39	53	-79	7	77	6	-3.0	0.2	-38	4	-80	9	83	7	-3.5	0.2
4	-32	3	-80	10	74	9	-2.9	0.1	-32	4	-110	12	105	10	-3.2	0.1
5	-35	4	-94	7	89	6	-3.0	0.1	-33	3	-80	9	80	6	-2.8	0.2
6	-39	3	-80	6	77	6	-3.3	0.1	-33	3	-81	9	78	8	-2.9	0.1
7	-34	4	-56	7	55	6	-2.9	0.2	-37	4	-79	7	73	8	-3.4	0.1
8	-38	3	-66	15	65	12	-3.2	0.1	-35	3	-83	8	95	8	-3.5	0.1
9	-34	4	-73	12	69	9	-2.9	0.3	-34	4	-71	17	60	7	-3.2	0.4
10	-39	3	-71	9	70	7	-3.2	0.1	-35	3	-79	9	82	6	-3.0	0.1
11	-33	4	-77	9	74	7	-2.8	0.1	-35	4	-78	9	77	8	-3.1	0.1
12	-37	2	-33	6	36	5	-2.9	0.1	-33	3	-26	5	31	4	-3.0	0.3
13	-31	4	-62	6	59	5	-2.8	0.3	-35	4	-76	7	63	10	-3.2	0.1
14	-35	3	-78	10	78	8	-2.8	0.1	-30	4	-85	9	72	9	-2.8	0.2
15	-36	3	-44	6	45	5	-2.7	0.1	-33	3	-28	4	32	8	-2.9	0.2
16	-27	2	-17	6	23	4	-1.9	0.1	-25	2	-21	3	16	5	-2.0	0.1
17	-23	2	-29	6	30	4	-2.0	0.1	-27	2	-21	4	26	5	-2.2	0.1
18	-24	2	-25	5	28	4	-2.1	0.1	-26	2	-38	10	19	3	-2.0	0.1

^a Shown are the averages in the calculated energetic changes upon binding over the production MD runs. Also included are corresponding standard deviations. All numbers are in units of kcal/mol.

**Figure 3.** Backbone traces of 10 snapshots taken from production MD trajectories of urokinase in complex with small-molecule compound **4** (not shown): (a) explicit water and (b) implicit solvent. The superpositions shown are taken at equally spaced intervals in time along the dynamics trajectories. The fluctuations in backbone position for the implicit-solvation runs appear to exhibit greater variation compared to explicit-solvation runs.

position for the implicit solvation, which is indicative of differing degrees of flexibility in backbone dynamics between the explicit (Figure 3a) and implicit (Figure 3b) representations. In fact, the average root-mean square deviation (rmsd) of backbone alpha-carbons from the crystal-structure coordinates of compound **4** is 1.22 Å for explicit solvent and 1.69 Å for implicit solvent. It is worth emphasizing that the explicit-water structures in Figure 3a are taken at equally spaced points in time along a trajectory of 500 ps, whereas the implicit-water structures in Figure 3b are taken from a trajectory of 10 ps.

From arguments based on flexibility differences, one might intuitively expect the implicit-solvent dynamics to produce a greater spread of values about the mean electrostatic contribution to the solvation energy. In fact, we see mostly greater standard deviations in the PB energies from the implicit-solvent runs. However, to illustrate this in a more

quantitative manner, we take a closer look at conformational motion over the dynamics trajectories. Listed in Table 4 are average rmsd values for the protein backbone and small-molecule ligands over the course of the production MD runs. Two different measures of rmsd are shown for the protein in Table 4, with only one rmsd measure shown for the ligands. These are designated as either *X-ray* or *ensemble*. The *X-ray* measure represents the average rmsd relative to the crystal structure. This is the alpha-carbon backbone positions for the protein, whereas for the ligands it is the positions of all atoms in the molecule. The second measure of rmsd (termed *ensemble*) characterizes the average pairwise separation of alpha-carbon positions for all protein structures in the conformational ensemble (the set of structures taken from production MD). In comparing the protein rmsd values for the *X-ray* measure we see that, more often than not, the implicit-water backbone motion exhibits a greater deviation

