# Molecular Recognition in a Diverse Set of Protein−Ligand Interactions Studied with Molecular Dynamics Simulations and End-Point Free Energy Calculations

Bo Wang,[‡,§] Liwei Li,[†] Thomas D. Hurley,[†] and Samy O. Meroueh*,[†,‡,§,∥]

[†]Indiana University Department of Biochemistry and Molecular Biology, [‡]Center for Computational Biology and Bioinformatics, [§]Department of Chemistry and Chemical Biology (IUPUI), [∥]Stark Neurosciences Research Institute, Indiana University School of Medicine, 535 Barnhill Drive, Indianapolis, Indiana 46202, United States

Ⓢ *Supporting Information*
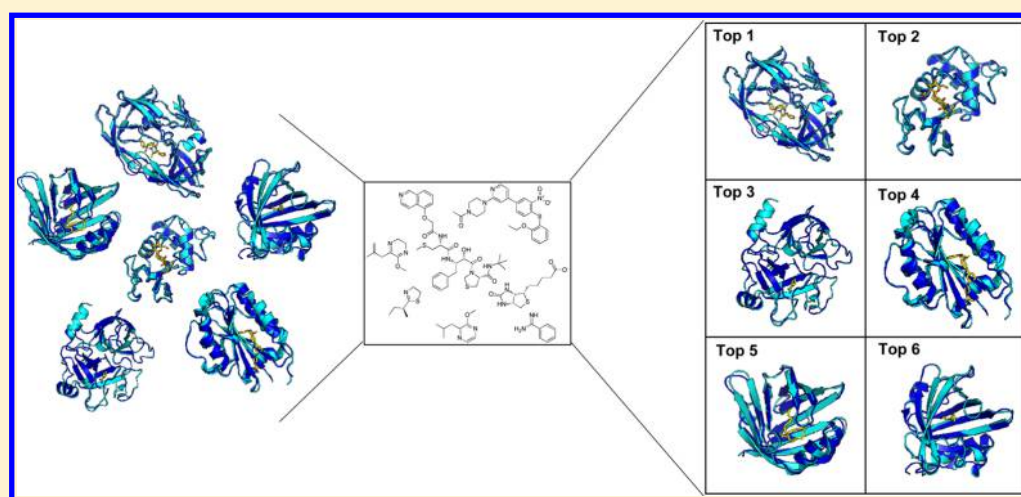
ABSTRACT: End-point free energy calculations using MM-GBSA and MM-PBSA provide a detailed understanding of molecular recognition in protein−ligand interactions. The binding free energy can be used to rank-order protein−ligand structures in virtual screening for compound or target identification. Here, we carry out free energy calculations for a diverse set of 11 proteins bound to 14 small molecules using extensive explicit-solvent MD simulations. The structure of these complexes was previously solved by crystallography and their binding studied with isothermal titration calorimetry (ITC) data enabling direct comparison to the MM-GBSA and MM-PBSA calculations. Four MM-GBSA and three MM-PBSA calculations reproduced the ITC free energy within 1 kcal·mol^{-1} highlighting the challenges in reproducing the absolute free energy from end-point free energy calculations. MM-GBSA exhibited better rank-ordering with a Spearman ρ of 0.68 compared to 0.40 for MM-PBSA with dielectric constant ($\varepsilon = 1$). An increase in ε resulted in significantly better rank-ordering for MM-PBSA ($\rho = 0.91$ for $\varepsilon = 10$), but larger ε significantly reduced the contributions of electrostatics, suggesting that the improvement is due to the nonpolar and entropy components, rather than a better representation of the electrostatics. The SVRKB scoring function applied to MD snapshots resulted in excellent rank-ordering ($\rho = 0.81$). Calculations of the configurational entropy using normal-mode analysis led to free energies that correlated significantly better to the ITC free energy than the MD-based quasi-harmonic approach, but the computed entropies showed no correlation with the ITC entropy. When the adaptation energy is taken into consideration by running separate simulations for complex, apo, and ligand (MM-PBSA$_{ADAPT}$), there is less agreement with the ITC data for the individual free energies, but remarkably good rank-ordering is observed ($\rho = 0.89$). Interestingly, filtering MD snapshots by prescoring protein−ligand complexes with a machine learning-based approach (SVMSP) resulted in a significant improvement in the MM-PBSA results ($\varepsilon = 1$) from $\rho = 0.40$ to $\rho = 0.81$. Finally, the nonpolar components of MM-GBSA and MM-PBSA, but not the electrostatic components, showed strong correlation to the ITC free energy; the computed entropies did not correlate with the ITC entropy.

■ INTRODUCTION

Molecular Dynamics (MD) simulation-based free energy calculations have been used extensively to predict the strength of protein−ligand interactions. Accurate rank-ordering of small molecules bound to protein structures can benefit every step of drug discovery from hit identification to lead optimization. When applied to a compound docked to the human proteome,

free energy calculations can be used for target discovery.[1] Several rigorous methods such as free energy perturbation and thermodynamic integration have been developed for accurate free energy calculations,[2−8] but these methods cannot easily be used for virtual screening of large chemical or combinatorial libraries that typically contain highly diverse compounds.[9] End-point methods such as molecular dynamics (MD)-based MM-GBSA or MM-PBSA[10] offer an alternative to rigorous free energy methods. Structurally diverse molecules can be considered in the calculations. The free energy consists of several terms that include a potential energy, a polar and nonpolar solvation energy, and an entropy.

The MM-GBSA or MM-PBSA free energy consists of several components that can be determined independently. There exists more than one approach for each of these components. For example, the potential energy, which typically includes electrostatic and van der Waals energies, can be obtained using different force fields.[11] The electrostatic component of the solvation energy can be performed using either Poisson−Boltzmann[12] (PB) or Generalized-Born (GB) model.[13] Two approaches are commonly used for the entropy, namely a normal-mode analysis or a quasi-harmonic approximation.[14,15] Finally, the calculations are performed on multiple snapshots collected from MD simulations.[16−18] The selection of different collections of structures is expected to affect the predicted free energy of binding.[19]

Here, we apply MM-GBSA and MM-PBSA calculations to determine the free energy of binding and rank-order a diverse set of protein−ligand complexes. The diversity in the structures of the ligand and targets distinguishes this work from previous efforts that have typically been limited to calculations on congeneric series of compounds on the same target protein. In addition, the use of structures whose binding was characterized with a single method, namely ITC, is expected to reduce the uncertainties in the comparisons between predicted and experimental data. We select 14 protein−ligand structures obtained from the PDBcal database (http://pdbcal.iu.edu) to provide high quality structural and thermodynamic binding data.[20] Extensive explicit-solvent MD simulations were performed, and binding to these proteins was studied using various implementations of MM-GBSA and MM-PBSA. We also tested our previously developed scoring functions for their ability to rank-order complexes by scoring MD structures. The effect of induced-fit conformational changes on rank-ordering these complexes was studied by performing separate simulations for ligand, protein, and protein−ligand complexes. Components of the MM-GBSA and MM-PBSA free energy are compared with the ITC free energy and entropy. To the best of our knowledge, this is the first comparison of MM-GBSA and MM-PBSA calculations to ITC data for a diverse set of proteins.

■ MATERIALS AND METHODS

Scoring Protein−Ligand Complexes. We previously reported the Support Vector Machine target SPecific (SVMSP) model[21] for enriching databases and Support Vector Regression Knowledge-Based (SVRKB) scoring[22] for rank-ordering protein-compound complexes based on their binding affinity. Unlike SVRKB, SVMSP is developed for each individual target protein as described previously.[21] SVMSP model was developed by using protein−ligand crystal structures from the sc-PDB database v2010 for the positive set and randomly selected compounds docked to the target of interest

as the negative set. The positive set was refined by removing crystal structures in which the ligand contains highly charged moieties such as sulfate or phosphate groups, resulting in a set of 4,677 structures. The negative set consisted of docking 5,000 randomly selected compounds from the ChemDiv library to a pocket within the target of interest. The random selection of these compounds from a large chemical library reduces the likelihood that active compounds exist in the negative set.

To develop the SVMSP or SVRKB models, we extended on our previous work on knowledge-based descriptors by using 14 distinct protein atom types and 16 ligand atom types (Table S1).[21] This resulted in 224 atom pair based potentials. We used 76 pair potentials for the vectors of SVMSP. A higher SVMSP score corresponds to a higher probability that the compound is active.

MD Simulations. A set of 14 complexes of small molecules bound to a protein was selected from the PDBcal database (Table 1).[20] The structures of proteins were obtained from RCSB and prepared using BIOPOLYMER in SYBYL 8.0 (Tripos International, St. Louis, Missouri, USA). Hydrogen atoms were added. Missing gaps were modeled. Residue orientation and protonation states were further adjusted using the REDUCE[23] program to optimize the hydrogen bonding network. The ligand structures extracted from crystal structures were prepared and visually checked in SYBYL. The compound was assigned AM1-BCC[24] charges using the *antechamber* program from the AMBER9 package.[25] Water molecules from crystal structures within 5 Å to any atoms on the protein or compound were retained. To perform MD simulations, the protein−ligand complexes were immersed in a box of TIP3P[26] water molecules such that no atom on the protein or ligand was within 14 Å from any side of the box. The solvated box was further neutralized with Na^+ or Cl^- counterions using the *leap* program from the AMBER9 package.

