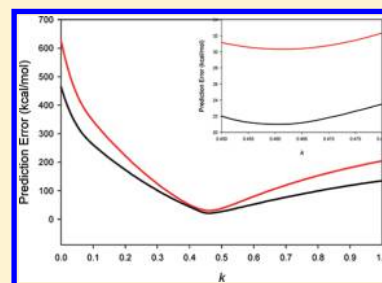


Develop and Test a Solvent Accessible Surface Area-Based Model in Conformational Entropy Calculations

Junmei Wang^{*,†} and Tingjun Hou[‡][†]Department of Biochemistry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390, United States[‡]Functional Nano & Soft Materials Laboratory (FUNSOM), Soochow University, 199 Ren'ai Road, Suzhou 215123, P. R. China

Supporting Information

ABSTRACT: It is of great interest in modern drug design to accurately calculate the free energies of protein–ligand or nucleic acid–ligand binding. MM-PBSA (molecular mechanics Poisson–Boltzmann surface area) and MM-GBSA (molecular mechanics generalized Born surface area) have gained popularity in this field. For both methods, the conformational entropy, which is usually calculated through normal-mode analysis (NMA), is needed to calculate the absolute binding free energies. Unfortunately, NMA is computationally demanding and becomes a bottleneck of the MM-PB/GBSA-NMA methods. In this work, we have developed a fast approach to estimate the conformational entropy based upon solvent accessible surface area calculations. In our approach, the conformational entropy of a molecule, S , can be obtained by summing up the contributions of all atoms, no matter they are buried or exposed. Each atom has two types of surface areas, solvent accessible surface area (SAS) and buried SAS (BSAS). The two types of surface areas are weighted to estimate the contribution of an atom to S . Atoms having the same atom type share the same weight and a general parameter k is applied to balance the contributions of the two types of surface areas. This entropy model was parametrized using a large set of small molecules for which their conformational entropies were calculated at the B3LYP/6-31G* level taking the solvent effect into account. The weighted solvent accessible surface area (WSAS) model was extensively evaluated in three tests. For convenience, TS values, the product of temperature T and conformational entropy S , were calculated in those tests. T was always set to 298.15 K through the text. First of all, good correlations were achieved between WSAS TS and NMA TS for 44 protein or nucleic acid systems sampled with molecular dynamics simulations (10 snapshots were collected for postentropy calculations): the mean correlation coefficient squares (R^2) was 0.56. As to the 20 complexes, the TS changes upon binding; $T\Delta S$ values were also calculated, and the mean R^2 was 0.67 between NMA and WSAS. In the second test, TS values were calculated for 12 proteins decoy sets (each set has 31 conformations) generated by the Rosetta software package. Again, good correlations were achieved for all decoy sets: the mean, maximum, and minimum of R^2 were 0.73, 0.89, and 0.55, respectively. Finally, binding free energies were calculated for 6 protein systems (the numbers of inhibitors range from 4 to 18) using four scoring functions. Compared to the measured binding free energies, the mean R^2 of the six protein systems were 0.51, 0.47, 0.40, and 0.43 for MM-GBSA-WSAS, MM-GBSA-NMA, MM-PBSA-WSAS, and MM-PBSA-NMA, respectively. The mean rms errors of prediction were 1.19, 1.24, 1.41, 1.29 kcal/mol for the four scoring functions, correspondingly. Therefore, the two scoring functions employing WSAS achieved a comparable prediction performance to that of the scoring functions using NMA. It should be emphasized that no minimization was performed prior to the WSAS calculation in the last test. Although WSAS is not as rigorous as physical models such as quasi-harmonic analysis and thermodynamic integration (TI), it is computationally very efficient as only surface area calculation is involved and no structural minimization is required. Moreover, WSAS has achieved a comparable performance to normal-mode analysis. We expect that this model could find its applications in the fields like high throughput screening (HTS), molecular docking, and rational protein design. In those fields, efficiency is crucial since there are a large number of compounds, docking poses, or protein models to be evaluated. A list of acronyms and abbreviations used in this work is provided for quick reference.



1.0. INTRODUCTION

1.1. MM-PBSA and MM-GBSA Methods. Free energy prediction is one of the central interests in computational chemistry. With continually increasing computer power, the calculations of free energy changes of biological events, such as protein folding, protein–ligand binding, and protein–protein association, have become popular. The methods of calculating free energies and entropies through molecular simulations have been summarized in many reviews.^{1–7} Among a

variety of approaches, MM-PBSA and MM-GBSA have gained popularity owing to their computational efficiency in comparison to the theoretically more rigorous methods such as free energy perturbation and thermodynamic integration.^{1,8–16} Critical assessment of the two techniques on modeling protein–ligand binding begins to emerge.^{17–22}

Received: February 2, 2012

Published: April 13, 2012

In a typical MM-PB/GBSA calculation, the molecular system of interest is first immersed in a water box and up to tens of nanoseconds of molecular dynamics simulations are performed using either the “single trajectory” or the “individual trajectories” protocol.³ In most scenarios, the “single trajectory protocol” is superior to the “individual trajectories protocol” as the former can achieve a better error cancellation. However, when the binding leads to a dramatic conformational change, the ensembles of the receptor, ligand, and complex need to be sampled separately (the “individual trajectories” protocol) since the conformational energies and entropies of the receptor and the ligand are significantly different in bound and unbound states.

In the second stage, post-free energy analysis is carried out after peeling off the solvent and counterions. In MM-PB/GBSA theory, the free energy of a molecule is calculated with eqs 1–3.

$$G = \langle H_{\text{gas}} \rangle + \langle G_{\text{solv}} \rangle - T \langle S_{\text{conf}} \rangle \quad (1)$$

$$G_{\text{solv}} = G_{\text{solv}}^{\text{pol}} + G_{\text{solv}}^{\text{nonpol}} \quad (2)$$

$$S_{\text{conf}} = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} \quad (3)$$

The first term in eq 1, H_{gas} , is replaced with E_{gas} , the gas phase MM energy, as the PV term is negligible for a molecule in condensed phase. $\langle \rangle$ indicates that those energy terms are ensemble averages. The second term in eq 1, G_{solv} , consists of two components, the polar and nonpolar solvation free energies (eq 2). The polar solvation energy is evaluated either by a PB or a GB model, while the nonpolar solvation free energy is typically estimated with the solvent accessible surface area assuming that the nonpolar contribution is proportional to SAS. The nonpolar term accounts for the entropy penalty associated with the reorganization of solvent molecules around the solute and the van der Waals interaction between the solute and solvent. The third term in eq 1, the conformational entropy, is further decomposed into three parts, the translational, the rotational, and the vibrational entropies (eq 3). The translational entropy (S_{trans}) and the rotational entropy (S_{rot}) can be approximated using the standard equations for rigid body translation and rotation,²³ and the vibrational part of conformational entropy (S_{vib}) is typically estimated by normal-mode analysis assuming that the vibrational movement around the energy well is harmonic. The S_{vib} term can also be obtained by conducting a quasi-harmonic analysis using the MD trajectories in the sampling phase. As we propose to predict the conformational entropic term through solvent accessible surface area calculation, in practice, as long as the two SAS calculations using the same radii, S_{conf} and $G_{\text{solv}}^{\text{nonpolar}}$ should be combined to avoid double work.