Table 4. Comparison of Average Values for Urokinase Backbone Alpha-Carbon rmsd and All-Atom rmsd for the Small-Molecule Ligands Taken over the Course of Production MD for Traditional (trad) and High-Throughput (HT) MM-PBSA^a

	protein				ligand	
	traditional		high-throughput		trad <rmsd> (X-ray)	HT <rmsd> (X-ray)
	<C _α rmsd> (X-ray)	<C _α rmsd> (ensemble)	<C _α rmsd> (X-ray)	<C _α rmsd> (ensemble)		
1	1.25	1.19	1.28	1.15	1.31	1.11
2	1.55	0.96	1.33	1.04	1.55	1.49
3	1.21	0.98	1.29	1.28	1.13	1.29
4	1.22	0.82	1.69	1.27	1.24	0.82
5	1.00	0.77	1.37	1.03	0.74	0.53
6	1.28	1.08	1.34	1.02	0.98	0.78
7	1.36	0.82	1.35	1.14	0.82	1.22
8	1.15	1.16	1.19	1.01	1.71	1.68
9	1.27	0.80	1.33	1.22	0.76	0.95
10	0.98	0.89	1.23	1.09	1.86	0.86
11	1.01	0.80	1.45	0.95	1.67	0.94
12	1.45	1.03	1.29	1.03	1.94	1.27
13	1.15	1.01	1.29	1.22	1.40	1.30
14	1.07	0.95	1.30	1.05	0.95	1.36
15	1.51	0.89	1.23	1.00	1.66	0.95
16	1.39	1.12	1.10	1.05	0.22	0.21
17	1.47	0.92	1.19	0.95	0.51	0.32
18	1.26	0.96	1.13	0.97	0.57	0.26

^a rmsd values designated as X-ray are relative to crystal-structure coordinates, and those designated as ensemble are the pairwise rmsd for a set of 100 structures taken at evenly spaced intervals in time from the production MD trajectories.

from the crystal structure than is seen for the explicit-water runs. The rmsd values for the *ensemble* measure also show a similar discrepancy between explicit and implicit solvent, with the majority of the implicit-solvent runs exhibiting larger average deviations. This is suggestive that 10 ps of backbone motion under implicit solvation contain a comparable degree of conformational motion as occurs over 500 ps of explicit-water dynamics, at least for average backbone fluctuations.

In contrast to the protein backbone average rmsd measures in Table 4, we find that ligand rmsd values for the explicit-water dynamics produce larger average deviations from crystal structure than occur over the course of the implicit-solvent dynamics. There are a number of possible reasons for this, such as small-molecule force-field inadequacies, poorly assigned water positions in the starting crystal-structure coordinates, or perhaps it is simply due to motion on different time scales; however, it is interesting to note that the removal of explicit water appears to result in implicit-solvent dynamics with less adjustment of the ligand position away from its crystallographic location.

We show in Figure 4 the time evolution of two measures of protein rmsd, with values displayed over the course of the production MD trajectories for both explicit and implicit solvent dynamics. Figure 4a shows the time dependence of backbone motion by plotting the alpha-carbon rmsd relative to crystal-structure coordinates. It is obvious that the presence of explicit waters produces a non-negligible effect on protein-backbone dynamics. A readily apparent difference in the dynamics in Figure 4a is the rapid fluctuations of backbone rmsd for explicit solvent (blue line). These fluctuations in the backbone are due to rapid oscillation of solvent forces caused by motion of the explicit waters. Typically these forces are not productive in promoting the exploration of backbone conformations; however, a more relevant aspect

of the plots in Figure 4a is the extent to which the backbone motion traverses different conformations over the course of the dynamics, i.e., the amplitude of the motion. It is the ability to sample different conformations that is of primary interest in this work. For the plots in Figure 4a there appear to be a qualitatively similar exploration of backbone space, that is, the low-frequency fluctuations in Figure 4a appear to be roughly comparable for explicit and implicit solvation.

Another measure of protein motion is shown in Figure 4b, in which the rmsd of the active-site side chains (relative to X-ray crystal structure) is monitored over the course of the production dynamics. The active-site side chains are roughly defined as those side chains lying within 5 Å of the ligand in the crystal structure. We see that differences in the rmsd of the active-site side chains exhibit similar characteristics as seen in the dynamics trajectories for backbone rmsd. It is interesting that, for the most part, the explicit water dynamics appears to cause a greater deviation of the active-site side chains from their crystal structure location than occurs with implicit solvent. This may explain the greater average ligand rmsd for the explicit-water dynamics seen in Table 4 and may be a consequence of the different time scales used in the simulation.