Simulations were carried out using the *pmemd* program in AMBER9 with ff03 force field[27] in periodic boundary conditions. All bonds involving hydrogen atoms were constrained by using the SHAKE algorithm,[28] and a 2 fs time step was used in simulation. The particle mesh Ewald (PME) method was used to treat long-range electrostatics. Simulations were run at 298 K under 1 atm in NPT ensemble employing Langevin thermostat and Berendsen barostat. Water molecules were first energy-minimized and equilibrated by running a short simulation with the complex fixed using Cartesian restraints. This was followed by a series of energy minimizations in which the Cartesian restraints were gradually relaxed from 500 kcal·$Å^{-2}$ to 0 kcal·$Å^{-2}$, and the system was subsequently gradually heated to 298 K via a 48 ps MD run. By assigning different initial velocities, 6 independent simulations in length of 4 ns were performed for each of the crystal structures. The first 2 ns in each trajectory were discarded. MD snapshots were saved every 1 ps yielding 4,000 structures per trajectory.

MD-Based Free Energy Calculations. MM-PBSA and MM-GBSA calculations were carried out as described previously.[29−31] It combines internal energy, solvation energy based on electrostatic and nonpolar contributions, and the entropy. These calculations are carried out on snapshots collected from an MD simulation. The binding free energy is expressed as

$$\Delta G_{\text{MM-PBSA}} = \Delta E_{\text{PBTOT}} - T\Delta S_{\text{NM or QHA}}$$

$$\Delta G_{\text{MM-GBSA}} = \Delta E_{\text{GBTOT}} - T\Delta S_{\text{NM}}$$

Table 1. Calculated Free Energies of Selected Protein−Ligand Complexes (Set 1)

target	complex PDB	$\Delta G_{\text{MM-GBSA}}$ (kcal·mol⁻¹)	$\Delta G_{\text{MM-PBSA}}$ (dielc = 1) (kcal·mol⁻¹)	$\Delta G_{\text{MM-PBSA}}$ (dielc = 2) (kcal·mol⁻¹)	$\Delta G_{\text{MM-SVRKB}}$ (kcal·mol⁻¹)	$\Delta G_{\text{SVMSP//MM-PBSA}}$ (dielc = 1) (kcal·mol⁻¹)	$\Delta G_{\text{SVMSP//MM-PBSA}}$ (dielc = 2) (kcal·mol⁻¹)	ΔG_{ITC} (kcal·mol⁻¹)	ligand no.
human cyclophilin A	1CWA	−17.4 ± 0.7	−13.1 ± 0.8	−36.6 ± 0.7	−7.3	−17.0 ± 0.4	−41.0 ± 0.7	−10.9 ± 0.03	1
HIV-1 protease	1HPV	−18.9 ± 0.8	−8.2 ± 0.8	−38.5 ± 0.8	−10.1	−15.6 ± 0.3	−43.7 ± 0.8	−12.6	2
HIV-1 protease	1HPX	−31.1 ± 0.8	−12.5 ± 0.8	−49.0 ± 0.8	−11.5	−21.1 ± 0.4	−58.9 ± 0.8	−13.3	3
HIV-1 protease	1HXW	−32.4 ± 0.8	−9.9 ± 0.8	−49.3 ± 0.8	−10.1	−24.6 ± 0.4	−65.8 ± 0.8	−13.63 ± 0.07	4
leukocyte function-associated antigen-1	1RD4	−14.9 ± 0.8	−9.8 ± 0.8	−30.9 ± 0.8	−11.1	−17.9 ± 0.2	−39.5 ± 0.8	−10.73	5
porcine odorant-binding protein	1DZK	−6.2 ± 0.6	−5.9 ± 0.6	−13.2 ± 0.6	−5.8	−6.0 ± 0.2	−13.0 ± 0.6	−9.2	6
mouse major urinary protein	1I06	−4.4 ± 0.7	−4.0 ± 0.7	−9.1 ± 0.7	−7.2	−3.2 ± 0.2	−9.0 ± 0.7	−8.38 ± 0.52	7
mouse major urinary protein	1QY1	−7.5 ± 0.7	−7.7 ± 0.7	−13.9 ± 0.7	−6.3	−8.8 ± 0.3	−14.9 ± 0.7	−8.1 ± 0.07	8
DNA gyrase subunit B	1KZN	−24.1 ± 0.9	−4.7 ± 1.0	−41.7 ± 0.9	−8.8	−14.0 ± 0.6	−49.8 ± 0.9	−9.785	9
hen lysozyme C	1LZB	−7.67 ± 0.7	3.6 ± 0.7	−28.5 ± 0.8	−9.0	−5.7 ± 0.3	−39.9 ± 0.7	−7 ± 0.01	10
human galectin-3	1KJL	−13.7 ± 0.6	−8.1 ± 0.7	−22.3 ± 0.6	−5.2	−12.1 ± 0.3	−25.1 ± 0.6	−5.6	11
purine nucleoside receptor A	2FQY	−19.4 ± 1.0	−13.1 ± 1.1	−33.8 ± 1.0	−6.9	−11.3 ± 0.6	−35.3 ± 1.0	−8.81 ± 0.09	12
bovine pancreatic trypsin	1S0R	−6.6 ± 0.8	−11.6 ± 0.9	4.9 ± 0.8	−5.5	−10.9 ± 0.5	5.2 ± 0.8	−6.35 ± 0.07	13
human brain fatty acid-binding protein	1FDQ	−10.4 ± 0.7	−12.8 ± 0.9	−40.3 ± 0.8	−9.8	−14.9 ± 0.6	−41.7 ± 0.7	−10.1	14

where $\Delta G_{\text{MM-PBSA}}$ and $\Delta G_{\text{MM-GBSA}}$ are binding free energies calculated by MM-PBSA and MM-GBSA methods, ΔE_{PBTOT} and ΔE_{GBTOT} are the combined internal and solvation energies, and T is the system temperature. $\Delta S_{\text{NM or QHA}}$ is entropy determined by normal mode calculation or quasi-harmonic analysis. The internal energy is determined using the Lennard-Jones and Coulomb potentials in the Amber force-field (ΔE_{GAS}). The solvation energy is determined using Poisson−Boltzmann or Generalized-Born solvation models (ΔE_{PBSOL} or ΔE_{PBSOL})

$$\Delta E_{\text{PBTOT}} = \Delta E_{\text{PBSOL}} + \Delta E_{\text{GAS}}$$

$$\Delta E_{\text{GBTOT}} = \Delta E_{\text{GBSOL}} + \Delta E_{\text{GAS}}$$

where ΔE_{PBSOL} and ΔE_{GBSOL} are the solvation free energies calculated with PB and GB models, and ΔE_{GAS} represents the molecular mechanical energies. The molecular mechanical energies are composed of three components

$$\Delta E_{\text{GAS}} = \Delta E_{\text{ELE}} + \Delta E_{\text{VDW}} + \Delta E_{\text{INT}}$$

where ΔE_{ELE} is the nonbonded electrostatic energy, ΔE_{VDW} is nonbonded van der Waals energy, and ΔE_{INT} is the internal energies composed of bond, angle, and dihedral energies. The solvation free energies can be calculated using the PB or GB model, expressed respectively by

$$\Delta E_{\text{PBSOL}} = \Delta E_{\text{PBSUR}} + \Delta E_{\text{PBCAL}}$$

$$\Delta E_{\text{GBSOL}} = \Delta E_{\text{GBSUR}} + \Delta E_{\text{GB}}$$

where ΔE_{PBSUR} and ΔE_{GBSUR} are hydrophobic contribution to desolvation energy, and ΔE_{PBCAL} and ΔE_{GB} are reaction field energies.[32]

All the binding energies are determined by

$$\Delta E = \Delta E^{\text{PL}} - \Delta E^{\text{P}} - \Delta E^{\text{L}}$$

where E^{PL}, E^{P}, and E^{L} are total energies corresponding to protein−ligand complex (PL), protein (P), and ligand (L), respectively.