Although successful stories of using MM-PB/GBSA for various biological systems have been reported, both accuracy and efficiency needs to improve for this promising free energy method. As there are many energy terms involved, simply improving one specific term, such as polar and nonpolar components of solvation energy or conformational entropy, may not necessarily improve the overall accuracy of this method. Thus, a step-by-step procedure must be applied to systematically improve the accuracies of all the terms.³ A good molecular mechanical force field for studying both biological and organic molecules with a high quality solvation model is mainly responsible for the success of this method. Certainly it is a challenging and time-consuming procedure to develop a good molecular mechanical force field and an implicit solvation model.

To improve the efficiency of the MM-PB/GBSA method, actions may be taken in both the sampling and the post free energy analysis phases. First of all, one may run MD simulations using an implicit water model, such as GBSA and PBSA, instead of an explicit water model. Although for large molecules, solving the Poisson–Boltzmann equation takes long time, the bottleneck of post free energy phase is to calculate the conformational entropy by NMA.

1.2. Normal Mode Analysis. To conduct normal mode analysis, a molecule must be fully minimized in order to make the harmonic assumption valid. Otherwise, the calculation result is meaningless or has a large computational error. To guarantee a real local minimum or the global minimum is found, a second derivative-based approach, such as Newton–Raphson, is usually applied to minimize the rms gradient to a very low value; otherwise, substantial errors can occur.^{24,25} After that, the mass-weighted Hessian matrix is diagonalized and the thermochemical properties are calculated in the same way as frequently done in quantum mechanics. Both the geometrical optimization and normal mode analysis are time-consuming and computer memory demanding for large biological molecules. For example, the CPU time for running conjugate gradient minimization followed by a Newton–Raphson minimization (the convergence criterion is the root-mean-square of gradient less than 1.0×10^{-12} kcal/(mol Å)) are 66, 549, and 1689 min for 1BE9 (120 residues), 8GCH (238 residues), and 1A9U (352 residues), respectively. The CPU time for Hessian matrix diagonalization are 99, 701, and 3037 s for the three proteins, correspondingly. The above benchmark is based on Intel Xeon X5660 (2.80 GHz) CPU. Apparently, full minimization for a middle-sized or large-sized protein is very expensive.

In order to break this bottleneck, a simplified computational strategy was proposed and routinely used in MM-PB/GBSA calculations: only residues within a short distance to the bound ligand, say 8 Å are kept and the other less important residues are truncated. The remaining residues, which are referred to as the active residues, are fully minimized using a distance-dependent dielectric constant.^{13,14,19,26} A big problem with this truncating strategy is that full minimization may lead to significant conformational changes in the molecular geometry. Kongsted and Ryde recently proposed to add a buffer region surrounding the active residues to maintain the overall molecular geometry and to reduce the standard deviation of the conformational entropy.²⁶ Nevertheless, as we will show later, the truncating strategy tends to underestimate the entropy changes of binding in our tests and should be used with caution.

1.3. Conformation Entropy Calculation. The major goal of this paper is to develop an efficient conformational entropy model for free energy calculations. In the following part of the introduction, we will first elaborate the definitions of conformational entropy and outline the conformational entropy calculations using various methods.

For a molecule in aqueous solution, the total entropy comes from three major contributions, namely, the external entropy due to the translational and rotational degrees of freedom, the internal entropy rooted in vibrational movement, and the entropy due to solvation. The solvation entropy along with the van der Waals interaction between the solute and solvent is well-predicted using the solvent accessible surface area as discussed above.²⁷ The vibrational part of conformational entropy is usually evaluated through normal-mode analysis, and the translational and rotational parts are estimated using rigid body/rotor approximation. Conformational entropy is also called configurational

entropy in many publications;^{2,28–31} here, we choose to use conformational entropy rather than configurational entropy in order to be consistent with our earlier publications.^{3,14,16} In this work, conformational entropy and configurational entropy are used interchangeably. It should be noted that in some papers conformational entropy is a measure of the number of occupied energy wells and has a different meaning from the conformational entropy defined here.³¹

Conformational entropy is an important term in the MM-PB/GBSA scoring function. In most cases of the protein–protein, protein–ligand binding, binding affinity is a combined function of binding enthalpy and binding entropy and high affinity is achieved when both terms contribute favorably to the binding. However, when the binding is dominated by entropic effect, conformational entropy becomes an indispensable term. A scoring function without an entropic term has a biased tendency to select large molecules in virtual screenings.^{32–34}

In molecular docking for which computation efficiency is crucial, conformational entropy is either omitted or simply estimated by the number of frozen rotatable bonds upon ligand binding.^{35–37} This simple entropy model has been further improved in some docking scoring functions so that the entropy loss of a given rotor depends on its environment.^{38,39} However, the chemical nature of rotors which can also affect the entropy is not differentiated. Evidently, entropy models based on the number of frozen rotatable bonds are very crude and they have no discrimination power at all when the major contribution of conformational entropy results from the narrower energy well in the bound state.² Gilson et al. found that, for a set of small host–guest systems, there was no clear correlation between entropy change and the number of rotatable bonds but the snugness of the guest's fit in the host's binding site correlated with entropy loss.^{40,41} In another study, Kongsted and Ryde found that the numbers of frozen rotatable bonds have no correlation to the conformational entropies calculated by normal mode analysis for biotin/avidin binding system.²⁶ In another efficient method, Abagyan et al. applied solvent accessible surface area to account for the conformational entropy of the side chains of proteins in conformational searches of peptides and proteins.⁴² However, their simple model may not be suitable to study protein–ligand binding.

More rigorous methods for calculating configurational entropy include quasi-harmonic analysis (QHA),⁴³ M2 (second generation mining minima) of Gilson et al.,^{44–46} FEP and TI through temperature-derivative calculations,⁴⁷ nonparametric methods using histogram⁴⁸ and *k*-nearest neighbor (kNN),^{28,49} hypothetical scanning approach,^{50,51} the adaptive anisotropic kernels and minimum information methods of Grubmüller et al.,^{52,53} and so on.

All the above-mentioned methods except the first two are time-consuming and not suitable to be implemented in docking scoring functions or to be used in high throughput screenings. Wlodek et al. recently proposed to use the Hessian matrix generated by the quasi-Newton optimizers rather than the exact analytical Hessian matrix calculated for the optimized compounds to calculate the vibrational entropy.⁵⁴ However, their method still requires massive structural minimization and therefore is also not suitable to be used in docking studies. To bridge the gap between the efficient but not accurate and the accurate but not efficient approaches, in this work, we proposed a very fast yet reasonably accurate approach for calculating conformational entropy based on solvent accessible surface area calculations. This new model can be used to estimate the conformational

entropies of protein models and rerank docking poses. It can also be integrated into a docking scoring function.