In closing, it is worth making a final comparison of the convergence rates of the calculated MM-PBSA free-energy estimates. This can be done in a quantitative manner by calculating a time correlation function of the free-energy estimates over the course of the dynamics trajectories, which naturally fluctuate due to intrinsic motions in the system. Shown in Figure 5 is the decay in the correlation of calculated free-energy values as a function of snapshot number for the explicit and implicit solvent runs. Fits to the data yield characteristic relaxation times of about 50 ps for explicit-water dynamics and about 1 ps for the implicit-solvent dynamics. These rates of decorrelation indicate that for both the explicit and implicit trajectories we are sampling approximately $10\tau_{AG}$.

CONCLUSIONS

The modifications we make to Case and Kollman's original formulation are strictly procedural in nature and do not change the underlying conceptual framework of the method. We use the recent GB implicit solvation method of Onufriev et al.³⁴ to facilitate shortened MD runs that are used to produce the ensemble averages for MM-PBSA binding-affinity estimates. The GB model allows for an order-of-magnitude reduction in the number of time steps used in MD compared to the traditional (explicit-water) formulation of MM-PBSA. Additionally, we employ a fast, smooth-permittivity method⁷³ for estimating the electrostatic and nonpolar contributions to solvation energy. The automated small-molecule force-field and partial-charge assignments are also an integral part of our methodological modifications. Additionally, the fact that we only need to process 10 snapshots, as compared to typical numbers for traditional MM-PBSA in the range of 50–100, is advantageous as the calculation consumes fewer resources. These modifications, in combination with a construction designed for deployment onto an enterprise grid, form the basis for performing high-throughput MM-PBSA.