The molecular mechanical gas phase energies were calculated by the *sander* program from the AMBER9 package, including the internal energy, van der Waals and electrostatic interactions. The dielectric constant for electrostatic interactions was set to 1.0. The polar contributions of the solvation free energy were calculated with the Poisson−Boltzmann (PB) method using the *pbsa* program[12] and the Generalized-Born (GB) method implemented in *sander*. The nonpolar contributions of the desolvation energy were determined with solvent-accessible surface area (SASA) dependent terms.[33] The surface area was calculated by the *molsurf* program.[34] The surface tension used to calculate the nonpolar contribution to the free energy of solvation is 0.0072. In the PB method, the reaction field energy was calculated with the dielectric constant for protein and solvent as 1.0 and 80.0, respectively. In the test of the contribution of the dielectric constant, we use various dielectric constants for the solute from 1 to 10, 15, and 20. The default value of the dielectric constant is 1. The solvent probe radius was set to 1.6 Å, which was optimized by Tan and Luo.[35] Atomic radii used were also optimized by Tan and Luo.[35] The calculation based on the GB method was performed with the Onufriev's GB model.[36,37] The SASA calculation was switched to the ICOSA method, surface area was computed by recursively approximating a sphere around an atom, starting from an icosahedra. Two different methods were applied for the
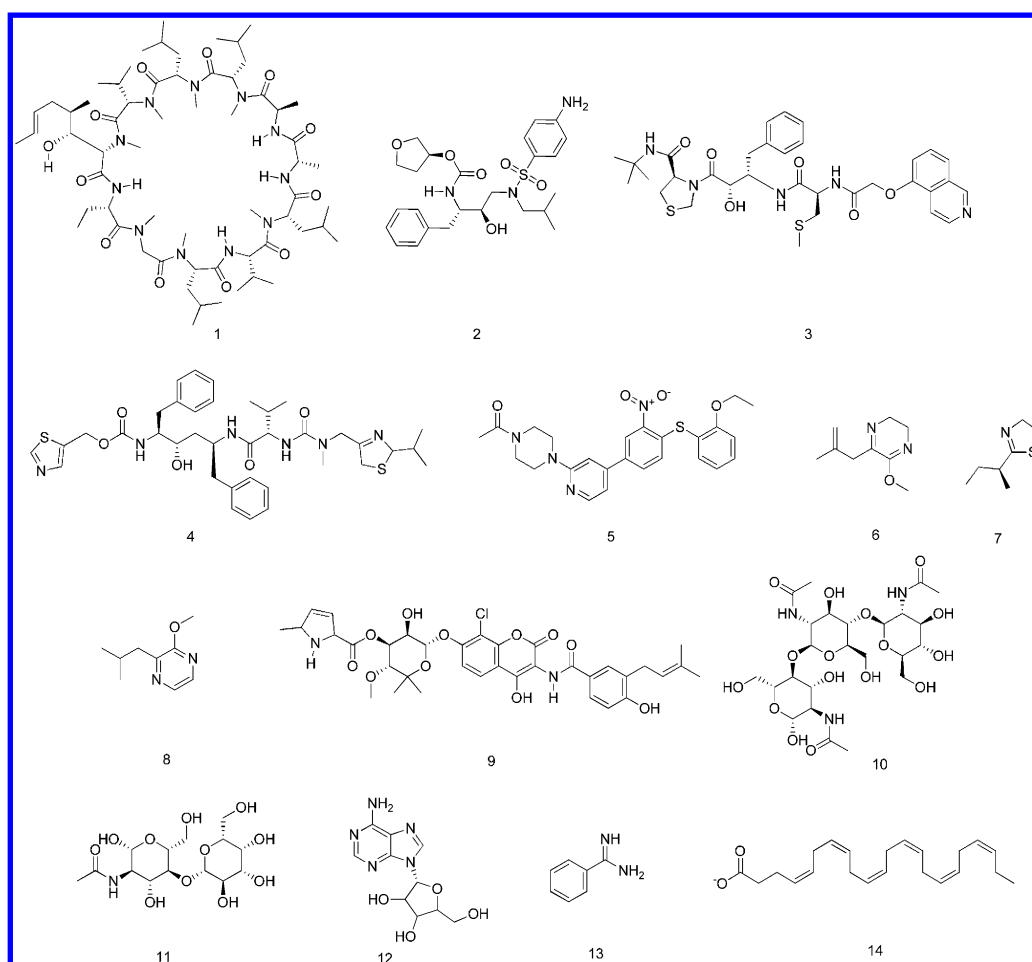
Figure 1. Chemical structure of bound ligands in protein−ligand complexes.

calculation of entropies of the protein−ligand complexes. The quasi-harmonic approximation was analyzed using the *ptraj* program in AMBER. Normal mode conformational entropies were estimated with the *nmode* module from AMBER. The distance-dependent dielectric constant was set to 4. The maximum number of cycles of minimization was set to 10,000. The convergence criterion for the energy gradient to stop minimization was 0.0001. Parameters for the MM-PBSA and MM-GBSA free energy calculation are summarized in Table S2.

For the MM-PBSA or MM-GBSA free energy calculations, a set of 500 structures for each protein−ligand complex was extracted from trajectories of MD simulations at regular intervals. For $\Delta G_{SVMSP//MMPBSA}$ and $\Delta G_{SVMSP//MM\text{-}GBSA}$, all snapshots from MD simulations were first scored by SVMSP. The top scored 500 structures were selected for free energy. For $\Delta G_{MM\text{-}SVRKB}$, all snapshots were scored by SVRKB first, the mean value of the SVRKB score of all snapshots was used to determine the binding affinity (pK_d) of the complex using

$$\Delta G = -2.303RT(pK_d)$$

where R is the gas constant, and T is room temperature (298.15 K).

Correlation Analysis. Three correlation metrics, Pearson's correlation coefficient R_p, Spearman correlation coefficient ρ, and Kendall tau τ, were used in model parametrization and performance assessment. All the correlation analysis was done using packages in R (version 1.12.1). The 95% confidence interval was calculated using 5000 replicate bootstrap sampling.

The Pearson product-moment correlation coefficient R_p is a measure of linear dependence between two variables x and y, giving a value between +1 and −1 inclusive. It was given by

$$R_p = \frac{\Sigma_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_i(x_i - \bar{x})^2 \cdot (y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the mean value for x_i and y_i, respectively. The Spearman's rank correlation coefficient ρ assesses how well the association of two variables can be described using a monotonic function. It was given by

$$\rho = 1 - \frac{6\Sigma_i(x_i' - y_i')^2}{n(n^2 - 1)}$$

where x_i' and y_i' denote the ranks of x_i and y_i, and n is the total number of x-y pairs. A perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other. Kendall tau rank correlation coefficient τ is a measure of the association between two measured quantities. It was given by

$$\tau = \frac{\Sigma_{i<j}sign(x_j - x_i) \cdot sign(y_j - y_i)}{\frac{1}{2}n(n - 1)}$$
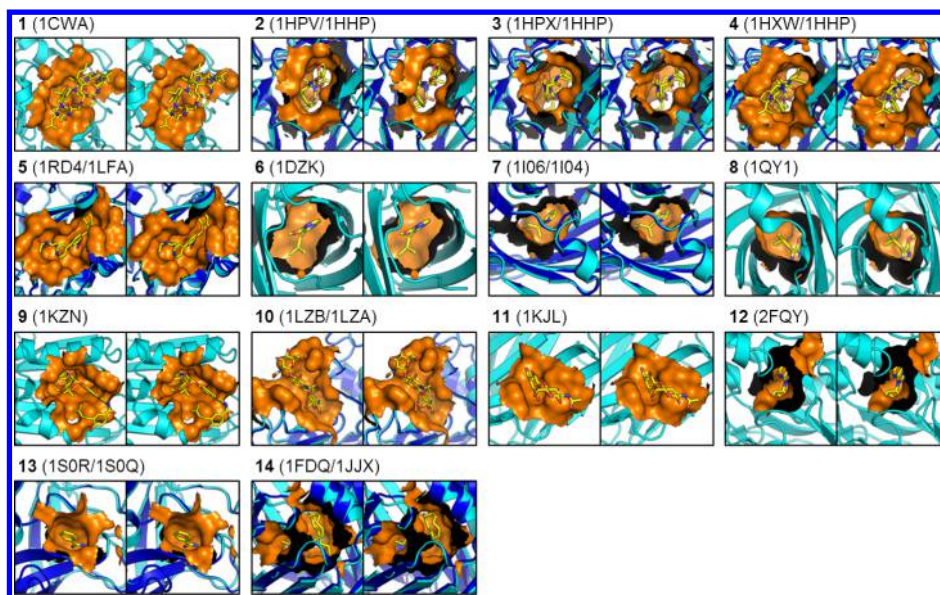
when the values of x_i and y_i are unique.

Figure 2. Stereoview of three-dimensional structures for apo proteins and protein−ligand complexes. The protein structure in complex is illustrated as a cartoon representation in cyan. The apo protein structure (if available) is illustrated as a cartoon representation in blue. Atoms within 5.5 Å of the ligands in complex is considered as pocket atoms and illustrated in a surface representation in orange. The ligand is illustrated as sticks representation in the color-coded by atom types (yellow, red, and blue for carbon, oxygen, and nitrogen). All illustrations were generated by PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