2. METHODOLOGY

2.1. Basic Theory. It is known that entropy is an additive property under some approximation, and as such, it is possible to calculate the conformational entropy of a molecule *S* by adding up the contributions of individual atoms. The interior atoms of a molecule are not fully free to move; therefore, they have less contribution to *S* than the exposed atoms. However, it is not proper to apply only solvent accessible surface area to measure an atom's contribution to *S* since a fully buried atom has zero SAS but it can still make non-negligible contributions to *S*. For small molecules, this problem may not be severe as most atoms are exposable. However, for macromolecules, omitting the contribution of buried atoms could lead to a significant underestimation of the conformational entropy *S* since a larger portion of atoms of macromolecules are not accessible to solvent. To address the above problem, we propose to use the following equations to calculate conformational entropy:

$$S = \sum_{i=1}^N w_i (\text{SAS}_i + k \text{BSAS}_i) \quad (4)$$

$$\text{BSAS}_i = 4\pi(r_i + r_{\text{prob}})^2 - \text{SAS}_i \quad (5)$$

Where, w_i is the weight for atom *i*; SAS_i is the solvent accessible surface area of atom *i*; *N* is the number of atoms in a molecule; BSAS_i , the buried solvent accessible surface area of atom *i*, is calculated using eq 5 (note that BSAS of atom *i* is complementary to SAS of atom *i*); and r_i is the radius of atom *i*. The probe radius, r_{prob} , was set to 0.8 Å. We did not use the standard radius of the water probe (1.4 Å), with an aim to explore more hidden areas not accessible by a large water probe. Using r_{prob} of 0.8 Å, a set of encouraging SAS-based models have been developed by us for a variety of molecular properties which include solvation free energy and aqueous solubility.^{55,56} The radius parameters used for SAS calculations are adopted from the MSMS program⁵⁷ and listed in Table 1.

The term *k* in eq 4 is an adjustable parameter to balance the contribution of buried atoms to *S*. Hereafter, we call conformational entropies calculated by eqs 4 and 5 WSAS entropies. To compare the entropy contributions at the energy level, the product of temperature and entropy, *TS*, is reported throughout the article (*T* is set to 298.15 K).

To develop a coherent WSAS model for both small molecules and macromolecules, the following strategy is proposed: (1) all the atoms having the same atom type share the same weight; (2) the weigh parameters are derived by linear regression analysis solely using the ab initio entropies of small molecules for a given *k* parameter; (3) the *k* parameter is systematically searched from 0 (buried SAS has no contribution to *S*) to 1 (buried SAS contributes equally to *S* as SAS), and an ideal value of *k* is recognized when both the performance of linear regression analysis is satisfactory and the rms error between the WSAS and NMA *TS* of macromolecules is as small as possible.

2.2. Data Sets. The objective of this paper is to develop an entropy model for both small organic and large biological molecules. Four types of data sets have been prepared to do parametrization and testing. Data set I collects 2756 small molecules that come from different sources: molecules used for force field parametrization; those with experimental entropies; model

Table 1. List of the Atom Type Definitions, Radius Parameters for SAS Calculations, and the Weights of the WSAS Entropy Model When k in Equation 4 Is Set to 0.461

atom type	radius (Å)	weight (cal/mol K Å ²)	definition	atom type	radius (Å)	weight (cal/mol K Å ²)	definition
h1	1.2	0.1676	hydrogen on aliphatic carbon with one electron-withdrawal group	na	1.54	0.0585	sp ² nitrogen with three substituents
h2	1.2	0.1539	hydrogen on aliphatic carbon with two electron-withdrawal groups	nb	1.54	0.1271	sp ² nitrogen in aromatic systems, such as pyridine
h3	1.2	0.1656	hydrogen on aliphatic carbon with three electron-withdrawal groups	nc/nd	1.54	0.1426	inner sp ² nitrogen in conjugated ring systems
h4	1.2	0.1708	hydrogen on aromatic carbon with one electron-withdrawal group	ne/nf	1.54	0.1398	inner sp ² nitrogen in conjugated chain systems
h5	1.2	0.1551	hydrogen on aromatic carbon with two electron-withdrawal groups	nh	1.54	0.0317	amine nitrogen bonded to aromatic rings
ha	1.2	0.1525	hydrogen bonded to sp ¹ and sp ² carbon	no	1.54	0.1343	nitrogen in nitro groups
hc	1.2	0.1670	hydrogen bonded to sp ³ carbon	o	1.4	0.2022	sp ² oxygen in C=O and COO [−]
hn	1.2	0.1605	hydrogen bonded to nitrogen	oh	1.4	0.2105	sp ³ oxygen in hydroxyl groups
ho	1.2	0.0984	hydrogen bonded to oxygen	os	1.4	0.1804	sp ³ oxygen in ethers and esters
hs	1.2	0.1225	hydrogen bonded to sulfur	s	2.0	0.1592	sp ¹ or sp ² sulfur (P=S, C=S, etc.)
hp	1.2	0.1274	hydrogen bonded to phosphorus	s4	2.0	0.1590	hypervalent sulfur, three substituents
hw	same as ho		hydrogen of water	s6	2.0	0.0900	hypervalent sulfur, four substituents
hx	same as h3		hydrogen on aliphatic carbon next to a positively charged group	sh	2.0	0.1674	sp ³ sulfur in thiol groups
c	1.74	0.0881	sp ² carbon in C=O and C=S	ss	2.0	0.1717	sp ³ sulfur other than “sh”, two substituents
c1	1.74	0.0974	sp ¹ carbon	sx	same as s4		conjugated sulfur, three substituents
c2	1.74	0.0374	sp ² carbon, aliphatic	sy	same as s6		conjugated sulfur, four substituents
c3	1.74	−0.0419	sp ³ carbon, aliphatic	p2	2.0	0.1821	sp ² phosphorus (C=P, etc.)
ca	1.74	0.0352	sp ² carbon, aromatic	p3	2.0	0.1435	sp ² phosphorus, three substituents
cc/cd	1.74	0.0308	inner sp ² carbon in conjugated ring systems	p5	2.0	0.0846	hypervalent phosphorus, four substituents
ce/cf	1.74	0.0321	inner sp ² carbon in conjugated chain systems	pc/pd	same as p2		inner sp ² phosphorus in conjugated ring systems
cg/ch	1.74	0.1080	inner sp ¹ carbon in conjugated ring systems	pe/pf	same as p2		inner sp ² phosphorus in conjugated chain systems
cp/cq	1.74	0.0260	bridge sp ² carbon in biphenyl	p4	same as p5		hypervalent phosphorus, three substituents
cu	same as c2		sp ² carbon in three-membered rings	px	same as p5		conjugated phosphorus, three substituents
cv	same as c2		sp ² carbon in four-membered rings	py	same as p5		conjugated phosphorus, four substituents
cx	same as c3		sp ³ carbon in three-membered rings	f	1.6	0.1992	any fluorine
cy	same as c3		sp ³ carbon in four-membered rings	cl	1.79	0.2101	any chlorine
cz	same as c2		sp ² carbon in guanidine	br	2.04	0.2025	any bromine
n	1.54	0.0194	sp ² nitrogen in amides	i	2.15	0.2012	any iodine
n1	1.54	0.1824	sp ¹ nitrogen	si	2.1	0.0594	any silicon
n2	1.54	0.1647	sp ² nitrogen with two substituents, real double bond formed	fe	2.0	0.0000	iron
n3	1.54	0.0393	sp ³ nitrogen with three substituents	constant		34.7177	intercept of multiple linear regression
n4	1.54	−0.0421	sp ³ nitrogen with four substituents				

compounds of different organic substituent groups and ring systems listed in CRC Handbook of Chemistry and Physics,⁵⁸ etc. Data set I was used to derive the weight parameters in eq 4. Since it covers most of if not all the chemical functional groups, our entropy model is unlikely to suffer from the missing parameter problem for arbitrary molecules.