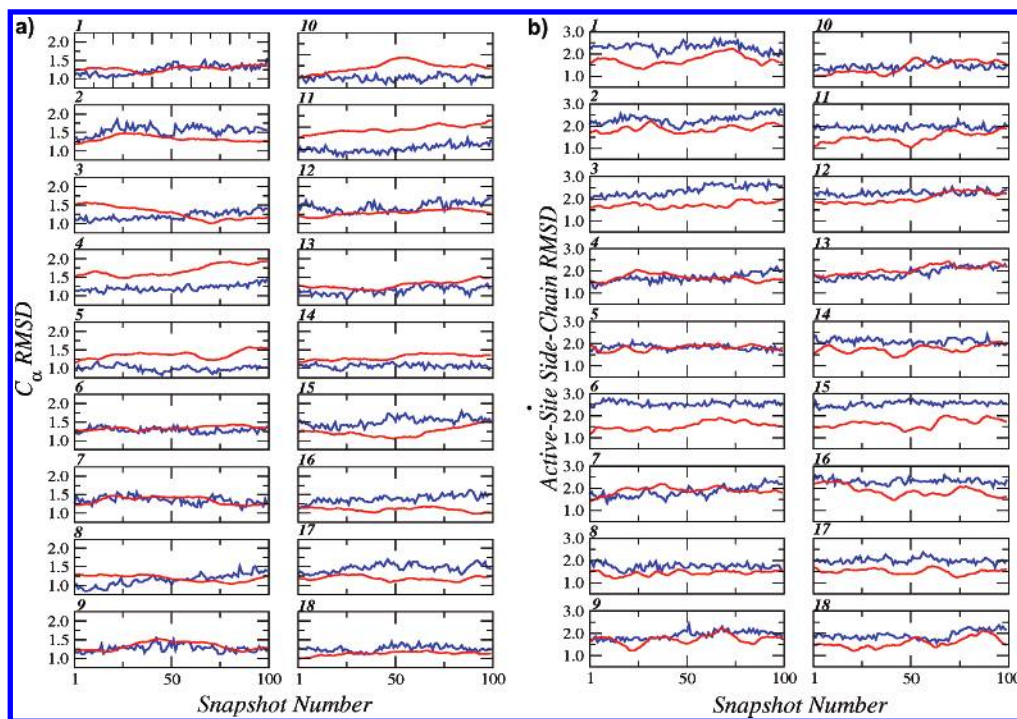


Figure 4. Measures of root mean-square deviation (rmsd) relative to the crystal structure for 100 urokinase snapshots taken over the course of the production MD runs for both explicit (blue line) and implicit (red line) solvent. (a) Protein alpha-carbon rmsd relative to crystal-structure coordinates. (b) All-atom rmsd of active-site side chains relative to crystal structure coordinates. The active site is defined as those protein atoms within 5.0 Å of the ligand. The identity of the complex in each panel is indicated by the number designation located in the upper left corner. The explicit-water trajectories are 500 ps, whereas the implicit-solvent trajectories are 10 ps.

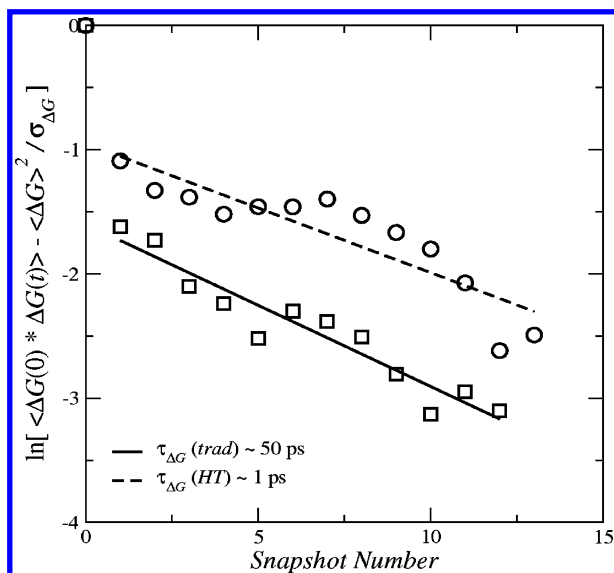


Figure 5. Log of time-correlation measure for MM-PBSA free-energy estimates $\ln[\langle \Delta G(0) * \Delta G(t) \rangle - \langle \Delta G \rangle^2 / \sigma_{\Delta G}^2]$ as a function of the snapshot number used from the production MD simulations for both traditional and high-throughput MM-PBSA protocols. $\sigma_{\Delta G}^2$ is a scale factor given by $\sigma_{\Delta G}^2 = \langle \Delta G^2 \rangle - \langle \Delta G \rangle^2$. The rough time constant for decay is obtained by linear fit to the data (indicated by the lines), with the corresponding values indicated in the legend. Note that the time scales are different for the two data sets (as we are plotting the number of snapshots as opposed to the absolute time).

We show that our intrinsically faster methodology is capable of performing rapid rank ordering of binding affinities with comparable quality to traditional MM-PBSA. To compare the overall performance of the two versions of MM-PBSA, we calculate estimates for the binding affinities of a set of 18 urokinase compounds and obtain correlation

coefficients of 0.90 for traditional MM-PBSA, compared to 0.87 for the high-throughput version of MM-PBSA. We found comparable results in our MM-PBSA comparison on a set of 7 biotin-derivative inhibitors of the protein target avidin.³⁶

The computation time for the high-throughput method is on the order of 100 CPU min per structure. This can be contrasted with a computation time for the traditional formulation of MM-PBSA that requires roughly 10 000 min of CPU time per structure. Additionally, as the high-throughput protocol uses an automated procedure for performing system parametrization as well as submission and reconciliation of jobs to and from the grid, it allows for the possibility of addressing larger-scale problems in pharmaceutical drug design, such as systematically processing compound libraries of sizes that require throughputs on the order of thousands of protein-ligand complexes per day. The ability to obtain rapid, quality results makes the high-throughput implementation of MM-PBSA a relevant tool for pharmaceutical drug discovery.

Based on the MM-PBSA results for the limited data set studied here, one might conclude that the calculation of solute entropy is not necessary to perform rank ordering of compounds for a particular protein target. It may be that this will not hold true for even a representative fraction of protein targets; however, on the other hand, it may well be a reasonable approximation for a significant fraction of pharmaceutically relevant disease targets. In any case, obtaining quality estimates of solute entropy is certainly of interest, and recent work in the field may lead to formalisms for calculating reasonable estimates of solute entropy.^{77,78}

Finally, the MM-PBSA methodology we use here may also be useful for investigating aspects of protein design, e.g.,

by employing a method analogous to Joughin et al.⁷⁹ or perhaps simply as a high-throughput method for delineating native-basin structures.^{25,80} In addition to investigating alternative applications of MM-PBSA, we also plan to perform the calculation on a much larger set of our in-house data for crystal structures and investigate the corresponding predicted correlation to measured binding affinities.