■ RESULTS

Calculations of Binding Free Energies and Comparison to ITC Data. Free energy calculations were carried out for a set of 14 protein−ligand interactions using MM-GBSA and MM-PBSA (Figure 2). The structure of these complexes was previously solved by crystallography and characterization of the binding was done by ITC. The set contains 11 unique proteins and 14 structurally different ligands. The ligands include a cyclic peptide (**1**), peptidomimetics (**2−4**), small organic molecules (**5−10** and **13**), carbohydrates (**10** and **11**), a nucleoside (**12**), and a fatty acid (**14**) (Figure 1). Among the small organic molecules, four were fragment-like (**6−8** and **13**) with molecular weight less than 200 Da. Calculations were carried out using the MM-GBSA and MM-PBSA approach on multiple MD structures collected from 12 ns of simulation. The computed MM-GBSA or MM-PBSA free energies were compared to experimental binding affinity data ΔG_{ITC} (Table 1, Figure 3A). Among the 14 complexes, the predicted $\Delta G_{MM-PBSA}$ were excellent (less than 1 kcal·mol^{-1}) for three of the ligands, namely for (i) **3** binding to HIV-1 protease (PDB code: 1HPX; $|\Delta\Delta G|$ = 0.8); (ii) **8** binding to mouse major urinary protein 1 (PDB code: 1QY1; $|\Delta\Delta G|$ = 0.4); and (iii) binding of **5** to human leukocyte function-associated antigen-1 (PDB code: 1RD4; $|\Delta\Delta G|$ = 0.9). The predicted binding affinities for another five ligands were between 2 and 4 kcal·mol^{-1}, namely for (i) **1** binding to human cyclophilin A (PDB code: 1CWA; $|\Delta\Delta G|$ = 2.2); (ii) **6** binding to porcine odorant-binding protein (PDB code: 1DZK; $|\Delta\Delta G|$ = 2.2); (iii) **14** binding to human brain fatty acid-binding protein (PDB code: 1FDQ; $|\Delta\Delta G|$ = 2.7); (iv) **4** binding to HIV-1 protease (PDB code: 1HXW; $|\Delta\Delta G|$ = 3.7); and (v) **11** bound to human galectin-3 (PDB code: 1KJL). The remaining predicted affinities for compounds **2**, **7**, **9**, **10**, and **13** were larger than 4 kcal·mol^{-1}. An overall measure of the deviation of the MM-PBSA free energy from the ITC free energy is provided by the root-mean-square of the calculated free energy deviation from

experimental energy $\Delta\Delta G_{RMS}$, which was 4.4 kcal·mol^{-1}. The median $\Delta\Delta G$ ($\Delta\Delta G_{MED}$) for MM-PBSA is 3.5. The effect of the dielectric constant on the MM-PBSA calculations was also investigated (Table 5). Doubling the dielectric constant from 1 to 2 resulted in a significantly worse agreement between the MM-PBSA and ITC free energy as evidenced by a 5-fold increase in $\Delta\Delta G_{RMS}$ and a 7-fold increase in ($\Delta\Delta G_{MED}$). This was also observed for calculations performed with larger dielectric constants (Table 5).

The above calculations are repeated using a GB model for the electrostatic solvation free energy (MM-GBSA). MM-GBSA free energies were significantly larger than MM-PBSA free energies. In some cases, MM-GBSA energies exceeded 18 kcal·mol^{-1}. Seven of the MM-GBSA free energies deviated from the ITC free energies by 5 kcal·mol^{-1} compared with only two for MM-PBSA. Overall the MM-GBSA free energy showed greater deviation from the ITC free energy ($\Delta\Delta G_{RMS}$ = 9.2 kcal·mol^{-1}) compared with MM-PBSA ($\Delta\Delta G_{RMS}$ = 4.4 kcal·mol^{-1}). The median $\Delta\Delta G$ for MM-GBSA is 5.2 kcal·mol^{-1}. Despite the large absolute values, MM-GBSA reproduced the free energy of binding remarkably well in four cases with $\Delta\Delta G$ less than 1 kcal·mol^{-1}: (i) **7** binding to the mouse major urinary protein 1 (PDB code: 1QY1; $|\Delta\Delta G|$ = 0.6); (ii) **10** binding to hen lysozyme C (PDB code: 1LZB; $|\Delta\Delta G|$ = 0.7); (iii) **13** binding to the bovine pancreatic trypsin (PDB code: 1S0R; $|\Delta\Delta G|$ = 0.3); and finally (iv) **14** bound to human brain fatty acid-binding protein (PDB code: 1FDQ; $|\Delta\Delta G|$ = 0.3).

Typically, MM-GBSA calculations are carried out by running a single simulation for the complex. Implicit in this approach is that the ligand will only select conformations of the apo protein that are similar to those that are sampled by the protein in the protein−ligand complex. However, there are numerous examples of ligand binding that lead to conformational change of the protein. The free energy of this conformational change, also known as adaptation energy, contributes to the overall free energy of binding.[38] We investigate the role of this adaptation energy for 6 of the 14 complexes (Table 2 and Figure 3B) for
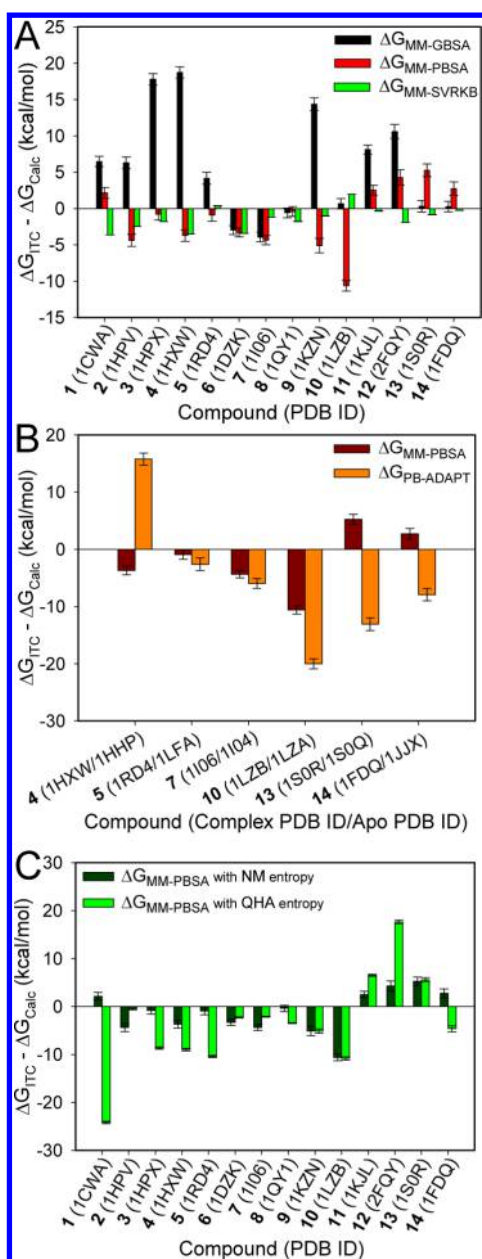
2663

dx.doi.org/10.1021/ci400312v | J. Chem. Inf. Model. 2013, 53, 2659−2670

Figure 3. Calculated free energy deviation from experimental energy in ΔG_{ITC}-ΔG_{Calc} of selected protein−ligand complexes with different methods: (A) $\Delta G_{MM\text{-}GBSA}$, $\Delta G_{MM\text{-}PBSA}$, and $\Delta G_{MM\text{-}SVRKB}$ for complexes in Set 1; (B) $\Delta G_{MM\text{-}PBSA}$ and $\Delta G_{PB\text{-}ADAPT}$ for complexes in Set 2; and (C) $\Delta G_{MM\text{-}PBSA}$ with normal mode calculated entropy and $\Delta G_{MM\text{-}PBSA}$ with quasi-harmonic analysis estimated entropy for complexes in Set 1.

which the crystal structure of the apo was solved independently from the complex structure. Starting with the structure of complex, apo, and ligand, three separate MD simulations were carried out. The root-mean-square deviation (RMSD) of the free protein and ligand were determined with respect to the crystal structure of the protein and ligand in the complex crystal structure (Supporting Information Figure S1). The protein and ligand sampled different structures in the free-state compared to the bound state.

The snapshots from the three separate simulations of complex, apo, and ligand are used to carry out MM-PBSA free energy calculations ($\Delta G_{PB\text{-}ADAPT}$) (Table 2). These are compared with the standard MM-PBSA free energies ($\Delta G_{MM\text{-}PBSA}$) (Table 2, Figure 3B). Overall, the root-mean-square deviation of $\Delta G_{PB\text{-}ADAPT}$ from the ITC free energies is $\Delta\Delta G_{RMS} = 12.4$ kcal·mol^{-1} with a median $\Delta\Delta G$ of 7.6 kcal·mol^{-1} (Table 3). Hence, $\Delta G_{PB\text{-}ADAPT}$ resulted in overall greater

Table 3. Correlation Coefficients for Free Energy Calculations

method	R_P	ρ	τ	$\Delta\Delta G_{RMS}$[a]	$\Delta\Delta G_{MED}$[b]
Set 1[c] Complexes					
MM-GBSA	0.75	0.68	0.52	9.16	5.23
MM-PBSA (dielc = 1)	0.37	0.40	0.25	4.37	3.51
MM-PBSA (dielc = 2)	0.76	0.80	0.60	23.53	23.24
MM-SVRKB	0.77	0.81	0.65	2.09	1.79
SVMSP//MM-GBSA	0.74	0.74	0.56	15.14	10.94
SVMSP//MM-PBSA (dielc = 1)	0.76	0.81	0.63	5.53	4.67
SVMSP//MM-PBSA (dielc = 2)	0.75	0.78	0.60	29.84	29.45
Set 2[d] Complexes					
MM-PBSA$_{ADAPT}$ (dielc = 1)	0.95	0.89	0.73	12.42	7.63
MM-PBSA$_{ADAPT}$ (dielc = 2)	0.92	0.94	0.87	27.70	20.92
MM-GBSA	0.89	0.83	0.73	8.00	2.30
MM-PBSA (dielc = 1)	0.42	0.14	0.20	5.50	4.03
MM-PBSA (dielc = 2)	0.82	0.89	0.73	23.00	20.83
MM-SVRKB	0.74	0.89	0.73	1.76	1.03
Set 1 Complexes[e]					
GBSA	0.44	0.47	0.27	2.32	1.49
PBSA	−0.51	−0.57	−0.45	2.22	1.64
SVRKB	0.83	0.82	0.69	1.43	0.59

[a]Root-mean-square deviation of the calculated free energy deviation from experimental energy. [b]Median of the calculated free energy deviation from experimental energy. [c]Complexes listed in Table 1. [d]Complexes listed in Table 2. [e]Correlation coefficients for free energy calculations with crystal structures.