Data set II has eight protein–ligand systems (PDB codes: 1A9U, 1ABE, 1AHA, 1FKG, 1FKI, 1HPV, 3PTB, and 4PHV), three protein–peptide systems (1BE9, 1HEF, and 8GCH), three unbound proteins (1F10, 1KOE, and 1LMI), one DNA–ligand complex (19SD), and one RNA (422D). For both 1A9U and 1HPV, four truncated systems were generated for each protein by only keeping residues within certain distances (8, 10, 12, and 15 Å) from any atom of the inhibitor. All the structures in this data set were downloaded from the Protein Data Bank (www.rcsb.org).⁵⁹ MD simulations were performed for each protein or nucleic acid system using the “single trajectory protocol”,

and 10 snapshots were collected for the postentropy calculations. In total, there are 20 × 10 bound and 24 × 10 unbound protein or nucleic acid models prepared (Table 2).

Data set III collects 12 protein decoys generated by the Rosetta software package (www.rosettacommons.org). Every protein decoy contains 31 protein models including the crystal one. This data set was used by Lee et al. to evaluate if the state-of-the-art explicit solvent molecular dynamics and implicit solvent free energy calculation can identify the native states from conformational decoys.¹¹ It can be downloaded from <http://depts.washington.edu/bakerpg/decoys/>. In total, there are 372 protein models in this data set.

Data set IV is used to evaluate how well the WSAS model in combination with MM-PBSA and MM-GBSA performs in binding free energy calculations. It comprises six protein systems and each has several to tens of inhibitors (the total number of inhibitors is 53). This data set was originally applied by Hou et al. to

Table 2. Comparison between NMA and WSAS in TS ($T = 298.15$ K) Calculations for 22 Proteins and 2 Nucleic Acids

PDB_ID	data type	TS by NMA*	TS by WSAS*	linear regression **		
				R ²	AUE	RMSE
protein–ligand complexes						
1A9U	complex	3949.05 ± 10.88	3990.36 ± 4.29	0.83	6.41	7.24
	ligand	47.97 ± 0.00	46.80 ± 0.02	0.00	0.01	0.02
	protein	3922.79 ± 11.72	3963.93 ± 4.57	0.76	7.43	8.10
	binding	21.71 ± 4.00	20.37 ± 1.83	0.61	2.05	2.76
1A9U truncated R = 8 Å	complex	337.24 ± 2.07	281.04 ± 0.92	0.61	1.20	1.40
	ligand	47.97 ± 0.00	46.80 ± 0.01	0.00	0.01	0.01
	protein	304.44 ± 2.86	248.29 ± 1.39	0.54	1.94	1.98
	binding	15.16 ± 3.72	14.05 ± 1.40	0.55	2.52	2.79
1A9U truncated R = 10 Å	complex	497.22 ± 2.16	435.34 ± 1.69	0.80	0.81	0.97
	ligand	47.97 ± 0.00	46.80 ± 0.01	0.00	0.01	0.01
	protein	465.78 ± 3.03	406.17 ± 1.97	0.82	1.11	1.39
	binding	16.53 ± 3.07	17.63 ± 2.56	0.72	1.35	1.65
1A9U truncated R = 12 Å	complex	698.80 ± 1.96	639.32 ± 1.11	0.32	1.26	1.62
	ligand	47.97 ± 0.00	46.80 ± 0.01	0.00	0.01	0.01
	protein	672.33 ± 2.87	612.42 ± 1.71	0.38	1.80	2.25
	binding	21.49 ± 3.02	19.90 ± 2.12	0.76	1.14	1.52
1A9U truncated R = 15 Å	complex	1125.27 ± 6.65	1042.01 ± 2.47	0.70	3.72	4.65
	ligand	47.97 ± 0.00	46.80 ± 0.01	0.00	0.01	0.01
	protein	1098.35 ± 3.04	1016.40 ± 1.98	0.45	1.77	2.25
	binding	21.05 ± 6.89	21.19 ± 2.85	0.80	3.39	4.51
1ABE	complex	3249.00 ± 4.45	3278.36 ± 3.18	0.63	2.04	2.74
	ligand	29.24 ± 0.04	28.33 ± 0.02	0.78	0.02	0.03
	protein	3232.34 ± 5.41	3265.31 ± 2.56	0.53	2.94	3.98
	binding	12.59 ± 2.53	15.29 ± 1.70	0.50	1.34	1.84
1AHA	complex	2715.46 ± 5.04	2726.63 ± 3.12	0.42	3.32	3.82
	ligand	25.22 ± 0.00	23.51 ± 0.01	0.00	0.01	0.01
	protein	2704.94 ± 4.31	2715.87 ± 3.26	0.37	2.99	3.48
	binding	14.71 ± 3.44	12.75 ± 2.17	0.80	1.44	1.65
1FKG	complex	1232.21 ± 3.38	1233.20 ± 1.46	0.00	3.01	3.60
	ligand	61.60 ± 0.00	60.93 ± 0.01	0.00	0.01	0.01
	protein	1193.46 ± 3.66	1193.86 ± 1.46	0.05	2.56	3.62
	binding	22.85 ± 2.45	21.59 ± 1.37	0.40	1.48	1.90
1FKI	complex	2427.95 ± 4.40	2440.44 ± 1.87	0.07	3.63	4.28
	ligand	59.59 ± 0.16	62.97 ± 0.13	0.99	0.03	0.04
	protein	2384.62 ± 5.92	2395.56 ± 1.71	0.62	4.03	4.70
	binding	16.26 ± 3.59	18.09 ± 2.16	0.71	1.79	2.22
1HPV	complex	2218.89 ± 5.06	2218.38 ± 2.84	0.42	2.90	3.88
	ligand	67.71 ± 0.02	64.29 ± 0.03	0.40	0.04	0.05
	protein	2181.30 ± 6.46	2182.79 ± 2.71	0.37	3.93	5.27
	binding	30.11 ± 4.83	28.71 ± 1.03	0.67	3.38	4.00
1HPV truncated R = 8 Å	complex	406.48 ± 2.66	340.28 ± 0.98	0.53	1.59	1.97
	ligand	67.71 ± 0.02	64.29 ± 0.03	0.40	0.04	0.05
	protein	359.91 ± 2.91	294.81 ± 2.02	0.59	1.58	1.87
	binding	21.14 ± 3.46	18.82 ± 2.00	0.16	2.95	3.30
1HPV truncated R = 10 Å	complex	582.00 ± 2.88	523.89 ± 0.95	0.01	2.56	2.98
	ligand	67.71 ± 0.02	64.29 ± 0.03	0.40	0.04	0.05
	protein	540.15 ± 3.14	482.95 ± 2.62	0.69	1.18	1.78
	binding	25.86 ± 4.99	23.35 ± 2.43	0.62	2.66	3.30
1HPV truncated R = 12 Å	complex	770.23 ± 1.62	712.70 ± 0.84	0.34	1.09	1.32
	ligand	67.71 ± 0.02	64.29 ± 0.03	0.40	0.04	0.05
	protein	727.74 ± 3.44	670.24 ± 1.86	0.88	1.47	1.70
	binding	25.22 ± 3.73	21.83 ± 1.96	0.71	1.74	2.18
1HPV truncated R = 15 Å	complex	1135.37 ± 3.60	1081.09 ± 1.57	0.68	1.89	2.43
	ligand	67.71 ± 0.02	64.29 ± 0.03	0.40	0.04	0.05
	protein	1093.99 ± 6.84	1041.29 ± 2.58	0.73	3.97	4.75
	binding	26.33 ± 5.54	24.49 ± 2.36	0.59	3.53	3.94
3PTB	complex	2296.37 ± 6.80	2302.52 ± 3.27	0.74	3.60	4.32
	ligand	25.81 ± 0.00	24.73 ± 0.01	0.00	0.08	0.00
	protein	2286.18 ± 7.17	2294.37 ± 3.35	0.78	3.37	4.51
	binding	15.63 ± 3.35	16.58 ± 0.93	0.87	2.05	2.55