REFERENCES AND NOTES

- Stillinger, F.; Rahman, A. Improved simulation in liquid water by molecular dynamics. *J. Chem. Phys.* **1974**, *60*, 1545–1557.
- Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; Hermans, J. In *Intermolecular Forces*; Reidel: Dordrecht, Holland, 1981; pp 331.
- Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- Jorgensen, W. L.; Madura, J. D. Temperature and size dependence for monte carlo simulations of TIP4P water. *Mol. Phys.* **1985**, *56*, 1381–1392.
- Berendsen, H.; Grigera, J.; Straatsma, T. The missing term in the effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.
- Rick, S. W. Simulation of ice and liquid water over a range of temperatures using the fluctuating charge model. *J. Chem. Phys.* **2001**, *114*, 2276–2283.
- Horn, H. W.; Swope, W. C.; Pitara, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173–99.
- Caldwell, J.; Kollman, P. Structure and properties of neat liquids using non-additive molecular dynamics: Water, methanol and N-methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208–6219.
- Ren, P.; Ponder, J. W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002**, *23* (16), 1497–506.
- Grossfield, A.; Ren, P.; Ponder, J. W. Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.* **2003**, *125* (50), 15671–82.
- Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L. 3rd CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25* (12), 1504–14.
- Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006**, *124* (2), 024503.
- Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, V. I. Water properties from first principles: Simulations by a general-purpose quantum mechanical polarizable force field. *Proc. Natl. Acad. Sci. U.S.A.* **2006**.
- Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99* (8), 2161–2200.
- Roux, B. In *Implicit Solvent Models*; Marcel Dekker: New York, 2001; p 133.
- Simonson, T. Electrostatics and dynamics of proteins. *Rep. Prog. Phys.* **2003**, *66*, 737–787.
- Srinivasan, J.; Cheatham, L.; Thomas E.; 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* (37), 9401–9409.
- Fogolari, F.; Tosatto, S. C. Application of MM/PBSA colony free energy to loop decoy discrimination: toward correlation between energy and root mean square deviation. *Protein Sci.* **2005**, *14* (4), 889–901.
- Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB-(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25* (2), 238–50.
- Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46* (1), 243–53.
- Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48* (12), 4040–8.
- Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43* (20), 3786–91.
- Lee, M. R.; Duan, Y.; Kollman, P. A. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins: Struct., Funct., Genet.* **2000**, *39* (4), 309–16.
- Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123* (22), 5221–30.
- Wang, W.; Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **2000**, *303* (4), 567–82.
- Steinbrecher, T.; Case, D. A.; Labahn, A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J. Med. Chem.* **2006**, *49* (6), 1837–44.
- Pearlman, D. A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48* (24), 7796–807.
- Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* **2006**, *49* (16), 4805–4808.
- Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49* (22), 6596–606.
- 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. 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889–97.
- Rizzo, R. C.; Toba, S.; Kuntz, I. D. A molecular basis for the selectivity of thiadiazole urea inhibitors with stromelysin-1 and gelatinase-A from generalized born molecular dynamics simulations. *J. Med. Chem.* **2004**, *47* (12), 3065–74.
- Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–94.
- Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- Brown, S. P.; Muchmore, S. W. High-Throughput Calculation of Protein-Ligand Binding Affinities: Modification and Adaptation of the MM-PBSA Protocol to Enterprise Grid Computing. *J. Chem. Inf. Model.* **2006**, *46*, 999–1005.
- Andreasen, P. A.; Kjoller, L.; Christensen, L.; Duffy, M. J. The urokinase-type plasminogen activator system in cancer metastasis: a review. *Int. J. Cancer* **1997**, *72* (1), 1–22.
- Duffy, M. J.; Maguire, T. M.; McDermott, E. W.; O'Higgins, N. Urokinase plasminogen activator: a prognostic marker in multiple types of cancer. *J. Surg. Oncol.* **1999**, *71* (2), 130–5.
- Nienaber, V.; Wang, J.; Davidson, D.; Henkin, J. Re-engineering of human urokinase provides a system for structure-based drug design at high resolution and reveals a novel structural subsite. *J. Biol. Chem.* **2000**, *275* (10), 7239–48.
- Wendt, M. D.; Rockway, T. W.; Geyer, A.; McClellan, W.; Weitzberg, M.; Zhao, X.; Mantei, R.; Nienaber, V. L.; Stewart, K.; Klinghofer, V.; Giranda, V. L. Identification of novel binding interactions in the development of potent, selective 2-naphthamide inhibitors of urokinase. Synthesis, structural analysis, and SAR of N-phenyl amide 6-substitution. *J. Med. Chem.* **2004**, *47* (2), 303–24.
- Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl.* **2002**, *41* (15), 2644–76.
- Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268* (5214), 1144–9.
- Froloff, N.; Windemuth, A.; Honig, B. On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. *Protein Sci.* **1997**, *6* (6), 1293–301.
- Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–52.