Table 2. Calculated Free Energies of Selected Protein−Ligand Complexes (Set 2)

			methods					
target	complex PDB	Apo PDB	$\Delta G_{PB\text{-}ADAPT}$ (dielc = 1) (kcal·mol^{-1})	$\Delta G_{PB\text{-}ADAPT}$ (dielc = 2) (kcal·mol^{-1})	$\Delta G_{MM\text{-}PBSA}$ (dielc = 1) (kcal·mol^{-1})	$\Delta G_{MM\text{-}PBSA}$ (dielc = 2) (kcal·mol^{-1})	ΔG_{ITC} (kcal·mol^{-1})	ligand no.
HIV-1 protease	1HXW	1HHP	−29.4 ± 1.1	−67.1 ± 2.8	−9.9 ± 0.8	−49.3 ± 0.8	−13.63 ± 0.07	4
leukocyte function-associated antigen-1	1RD4	1LFA	−8.1 ± 1.1	−35.8 ± 3.1	−9.8 ± 0.8	−30.9 ± 0.8	−10.73	5
mouse major urinary protein	1I06	1I04	−2.4 ± 0.9	−9.7 ± 3.1	−4.0 ± 0.7	−9.1 ± 0.7	−8.38 ± 0.52	7
hen lysozyme C	1LZB	1LZA	13.0 ± 0.9	−22.6 ± 2.3	3.6 ± 0.7	−28.5 ± 0.8	−7 ± 0.01	10
bovine pancreatic trypsin	1S0R	1S0Q	6.7 ± 1.1	15.0 ± 2.7	−11.6 ± 0.9	4.9 ± 0.8	−6.35 ± 0.07	13
human brain fatty acid-binding protein	1FDQ	1JJX	−2.2 ± 1.1	−30.6 ± 2.8	−12.8 ± 0.9	−40.3 ± 0.8	−10.1	14

deviation from the experimental free energy than both MM-GBSA ($\Delta G_{MM\text{-}GBSA}$) and MM-PBSA ($\Delta G_{MM\text{-}PBSA}$). Only one out of the 6 complexes, namely **5** in complex with human leukocyte function-associated antigen-1 (PDB code: 1RD4; $|\Delta\Delta G| = 2.6$), showed reasonable agreement with experiment (<3 kcal·mol^{-1}). The remaining five exhibited binding free energies that were substantially different from the ITC data.

A question of interest is whether scoring functions can generate reliable binding affinities when carried out on multiple structures sampled from MD simulations instead of crystal structures. To address this question, we applied our recently developed scoring function, SVRKB,[22] to snapshots from MD simulations. The empirical scoring function is trained on three-dimensional protein−ligand crystal structures and experimentally measured binding affinity data. SVRKB is used to score MD snapshots of the 14 complexes considered for MM-GBSA and MM-PBSA calculations (Tables 1 and 2). We refer to this approach as MM-SVRKB to emphasize the use of multiple MD structures in the scoring. MM-SVRKB ($\Delta\Delta G_{RMS} = 2.1$ kcal·mol^{-1}) showed better agreement with the experimental free energies than MM-PBSA ($\Delta\Delta G_{RMS} = 4.4$ kcal·mol^{-1}). In fact, $|\Delta\Delta G_{MM\text{-}SVRKB}|$ was less than 2 kcal·mol^{-1} for 10 of the targets, compared with three for the MM-PBSA calculations. None of the predicted MM-SVRKB binding affinities were greater than 5 kcal·mol^{-1} than the experimentally measured affinity.

Finally, we compared calculations performed using harmonic versus quasi-harmonic approaches for the entropy of binding. Two approaches were considered, namely normal-mode analysis, or the use of a quasi-harmonic approach where the entropies are determined by a covariance analysis of the fluctuations obtained from the MD simulations. The MM-PBSA free energies obtained with the normal-mode analysis resulted in a $\Delta\Delta G_{RMS} = 4.4$ kcal·mol^{-1} when compared with the ITC free energy and a median of 3.5 kcal·mol^{-1} for $\Delta\Delta G$ (Figure 3C). On the other hand, the MM-PBSA free energies for the quasi-harmonic approach led to a $\Delta\Delta G_{RMS}$ of 10.1 kcal·mol^{-1} and a median value of 6.1 kcal·mol^{-1}.

Rank-Ordering Protein−Ligand Complexes. Performance to rank-order complexes was evaluated using three correlation metrics, namely the Pearson's correlation coefficient (R_p), Spearman's rho (ρ), and Kendall's tau (τ). Pearson's coefficient is the more traditional metric used to measure the correlation between observed and predicted affinities. Spearman's rho is a nonparametric measure of the correlation between the *ranked lists* of the experimental binding affinities and the scores. It ranges between −1 and 1. A negative value corresponds to anticorrelation, while a positive value suggests correlation between the variables. Kendall's tau (τ) was also considered to assess rank-ordered correlation as suggested by Jain and Nicholls.[39] τ has the advantage of being more robust and can be more easily interpreted. It corresponds to the probability of having the same trend between two rank-ordered lists.

It is interesting that despite the better performance of MM-PBSA in predicting the absolute free energy, the opposite is observed for rank-ordering. All three correlation coefficients metrics were significantly higher for MM-GBSA ($R_p = 0.75$; $\rho = 0.68$; $\tau = 0.52$) compared with MM-PBSA ($R_p = 0.37$; $\rho = 0.40$; $\tau = 0.25$) (Table 3, Figure 4A). At higher dielectric constants, the correlations for MM-PBSA significantly improve (Table 5). A mere doubling of the dielectric constant from 1 to 2 led to a similar increase in the correlation factors ($R_p = 0.77$; $\rho = 0.81$; $\tau = 0.65$). Further increase of the dielectric beyond two results in
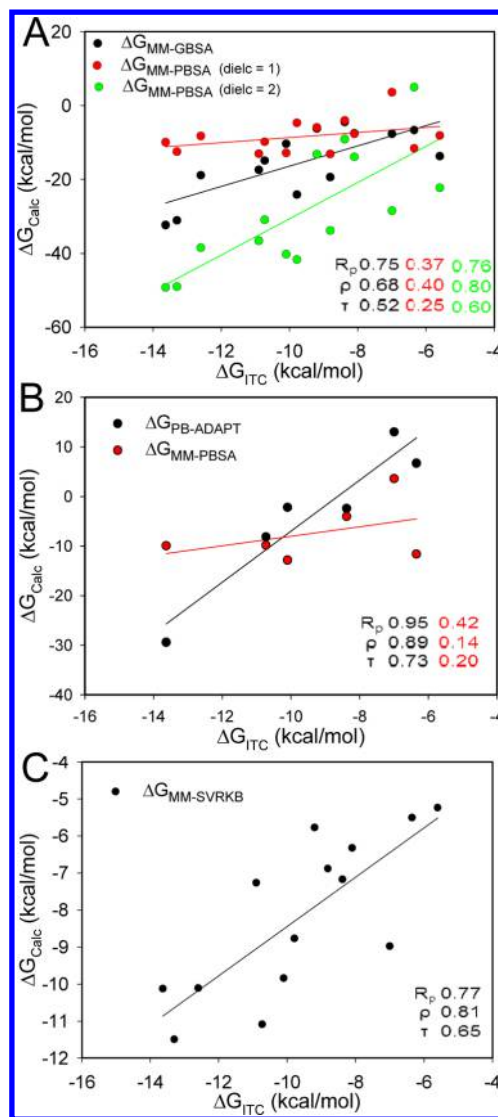

Figure 4. Regression plots between experimental free energy (ΔG_{ITC}) and different calculated free energy of selected protein−ligand complexes: (A) $\Delta G_{MM\text{-}GBSA}$, $\Delta G_{MM\text{-}PBSA}$ (dielc = 1), and $\Delta G_{MM\text{-}PBSA}$ (dielc = 2) for complexes in Set 1; (B) $\Delta G_{MM\text{-}PBSA}$ and $\Delta G_{PB\text{-}ADAPT}$ for complexes in Set 2; and (C) $\Delta G_{MM\text{-}SVRKB}$ for complexes in Set 1.