Table 2. continued

PDB_ID	data type	TS by NMA*	TS by WSAS*	linear regression**		
				R ²	AUE	RMSE
protein–ligand complexes						
4PHV	complex	2246.00 ± 5.68	2237.64 ± 2.13	0.77	3.27	3.95
	ligand	78.26 ± 0.38	73.52 ± 0.09	0.80	0.28	0.30
	protein	2195.60 ± 7.73	2192.79 ± 3.34	0.85	3.98	4.83
	binding	27.85 ± 8.64	28.67 ± 2.76	0.86	5.65	6.21
protein–peptide complexes						
1BE9	complex	1306.77 ± 4.98	1315.13 ± 2.34	0.25	3.72	4.31
	peptide	84.32 ± 0.29	79.40 ± 0.12	0.71	0.16	0.20
	protein	1255.68 ± 4.74	1259.23 ± 1.35	0.07	3.60	4.58
	binding	33.23 ± 6.03	23.51 ± 2.33	0.82	2.81	3.33
1HEF	complex	1179.50 ± 5.45	1178.84 ± 2.67	0.59	3.25	3.79
	peptide	90.76 ± 0.57	86.93 ± 0.18	0.07	0.46	0.55
	protein	1115.50 ± 4.54	1114.67 ± 2.58	0.89	1.97	2.27
	binding	26.76 ± 6.74	22.76 ± 2.89	0.83	3.05	3.87
8GCH	complex	2491.53 ± 4.47	2489.09 ± 3.05	0.33	2.99	3.68
	peptide	50.23 ± 0.90	46.70 ± 0.30	0.82	0.50	0.62
	protein	2467.12 ± 3.30	2463.14 ± 3.32	0.69	1.54	1.93
	binding	25.82 ± 5.62	20.75 ± 2.93	0.85	1.89	2.67
protein receptors						
1FLO	protein	1817.62 ± 6.65	1823.04 ± 2.52	0.47	4.19	5.25
1KOE	protein	1859.37 ± 7.42	1859.16 ± 2.78	0.85	4.16	4.97
1LMI	protein	1380.58 ± 3.97	1381.46 ± 2.38	0.89	1.63	1.89
nucleic acids						
19SD	complex	645.24 ± 2.28	658.07 ± 1.45	0.80	0.98	1.19
	ligand	65.72 ± 0.50	57.99 ± 0.08	0.63	0.38	0.43
	DNA	611.69 ± 3.06	625.47 ± 1.81	0.42	2.17	2.34
	binding	32.18 ± 3.64	25.39 ± 2.10	0.50	2.01	2.57
422D	RNA	668.23 ± 5.02	658.56 ± 2.55	0.88	2.32	2.75

* For each TS by NMA or WSAS, the mean and the rms deviation were estimated using 10 data points calculated using 10 MD snapshots. ** The performance of a linear regression is measured by correlation coefficient squared, the average unsigned error (AUE), and root-mean-square error (RMSE).

assess the performance of MM-PBSA and MM-GBSA methods in binding free energy calculations.¹⁹

2.3. Ab initio Calculations. Two types of ab initio calculations were performed in this work. For all the molecules in data set I, the conformational entropies were calculated using the Jaguar software package of Schrodinger LLC (www.schrodinger.com). Structural optimization was first performed at the B3LYP/6-31G* level taking the aqueous solvent effect into account using a Poisson–Boltzmann model. It has been shown by Baron et al. that solvent has effect on the solute conformational entropy.⁶⁰ The Hessian matrixes calculated in the first step were used to conduct frequency analysis. The B3LYP/6-31G* frequencies were scaled down using a scaling factor of 0.9945 to better reproduce the experimental values.⁶¹ The thermochemical properties at 295.15 K were then calculated using the scaled frequencies. The total entropy *S* including the contributions from the translational (*S*_{trans}), rotational (*S*_{rot}), and vibrational (*S*_{vib}) movements was taken as the conformational entropy of a molecule.

For the second type of ab initio calculations, the inhibitors of 19SD, 1A9U, 1ABE, 1AHA, 1FKG, 1FKI, 1HPV, 3PTB, and 4PHV were extracted from the complexes and optimized at the HF/6-31G* level using the Gaussian 03 software package.⁶² The Merz–Singh–Kollman scheme was used to generate electrostatic potential (ESP).⁶³ Then the RESP (restrained electrostatic potential) charges⁶⁴ were derived using the RESP program in AMBER 11⁶⁵ taking the ab initio ESP as input.

2.4. Molecular Dynamics Simulations. The Parm99SB biomolecular force field^{66,67} and the general AMBER force field (GAFF)⁶⁸ were used for all the molecular mechanics calculations. The topologies of nonstandard residues were prepared using the Antechamber module⁶⁹ in AMBER 11.⁶⁵ All MD simulations were performed with the periodic boundary condition to produce isothermal–isobaric ensembles at 300 K using the Sander program in AMBER11. The particle mesh Ewald (PME) method^{70–72} was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images. The integration of the equations of motion was conducted at a time step of 2 fs. The covalent bonds involving hydrogen atoms were frozen with the SHAKE algorithm.⁷³ Temperature was regulated using the Langevin dynamics⁷⁴ with the collision frequency of 5 ps^{−1}.^{75–77} Pressure regulation was achieved with isotropic position scaling and the pressure relaxation time was set to 1.0 ps. After the systems were well equilibrated, 1 ns MD simulations were conducted and 10 snapshots were evenly recorded for the following entropy calculations using NMA and WSAS.