- (46) Onufriev, A.; Bashford, D.; Case, D. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (47) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23* (14), 1297–304.
- (48) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. 3rd Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25* (2), 265–84.
- (49) Geney, R.; Layten, M.; Gomperts, R.; Hornak, V.; Simmerling, C. Investigation of Salt Bridge Stability in a Generalized Born Solvent Model. *J. Chem. Theory Comput.* **2006**, *2*, 115–127.
- (50) Nymeyer, H.; Garcia, E. Simulation of the folding equilibrium of alpha-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (24), 13934–13939.
- (51) Gilson, M. K.; Sharp, K. A.; Honig, B. H. Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment. *J. Comput. Chem.* **1987**, *9* (4), 327–335.
- (52) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct., Funct., Genet.* **1988**, *4* (1), 7–18.
- (53) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **1991**, *252* (5002), 106–9.
- (54) Hermann, R. B. Theory of Hydrophobic Bonding. II. The Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, *76* (19), 2754–2759.
- (55) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71* (8), 2925–2927.
- (56) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **1991**, *30* (40), 9686–97.
- (57) Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (58) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8* (3), 243–56.
- (59) Case, D. A.; Darden, T. A.; Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California, San Francisco: San Francisco, CA, 2004.
- (60) 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 point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.
- (61) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049–1074.
- (62) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model I: Method. *J. Comput. Chem.* **2000**, *21* (2), 132–146.
- (63) 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* (9), 1157–74.
- (64) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247–60.
- (65) Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86* (1 Pt 1), 67–74.
- (66) Ryckaert, J. P.; Ciccotti, 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.
- (67) Hockney, R. W. The potential calculation and some applications. *Methods Comput. Phys.* **1970**, *9*, 136–211.
- (68) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald—an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (69) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (71) Pastor, R. W.; Brooks, B. R.; Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **1988**, *65*, 1409–1419.
- (72) OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com/> (accessed month year).
- (73) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function of Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22* (6), 608–640.
- (74) Sitkoff, D.; Sharp, K. A.; Honig, B. Correlating solvation free energies and surface tensions of hydrocarbon solutes. *Biophys. Chem.* **1994**, *51* (2–3), 397–409.
- (75) Thain, D.; Tannenbaum, T.; Livny, M. In *Condor and the Grid*; John Wiley & Sons Inc., 2003; p 1080.
- (76) Zhou, R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins: Struct., Funct., Genet.* **2003**, *53* (2), 148–61.
- (77) Woo, H. J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6825–30.
- (78) Chang, C. E.; Potter, M. J.; Gilson, M. K. Calculation of molecular configuration integrals. *J. Phys. Chem. B: Condens. Matter. Mater. Surf. Interfaces Biophys.* **2006**, *110* (13), 7083.
- (79) Joughin, B. A.; Green, D. F.; Tidor, B. Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* **2005**, *14* (5), 1363–9.
- (80) Lee, M. C.; Yang, R.; Duan, Y. Comparison between Generalized-Born and Poisson-Boltzmann methods in physics-based scoring functions for protein structure prediction. *J. Mol. Model.* **2005**, *12*, 101–110.

CI700041J