Table 4. Correlation Coefficients for Components of Free Energy Calculations

component	R_p	ρ	τ
	Set 1 Complexes		
ΔE_{VDW}	0.89	0.90	0.76
ΔE_{NP}	0.88	0.89	0.74
ΔE_{ELE}	0.20	0.13	0.16
ΔE_{PB}	−0.47	−0.46	−0.38
$T\Delta S_{NM}$	−0.63	−0.55	−0.43
$T\Delta S_{QHA}$	−0.47	−0.45	−0.30
$T\Delta S_{NM}^{Lig}$	−0.45	−0.10	−0.07
$T\Delta S_{NM}^{Apo}$	0.02	−0.12	−0.07

smaller increases in performance, as illustrated by the correlations at a dielectric constant of 20 ($R_p = 0.90$; $\rho = 0.91$; $\tau = 0.76$), but inspection of the components of the free energy (Table 6) reveals that this increase in performance is not due to more accurate representation of the electrostatic

Table 5. Correlation Coefficients for MM-PBSA Calculations with Different Dielectric Constants

dielectric constant	R_p	ρ	τ	$\Delta\Delta G_{RMS}$[a]	$\Delta\Delta G_{MED}$[b]
		Set 1 Complexes			
1	0.37	0.40	0.25	4.37	3.51
2	0.76	0.80	0.60	23.53	23.24
3	0.81	0.83	0.65	25.04	23.69
4	0.84	0.85	0.69	24.63	23.13
5	0.85	0.89	0.76	24.03	21.89
6	0.86	0.89	0.76	23.49	20.44
7	0.87	0.89	0.76	23.04	19.29
8	0.88	0.89	0.76	22.67	18.36
9	0.88	0.91	0.78	22.36	17.59
10	0.89	0.91	0.78	22.10	16.96
15	0.89	0.91	0.76	21.26	14.92
20	0.90	0.91	0.76	20.82	13.83

[a]Root-mean-square deviation of the calculated free energy deviation from experimental energy. [b]Median of the calculated free energy deviation from experimental energy.

component of the free energy. Instead, it is attributed to the significantly smaller contribution of the electrostatic energy at higher dielectric constants. An increase in the dielectric constant reduced ΔE_{ELE} and ΔE_{PB} by a factor of $1/\epsilon$ and $1/\epsilon^2$, respectively, where ϵ is the dielectric constant. As a result, the lower contributions from the electrostatic component results in a free energy component that is dominated by the nonpolar and entropy terms. SVRKB applied to MD structures (MM-SVRKB) showed better performance than MM-GBSA ($R_p = 0.77$; $\rho = 0.81$; $\tau = 0.65$) (Figure 4C, Figure 6A). Interestingly, free energies that included the adaptation energy ($\Delta G_{PB-ADAPT}$) exhibited dramatic improvement over MM-PBSA ($R_p = 0.95$; $\rho = 0.89$; $\tau = 0.73$) (Table 3, Set 2, Figure 4B). $\Delta G_{PB-ADAPT}$ correlations are also better than MM-SVRKB ($R_p = 0.74$; $\rho = 0.89$; $\tau = 0.73$).

MM-GBSA and MM-PBSA calculations are performed on multiple structures collected from MD simulations. Typically, snapshots are selected at regular intervals. We wondered how MD snapshots can be prescored to improve the MM-PBSA or MM-GBSA results. We had previously developed a scoring approach (SVMSP) to distinguish between native and non-native binding modes.[21] Scoring of MD snapshots with SVMSP

is expected to enrich these structures for native-like complexes. SVMSP was used to score all snapshots from MD simulations for each of the 14 targets considered in this work. A total of 500 complexes with the top SVMSP scores were selected for MM-GBSA calculations. The combined SVMSP//MM-GBSA scoring did not improve the predictive abilities of MM-GBSA suggesting that the GB method is less sensitive to the structure used in the calculation (Table 3, Figure 6B). However, a dramatic boost in performance is observed for SVMSP//MM-PBSA (Table 3, Figure 6B). In fact, an increase of 0.39, 0.41, and 0.38 is seen for R_p, ρ, and τ, respectively. In Set 2, SVMSP//MMPBSA's prediction of the binding affinity trend is as good as MM-PBSA$_{ADAPT}$. Components of the MM-PBSA or MM-GBSA calculations are insightful as they provide insight into the free energy of binding (Table 6), but an important question is whether these components correlate with the experimentally determined thermodynamic parameters provided by ITC. Each component of the MM-GBSA and MM-PBSA calculations is plotted against the ITC free energy. It was interesting, but not completely surprising,[38,40] that the nonpolar components of the binding affinity correlated with the ITC free energy (Table 4 and Figure 5). The correlation coefficients were $R_p = 0.89$, $\rho = 0.90$, $\tau = 0.76$ and $R_p = 0.88$, $\rho = 0.89$, $\tau = 0.74$ for the van der Waals energy (ΔE_{VDW}) and the nonpolar component of the solvation free energy (ΔE_{NP}), respectively. There was no correlation between the electrostatic components of the free energy (ΔE_{ELE}) and the ITC free energy. There was also no correlation between the reaction field energy calculated by PB (ΔE_{PB}) and the ITC free energy. This is consistent with previous results that showed that the nonpolar component of the free energy was a significantly better predictor of the stability of protein−ligand complexes than the electrostatic component.[38,40] Finally, there was no correlation between the molecular weight of ligand and the binding affinity ($R_p = -0.51$, $\rho = -0.65$, $\tau = -0.51$).

The entropy component of the MM-GBSA and MM-PBSA calculations follows a similar trend to the true entropy of binding. The availability of ITC data for each of our systems provides an opportunity to compare computed versus experimental entropy. For MM-PBSA and MM-GBSA, the entropy is typically determined using either normal mode or a quasi-harmonic analysis. Figure 3C shows that these two

Table 6. Components of Free Energy Calculations

| PDB	ligand no.	ΔE_{ELE}		ΔE_{VDW}	ΔE_{NP}	ΔE_{PB}		$T\Delta S_{NM}$	$T\Delta S_{QHA}$
		dielc = 1	dielc = 2			dielc = 1	dielc = 2		
1CWA	1	−34.5 ± 0.3	−17.2 ± 0.2	−58.1 ± 0.1	−6.7 ± 0.01	54.9 ± 0.3	14.4 ± 0.1	−31.3 ± 0.7	−54.7
1HPV	2	−35.2 ± 0.3	−17.4 ± 0.1	−60.4 ± 0.2	−6.8 ± 0.01	63.4 ± 0.2	15.2 ± 0.0	−30.9 ± 0.8	−14.9
1HPX	3	−43.9 ± 0.4	−21.3 ± 0.2	−70.7 ± 0.2	−8.1 ± 0.01	77.6 ± 0.4	18.0 ± 0.1	−32.7 ± 0.8	−35.8
1HXW	4	−41.8 ± 0.3	−20.9 ± 0.2	−75.2 ± 0.2	−8.5 ± 0.01	79.4 ± 0.3	19.0 ± 0.1	−36.2 ± 0.7	−30.9
1RD4	5	−7.4 ± 0.2	−4.2 ± 0.1	−53.9 ± 0.2	−6.4 ± 0.01	33.3 ± 0.2	8.8 ± 0.1	−24.6 ± 0.8	−41.3
1DZK	6	−3.3 ± 0.1	−1.6 ± 0.0	−27.5 ± 0.1	−3.9 ± 0.01	13.0 ± 0.1	4.0 ± 0.0	−15.8 ± 0.6	−43.1
1I06	7	−1.3 ± 0.1	−0.6 ± 0.0	−21.4 ± 0.1	−3.6 ± 0.01	7.7 ± 0.1	2.2 ± 0.0	−14.6 ± 0.7	−13.4
1QY1	8	−3.5 ± 0.1	−1.7 ± 0.0	−26.9 ± 0.1	−3.9 ± 0.01	10.8 ± 0.1	3.0 ± 0.0	−15.9 ± 0.7	−22.1
1KZN	9	−38.1 ± 0.4	−18.5 ± 0.2	−63.5 ± 0.2	−6.8 ± 0.02	74.2 ± 0.4	17.8 ± 0.1	−29.4 ± 0.9	−43.0
1LZB	10	−57.3 ± 0.9	−28.9 ± 0.4	−38.5 ± 0.3	−5.3 ± 0.02	78.4 ± 0.9	18.7 ± 0.2	−26.5 ± 0.6	−26.9
1KJL	11	−61.6 ± 0.3	−31.2 ± 0.2	−21.2 ± 0.1	−3.7 ± 0.01	57.7 ± 0.3	13.0 ± 0.1	−20.7 ± 0.6	−18.5
2FQY	12	−65.1 ± 0.3	−32.4 ± 0.1	−34.7 ± 0.1	−4.3 ± 0.01	69.7 ± 0.4	16.6 ± 0.1	−21.2 ± 0.9	−35.3
1S0R	13	42.7 ± 0.6	21.0 ± 0.3	−19.3 ± 0.1	−3.0 ± 0.01	−48.9 ± 0.5	−11.2 ± 0.1	−16.9 ± 0.8	−15.2
1FDQ	14	−93.8 ± 1.0	−45.4 ± 0.5	−41.2 ± 0.2	−5.8 ± 0.01	100.7 ± 0.6	24.5 ± 0.1	−27.3 ± 0.6	−24.0
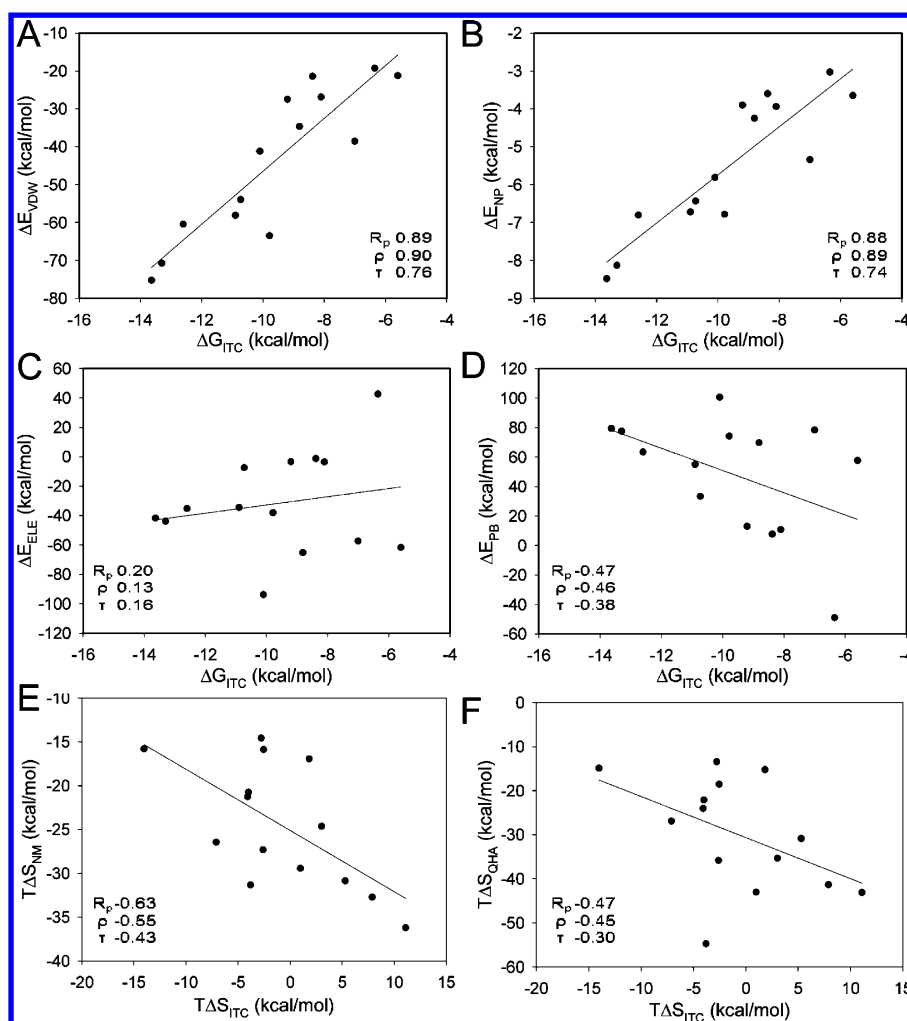
Figure 5. Regression plots between experimental free energy (ΔG_{ITC}) and different component of calculated free energy: (A) ΔE_{VDW}; (B) ΔE_{NP}; (C) ΔE_{ELE}; (D) ΔE_{PB}; (E) $T\Delta S_{NM}$; and (F) $T\Delta S_{QHA}$.