2.5. Normal Model Analysis. The normal mode analysis was conducted using the NAB module of the AMBER11 software package.⁶⁵ Unless stated otherwise, the whole structures of biomolecules were used for normal mode analysis. For each snapshot, a maximum of 20 000 step-conjugate gradient minimization was first performed and the converge criterion of the gradient was set to 0.0001 kcal/(mol Å). Then a Newton–Raphson

minimization was performed until either the maximum of 200 steps was reached or the root-mean-square of gradient was less than 1.0×10^{-12} kcal/(mol Å). The generalized Born model of Hawkins, Cramer, and Truhlar⁷⁸ was utilized during the minimization and the followed normal mode analysis. The interior and exterior dielectric constants for the GB calculation were set to 1 and 78.5, respectively.

2.6. MM-PBSA/MM-GBSA Binding Free Energy Calculations. The computational details of MM-PBSA and MM-GBSA binding free calculations for Data set IV were presented in Hou's work.¹⁹ All the energy terms were adopted from the paper except for the conformational entropies, which were recalculated by our WSAS model as described below.

2.7. Model Construction. A two-step systematic search was applied to locate the ideal k parameter in eq 4. First of all, k was scanned from 0 to 1.0 at a step of 0.01. In the next fine-tuning step, a much smaller step, 0.001, was used to scan a focused range of k . For a given k , the weight parameters were derived by linear regression analysis solely using the ab initio entropies of small molecules. Then the TS of macromolecules in data sets II and III were calculated by the current WSAS entropy model. An ideal k was recognized if it led to a good regression performance and simultaneously minimized the difference between the WSAS and NMA TS of macromolecules in data sets II and III.

3. RESULTS AND DISCUSSION

In this section, we present the linear regression model of conformational entropy based upon solvent accessible surface area calculation, followed by assessing the performance of this model in three tests.

3.1. Development of a Conformational Entropy Model Based on Weighted Solvent Accessible Surface Area. In our algorithm, k in eq 4 is a tunable parameter. Without this parameter, it is nearly impossible to develop one single entropy model to make satisfactory predictions for both small molecules and macromolecules. An ideal k value minimizes the rms error between the WSAS and NMA TS of macromolecules in data sets II and III without sacrificing the performance of linear regression for the small molecules in data set I. A systemic search was applied to identify the ideal k value. As shown in Figure 1, for small molecules in data set I, the AUE and RMSE of TS predicted by the linear regression models do not change dramatically and the maximum difference of RMSE is only 0.01 kcal/mol for k ranged from 0.4 to 0.8 (the minimum of 0.698 kcal/mol is at $k = 0.55$). While for macromolecules, the RMSE of TS dramatically reduces from 625.8 to 30.3 kcal/mol when k increases from 0 to 0.461; and then, the RMSE significantly increases to 204.6 kcal/mol when k approaches 1.0. We finally set k to 0.461 as the ideal value. When k is set to the ideal value, the RMSE of TS for the small molecules in data set I is 0.70 kcal/mol, while the RMSE of TS for the macromolecules in data sets II and III are 30.3 kcal/mol.

As illustrated in Figure 1, the buried solvent accessible surface areas make a big contribution to the TS of a given molecule. When this contribution is totally ignored ($k = 0$), not only does the RMSE of TS between WSAS and NMA for macromolecules dramatically increase (625.8 kcal/mol), but the RMSE of TS between WSAS and ab initio for small molecules (0.84 kcal/mol) also increases.

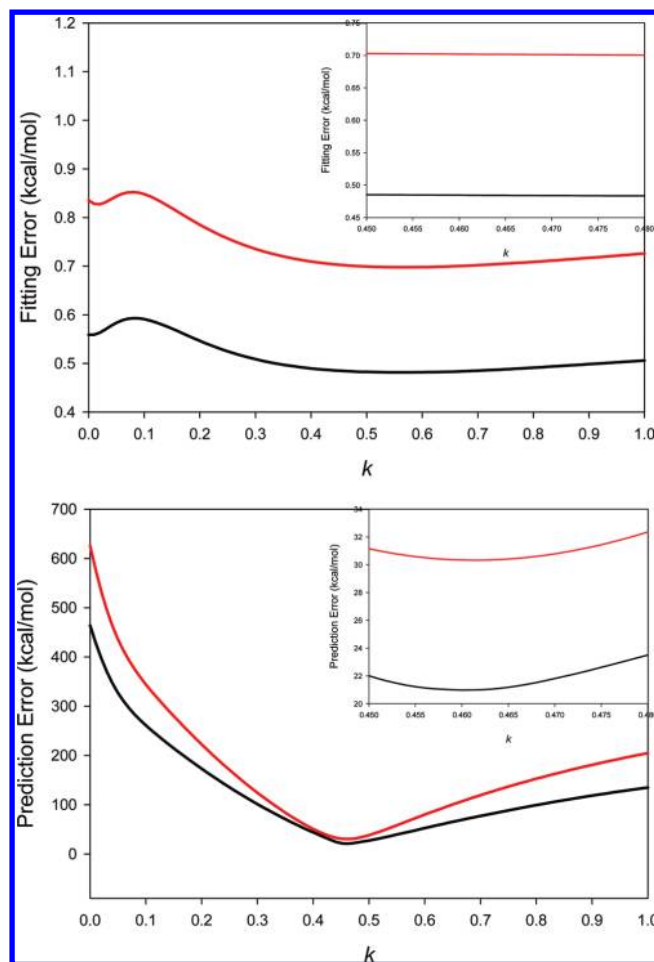


Figure 1. Application of a two-step systematic search to scan the k parameter in eq 4. An ideal k value is recognized when the rms error of conformational entropy by linear regression analysis for small molecules in data set I and the rms error between the NMA TS and WSAS TS of biomolecules in data sets II and III are simultaneously minimized. It is shown that the AUE (black) and RMSE (red) of TS by linear regression analysis using data set I are almost unchanged for k from 0.45 to 0.48 (top panel); similarly, the AUE (black) and RMSE (red) between the NMA TS and WSAS TS of biomolecules in data sets II and III approach their minimum for k from 0.45 to 0.48 (bottom panel). Thus, the ideal k parameter is set to 0.461.

The weight parameters for $k = 0.461$ are listed in Table 1. The atom type definitions are the same as those for GAFF.⁶⁸ The performance of the linear regression using data set I is demonstrated in Figure 2. The correlation coefficient square is 0.99, and the AUE and RMSE of TS are 0.48 and 0.70 kcal/mol ($T = 298.15$ K), respectively. The ab initio and WSAS entropies, the SAS and BSAS as well as the SMILES and SLN (Sybyl Line Notation) notations of the 2756 molecules are listed in Table S1 of the Supporting Information.

It is interesting to investigate the performance of the regression model with all the atom types sharing the same weight. We have found that this simplified SAS model is far inferior to the above WSAS model: the R^2 is 0.92, and the AUE and RMSE of TS are 1.24 and 1.73 kcal/mol, respectively. Therefore the approach to allowing different atom types to have different weights is well-justified.