approaches result in different free energies with overall better agreement for the free energy from the normal-mode analysis. The correlations MM-PBSA free energies using normal mode were $R_p = 0.37$, $\rho = 0.40$, $\tau = 0.25$, compared with $R_p = -0.20$, $\rho = -0.30$, $\tau = -0.25$ for the quasi-harmonic analysis. The normal mode and quasi-harmonic entropies are compared to the experimental entropy. A plot of $T\Delta S_{ITC}$ versus $T\Delta S_{NM}$ or $T\Delta S_{QHA}$ shows that computed and experimental entropies are anticorrelated with correlation coefficients of ($R_p = -0.63$; $\rho = -0.55$; $\tau = -0.43$) and ($R_p = -0.47$; $\rho = -0.45$; $\tau = -0.30$), respectively (Figures 5E and 5F). No change is observed when the entropy change of ligand only ($T\Delta S_{NM}^{Lig}$) or receptor only ($T\Delta S_{NM}^{Apo}$) are compared to the ITC entropy (Table 4).

The performance of MM-GBSA and MM-PBSA is compared to GBSA and PBSA, which correspond to calculations performed on a single crystal structure for each of the complexes in Table 1. Correlation coefficients reveal that both GBSA and PBSA perform poorly in rank-ordering complexes when a single crystal structure is used (Figure 6C). For GBSA all three correlation factors were smaller than 0.5 ($R_p = 0.44$; $\rho = 0.47$; $\tau = 0.27$), and for PBSA, predicted and experimental data were anticorrelated ($R_p = -0.51$; $\rho = -0.57$; $\tau = -0.45$). SVRKB, on the other hand, performed well consistent with our previous study[22] ($R_p = 0.83$; $\rho = 0.82$; $\tau = 0.69$).

■ DISCUSSION

MM-GBSA and MM-PBSA calculations are applied to a diverse set of 14 ligands bound to 11 different proteins. A unique aspect of this work is that (i) a diverse set of proteins and ligands are used in contrast to most studies that compare ligands bound to the same protein and (ii) all complexes were previously solved by X-ray crystallography and binding was characterized by ITC. The ligands included small organic compounds, cyclic and linear peptides, fragment-like small molecules, and carbohydrates. Most free energy calculations did not accurately reproduce the ITC free energy, but there were several cases that were in excellent agreement with ITC: three complexes for MM-PBSA calculations and four for MM-GBSA. Overall, MM-PBSA resulted in less deviation from the experimental data than MM-GBSA, but the opposite was observed for rank-ordering. MM-GBSA correlated significantly better with the ITC free energy ($R_p = 0.75$; $\rho = 0.68$; $\tau = 0.52$) when compared to MM-PBSA ($R_p = 0.37$; $\rho = 0.40$; $\tau = 0.25$). The nonpolar terms (ΔG_{VDW} and ΔG_{NP}) showed strong correlation with the experimental free energy ($R_p = 0.89$; $\rho = 0.90$; $\tau = 0.76$ and $R_p = 0.88$; $\rho = 0.89$; $\tau = 0.74$ for ΔG_{VDW} and ΔG_{NP}, respectively). An increase in the dielectric constant for the MM-PBSA calculations worsened agreement of the computed and experimental free energies. However, rank-ordering appeared to significantly improve upon increase of the
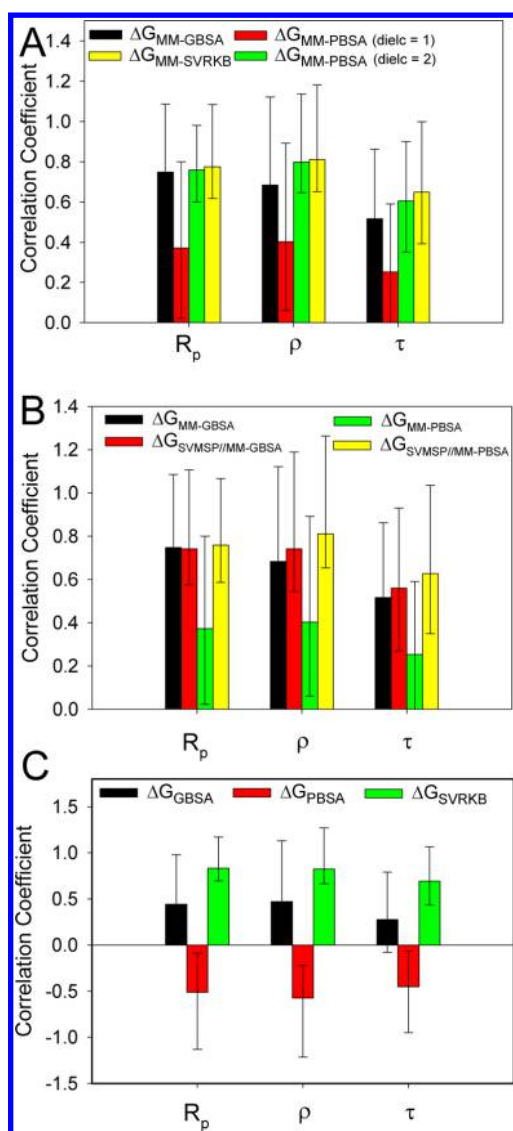
Figure 6. Performance on Pearson's correlation coefficient R_p, Spearman's ρ, and Kendall's τ correlation coefficients: (A) $\Delta G_{MM\text{-}GBSA}$, $\Delta G_{MM\text{-}PBSA}$ (dielc = 1), $\Delta G_{MM\text{-}PBSA}$ (dielc = 2), and $\Delta G_{MM\text{-}SVRKB}$ for multiple MD structures; (B) $\Delta G_{MM\text{-}GBSA}$, $\Delta G_{SVMSP//MM\text{-}GBSA}$, $\Delta G_{MM\text{-}PBSA}$, and $\Delta G_{SVMSP//MM\text{-}PBSA}$ for multiple MD structures; and (C) ΔG_{GBSA}, ΔG_{PBSA}, and ΔG_{SVRKB} for single crystal structure.

dielectric constant, but close inspection of the components of the free energy reveals that this increase is attributed to the lower contribution of electrostatics as a result of an increase of the dielectric constant. The Coulomb and electrostatic terms are inversely proportional to the dielectric constant and to the square of the dielectric constant, respectively. Less contribution from electrostatics leads to a free energy that is dominated by the nonpolar and entropy components resulting to better performance. There was no correlation between the electrostatic components of the MM-GBSA and MM-PBSA free energy and the ITC free energy.

Two models for the entropy were considered: normal mode and quasi-harmonic. Normal mode analysis assumes that each structure is at a potential energy minimum. Quasi-harmonic analysis is based on a covariance analysis of the atomic fluctuation. Our data showed that the free energies using normal-mode analysis correlated significantly better than free energies using quasi-harmonic analysis. A possible explanation

is that the simulations used in this study may not have been sufficiently long to ensure convergence of quasi-harmonic analysis. Neither the normal mode entropy nor the quasi-harmonic entropies correlated with the ITC entropy. This is likely due to the fact that the ITC entropy includes both solvation and configurational entropy,[41] while the computed entropy only includes the configurational entropy. The solvation entropy may be indirectly captured by the other terms of the MM-GBSA or MM-PBSA free energy.

Small-molecule binding often induces conformational change to the target protein. This adaptation energy is often ignored in MM-GBSA or MM-PBSA calculations as a single simulation is carried out starting with the complex structure. We studied the effect of this adaptation energy by running separate simulations for ligand, apo, and complex structures. We did this for 6 of the 14 complexes whose apo structure was solved independently by X-ray crystallography. Overall, this resulted in poorer agreement with the ITC data when comparing the absolute values of the free energies. However, the adaptation energy resulted in a significant boost in rank-ordering. The $\Delta G_{PB\text{-}ADAPT}$ resulted in a Pearson (Spearman) correlation of 0.95 (0.89) compared with 0.89 (0.83) for $\Delta G_{MM\text{-}GBSA}$ and 0.42 (0.14) for $\Delta G_{MM\text{-}PBSA}$. The $\Delta G_{PB\text{-}ADAPT}$ showed the best rank-ordering among all methods that were tested in this work.

Typically snapshots for MM-GBSA or MM-PBSA calculations are selected at regular intervals in an MD simulation. We wondered whether different approaches for selecting structures will influence the free energy of binding. We used a recently developed machine learning-based scoring approach (SVMSP) to prescore all the snapshots in a trajectory. SVMSP is trained using crystal protein−ligand and docked protein-decoy structures.[21] It is therefore expected that the method will enrich MD snapshots for native-like structures. Rank-ordering of snapshots (Table 3) had little influence on the MM-GBSA free energies (Table 3). However, rank-ordering with MM-PBSA calculations improved significantly from $R_p = 0.37$, $\rho = 0.40$, $\tau = 0.25$ for snapshots selected at regular intervals to $R_p = 0.76$, $\rho = 0.81$, $\tau = 0.63$ for SVMSP-selected snapshots. These results indicate that the Poisson−Boltzmann calculations are more sensitive to the quality of the structure than MM-GBSA.

In sum, MM-GBSA and MM-PBSA methods come short in reliably reproducing the free energy of binding. However, these methods can perform remarkably well for rank-ordering diverse set of compounds. MM-GBSA can perform well by merely using snapshots from an MD simulation of the complex, while MM-PBSA is significantly more sensitive to the structures used. Filtering MD structures with scoring functions to enrich for native-like complexes results in excellent rank-ordering by MM-PBSA. In addition, running separate simulations of the receptor also improves the rank-ordering abilities of MM-PBSA. While previous studies have found that rank-ordering performance for MM-PBSA improves with increasing dielectric constant, we found that this is mainly due to the smaller contributions of electrostatics as a result of increasing the dielectric constant (at $\varepsilon = 5$, for example, the Coulomb energy is reduced by a factor of 1/5 and the PB solvation energy by 1/25). It was remarkable that the nonpolar components correlated very well with the free energy. The combination of nonpolar and entropy also correlated very well with the free energy, which is why overall correlation improved at higher dielectric constants for MM-PBSA. Finally, the MM-PBSA entropy does not correlate with the ITC entropy.

ASSOCIATED CONTENT

ⓢ Supporting Information

Tables listing parameters that were used for the SVMSP and SVRKB scoring as well as MM-PBSA and MM-GBSA calculations. Root-mean-square deviations are also provided for the structures subjected to molecular dynamics simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 317-274-8315. Fax: 317-278-9217. E-mail: smeroueh@iupui.edu. Corresponding author address: Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W. 10th Street, HITS 5000, Indianapolis, IN 46202.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Li, L.; Li, J.; Khanna, M.; Jo, I.; Baird, J. P.; Meroueh, S. O. Docking to erlotinib off-targets leads to inhibitors of lung cancer cell proliferation with suitable in vitro pharmacokinetics. *ACS Med. Chem. Lett.* **2010**, *1*, 229–233.

(2) De Ruiter, A.; Oostenbrink, C. Efficient and accurate free energy calculations on trypsin inhibitors. *J. Chem. Theory Comput.* **2012**, *8*, 3686–3695.

(3) Wan, S.; Coveney, P. V.; Flower, D. R. Peptide recognition by the T cell receptor: Comparison of binding free energies from thermodynamic integration, Poisson-Boltzmann and linear interaction energy approximations. *Philos. Trans. R. Soc., A* **2005**, *363*, 2037–2053.

(4) Golemi-Kotra, D.; Meroueh, S. O.; Kim, C.; Vakulenko, S. B.; Bulychev, A.; Stemmler, A. J.; Stemmler, T. L.; Mobashery, S. The importance of a critical protonation state and the fate of the catalytic steps in class A β-lactamases and penicillin-binding proteins. *J. Biol. Chem.* **2004**, *279*, 34665–34673.

(5) Oostenbrink, C. B.; Pitera, J. W.; Van Lipzig, M. M. H.; Meerman, J. H. N.; Van Gunsteren, W. F. Simulations of the estrogen receptor ligand-binding domain: Affinity of natural ligands and xenoestrogens. *J. Med. Chem.* **2000**, *43*, 4594–4605.

(6) Lawrenz, M.; Wereszczynski, J.; Ortiz-Sánchez, J.; Nichols, S.; McCammon, J. A. Thermodynamic integration to predict host-guest binding affinities. *J. Comput.-Aided. Mol. Des.* **2012**, *26*, 569–576.

(7) Steinbrecher, T.; Case, D. A.; Labahn, A. Free energy calculations on the binding of novel thiolactomycin derivatives to E. coli fatty acid synthase I. *Bioorg. Med. Chem.* **2012**, *20*, 3446–3453.

(8) Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A. Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2523–2532.

(9) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.

(10) Chong, L. T.; Dempster, S. E.; Hendsch, Z. S.; Lee, L.-P.; Tidor, B. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Sci.* **1998**, *7*, 206–210.

(11) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(12) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2002**, *23*, 1244–1253.

(13) Still, C. W.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(14) Karplus, M.; Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, *14*, 325–332.

(15) 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*, 5221–5230.

(16) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

(17) Basdevant, N.; Weinstein, H.; Ceruso, M. Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J. Am. Chem. Soc.* **2006**, *128*, 12766–12777.

(18) Grünberg, R.; Nilges, M.; Leckner, J. Flexibility and conformational entropy in protein-protein binding. *Structure* **2006**, *14*, 683–693.

(19) Lill, M. A.; Thompson, J. J. Solvent interaction energy calculations on molecular dynamics trajectories: Increasing the efficiency using systematic frame selection. *J. Chem. Inf. Model.* **2011**, *51*, 2680–2689.

(20) Li, L.; Dantzer, J. J.; Nowacki, J.; O'Callaghan, B. J.; Meroueh, S. O. PDBcal: A comprehensive dataset for receptor-ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. *Chem. Biol. Drug Des.* **2008**, *71*, 529–532.

(21) Li, L.; Khanna, M.; Jo, I.; Wang, F.; Ashpole, N. M.; Hudmon, A.; Meroueh, S. O. Target-specific support vector machine scoring in structure-based virtual screening: Computational validation, in vitro testing in kinases, and effects on lung cancer cell proliferation. *J. Chem. Inf. Model.* **2011**, *51*, 755–759.

(22) Li, L.; Wang, B.; Meroueh, S. O. Support vector regression scoring of receptor–ligand complexes for rank-ordering and virtual screening of chemical libraries. *J. Chem. Inf. Model.* **2011**, *51*, 2132–2138.

(23) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

(24) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(25) 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.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S. R.; Tsui, V.; Schafmeister, H.; Ross, W. S.; Kollman, P. A. *AMBER9*; University of California: San Fransico, 2006.

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

(27) 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*, 1999–2012.

(28) 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.

(29) Srinivasan, J.; Cheatham, T. 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*, 9401−9409.

(30) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

(31) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.

(32) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144−1149.

(33) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(34) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.

(35) Tan, C.; Yang, L.; Luo, R. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *J. Phys. Chem. B* **2006**, *110*, 18680−18687.

(36) 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*, 383−394.

(37) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.

(38) Liang, S.; Li, L.; Hsu, W.-L.; Pilcher, M. N.; Uversky, V.; Zhou, Y.; Dunker, K. A.; Meroueh, S. O. Exploring the molecular design of protein interaction sites with molecular dynamics simulations and free energy calculations. *Biochemistry* **2008**, *48*, 399−414.

(39) Jain, A.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133−139.

(40) Li, L.; Liang, S.; Pilcher, M. M.; Meroueh, S. O. Incorporating receptor flexibility in the molecular design of protein interfaces. *Protein Eng., Des. Sel.* **2009**, *22*, 575−586.

(41) Lee, K. H.; Xie, D.; Freire, E.; Amzel, M. L. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. *Proteins: Struct., Funct., Bioinf.* **1994**, *20*, 68−84.