The contribution of translational and rotational entropies dominates the conformational entropies of a small molecule as

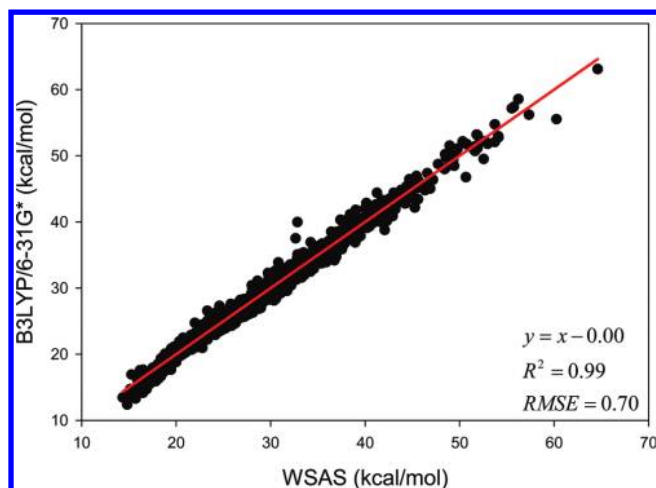


Figure 2. Performance of the WSAS entropy model in reproducing the B3LYP/6-31G* TS for the 2756 small molecules in data set I.

indicated by Supporting Information Table S1. However, it is not justified to neglect the contribution of vibrational entropy: the correlation coefficient square is only 0.79 between the translational/rotational entropies and the total entropies for data set I. As long as a macromolecule is concerned, the dominated component of conformational entropy is from vibrational frequencies and the vibrational entropy is certainly not negligible.

Is it a better idea to model vibrational entropy separately? To find out the answer, we constructed a WSAS model only using the vibrational entropies (Table S1). It is clear that the regression model is also inferior to the best WSAS model which calibrates the translational, rotational and vibrational entropies altogether: $R^2 = 0.97$ and the AUE and RMSE of TS are 0.67 and 0.96 kcal/mol, respectively. In another word, the AUE and RMSE increase about 40% compared to the best WSAS model ($k = 0.461$).

Compared to normal mode analysis, the CPU time of WSAS is negligible. For the three aforementioned protein systems, the average CPU time is 0.7, 1.8, and 5.1 s for 1BE9, 8GCH, and 1A9U, respectively.

3.2. Model Validation for Small Molecules. In order to test the predictability of the WSAS model described above, two validation analyses were conducted for data set I. First of all, the leave-one-out (LOO) analysis achieves a correlation coefficient square (q^2), AUE, and RMSE of 0.99, 0.50, and 0.79 kcal/mol, respectively. Then 10 000 cross validation (CV) runs were performed. For each run, 5% entries were randomly selected to enter the test set and their entropies were calculated with the model generated using the remaining entries. The mean q^2 , AUE, and RMSE for the 10 000 cross validation are 0.99 ± 0.01 , 0.50 ± 0.05 , and 0.76 ± 0.23 , respectively. The distributions of cross validation q^2 , AUE, and RMSE are shown in Figure 3. As the AUE and RMSE in both validation tests are only marginally larger than those of full regression analysis and the q^2 of LOO and CV are essentially equal to the correlation coefficient square of the full regression analysis (R^2), we are confident that our WSAS model is very reliable.

3.3. Model Validation for Macromolecules. It should be emphasized that the weight parameters in eq 4 were determined using the small molecule data set (data set I). The macromolecule data sets (data sets II and III) were only used to determine the ideal value of the k parameter. The performance of the WSAS model in reproducing the NMA TS of macromolecules

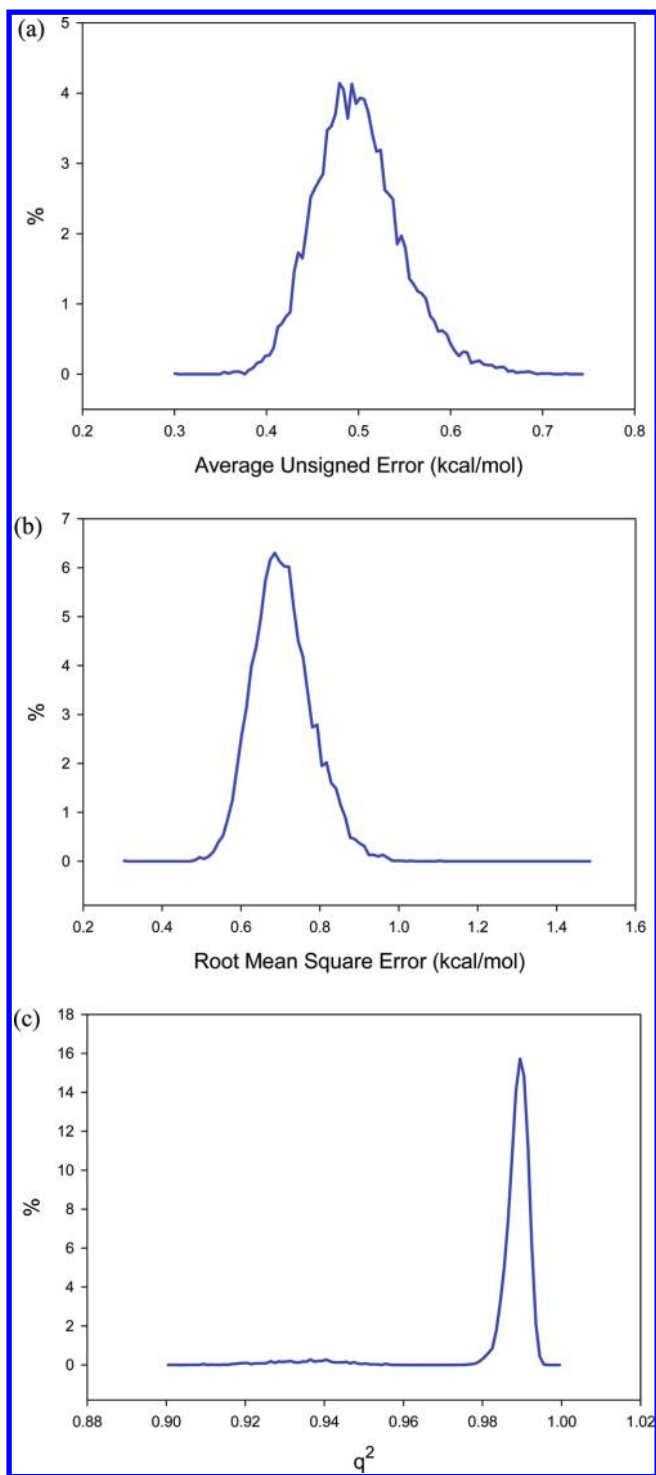


Figure 3. Distributions of the key statistical parameters of the 10 000 cross validation runs on the best WSAS model ($k = 0.461$): (a) average unsigned error of TS, (b) root-mean-square error of TS, (c) q^2 .

is illustrated in Figure 4. The correlation coefficient square is 1.00, and the AUE and RMSE are 21.0 and 30.3 kcal/mol, respectively.

For the normal mode analysis, attention must be paid to the first several frequencies. The first six frequencies are for the translational and rotational modes. Their values should be close to 0.0. If one or several negative vibrational frequencies are observed, it suggests that the structure is optimized to a transition state, or to a higher-order point. To guarantee that a

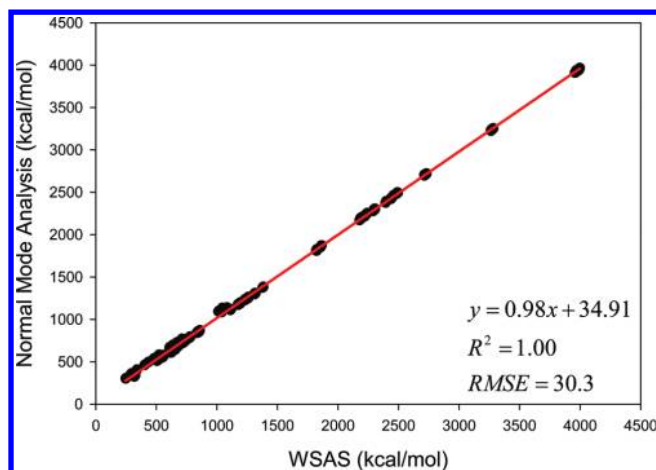


Figure 4. How the WSAS entropies reproduce the TS by normal mode analysis for 812 protein and nucleic acid models in data sets II and III.

local minimum or the global minimum is found, two types of algorithms, conjugate gradient followed by Newton–Raphson, were used to minimize the rms gradient to a very low value. In this work, the final rms gradients are all smaller than 1.0×10^{-12} kcal/(mol Å). In our experience, when rms gradients are larger than 1.0×10^{-4} kcal/(mol Å), the calculated thermodynamics properties may not be reliable.

3.4. Application of the WSAS Entropy in Protein Structure Prediction. In the following section, the WSAS model was challenged in more realistic applications. In the first test, we examine how WSAS performs in reproducing the NMA entropies of different conformations sampled by MD simulations for a single macromolecule. For each protein or nucleic acid of data set II, 10 snapshots were collected in 1 ns MD simulations. The statistics of the means and the root-mean-square deviations of TS are summarized in Table 2. Although the absolute difference of conformational entropies by WSAS and NMA can be large, reasonably good correlations are achieved in most cases. The mean correlation coefficient squares R^2 is 0.56 for 44 macromolecules (24 receptors and 20 complexes), while R^2 of 20 ligands is 0.34 between WSAS and NMA. However, R^2 for the ligands can be dramatically increased to 0.69 if those rigid ligands with the rms deviations of TS by normal mode analysis smaller than 0.03 kcal/mol are eliminated. The R^2 , AUE, and RMSE of the linear regressions analysis are listed in Table 2.

The changes of TS in ligand binding were calculated using the following equation: $T\Delta S = T(S_{\text{complex}} - S_{\text{receptor}} - S_{\text{ligand}})$. Good correlations were achieved between the WSAS $T\Delta S$ and NMA $T\Delta S$ as shown in Table 2. The mean R^2 , AUE, and RMSE of 20 protein and nucleic systems are 0.67, 2.41, and 2.94 kcal/mol, respectively. The significantly better correlation of $T\Delta S$ than that of the TS of macromolecules is understandable since the former has achieved a better error cancellation.

The root-mean-square deviations (rmsd) of the NMA and WSAS TS were calculated and listed in Table 2. For the macromolecules (receptors and complexes), the average, minimum, and maximum rmsd of NMA are 4.3, 0.3 and 11.7, respectively; while the three deviations are 2.1, 0.1, and 4.6 kcal/mol correspondingly for WSAS. As long as the rmsd for the binding entropy $T\Delta S$ are compared, the deviations for WSAS (mean = 2.1, min = 0.9, max = 2.9 kcal/mol) are also much smaller than those for NMA (mean = 4.5, min = 2.4, max = 8.6 kcal/mol). It is also encouraging that the RMSE of linear regression analysis are always smaller than the rms deviations of TS by NMA for biomolecules.

For large macromolecules, the widely used normal-mode analysis is very time-consuming. A common strategy in MM-PB/GBSA analysis is to truncate the residues that are relatively far away from the binding site in order to save computer time. Here we investigated how this strategy affects the calculation results using two protein systems, MAP kinase P38 (PDB code: 1A9U)⁷⁹ and HIV-1 protease (PDB code: 1HPV).⁸⁰ Each protein was truncated by only keeping residues within 8, 10, 12, and 15 Å of any atoms of the inhibitors. As demonstrated by Table 2, the TS by NMA can be dramatically different between untruncated and truncated proteins. For the first truncating scheme (truncating radius $R = 8$ Å), the $T\Delta S$ of ligand binding by NMA are underestimated by 6.6 and 8.9 kcal/mol for 1A9U and 1HPV, respectively. For the second truncating scheme (truncating radius $R = 10$ Å), the $T\Delta S$ are underestimated by 5.2 and 4.2 kcal/mol for 1A9U and 1HPV, respectively. Even when the truncating radius is as large as 15 Å, the binding entropy is still underestimated by more than 3.8 kcal/mol for 1HPV. Therefore, the strategy of truncating less important residues to enhance computational efficiency should be exercised with great caution.

Unlike data set II for which the entries were sampled by MD simulations, Data set III collects conformational decoys generated by the program of conducting ab initio protein predictions in the Rosetta software package (www.rosettacommons.org). An encouraging performance has been achieved by WSAS in comparison to NMA calculated by the NAB program. The mean R^2 for 12 small protein systems is 0.73, and the AUE and RMSE of TS are 2.4 and 3.0 kcal/mol, respectively. Figure 5 illustrates the performance of linear regressions between WSAS and NMA for each protein decoy set.

In summary, our WSAS entropy model has achieved a very good performance in predicting the conformational entropies for a variety of protein and nucleic acids systems. Given the fact that WSAS is computationally very efficient, it could be implemented in protein structural prediction or protein design packages to take the entropic effect into account.

3.5. Application of the WSAS Model in Rational Drug Design. Unlike the above sections of using entropies by normal mode analysis as references, in this evaluation test, we investigate how WSAS in combination with MM-PB/GBSA performs when compared to experimental binding free energies. In a recent work by Hou et al.,¹⁹ different MM-PBSA and MM-GBSA schemes, such as using different GB models and utilizing different intrinsic dielectric constants in solvation free energies calculations, were explored for six protein–ligand systems. Each protein–ligand system has several to tens of inhibitors and their binding free energies are known. Four binding free energy calculation schemes, namely, MM-PBSA-NMA, MM-PBSA-WSAS, MM-GBSA-NMA, and MM-GBSA-WSAS, were assessed with the experimental data. All the energetic terms except the conformational entropies by WSAS are adopted from the work of Hou et al.¹⁹ It should be pointed out that the WSAS entropies were calculated using the MD sampled conformations without further minimization.

The multiple linear regression analysis results for the two MM-GBSA and two MM-PBSA schemes are listed in Tables 3 and 4, respectively. The linear fitting performance of the two MM-GBSA schemes is shown in Figure 6 and that of the two MM-PBSA schemes is illustrated in Figure S1 of the Supporting Information. As shown in Figures 6 and the S1, the absolute binding free energies of the first two systems (α -thrombin and avidin) are poorly predicted; however, very good correlations are achieved between the MM-PB/GBSA and experimental binding free energies for these two systems. The seven

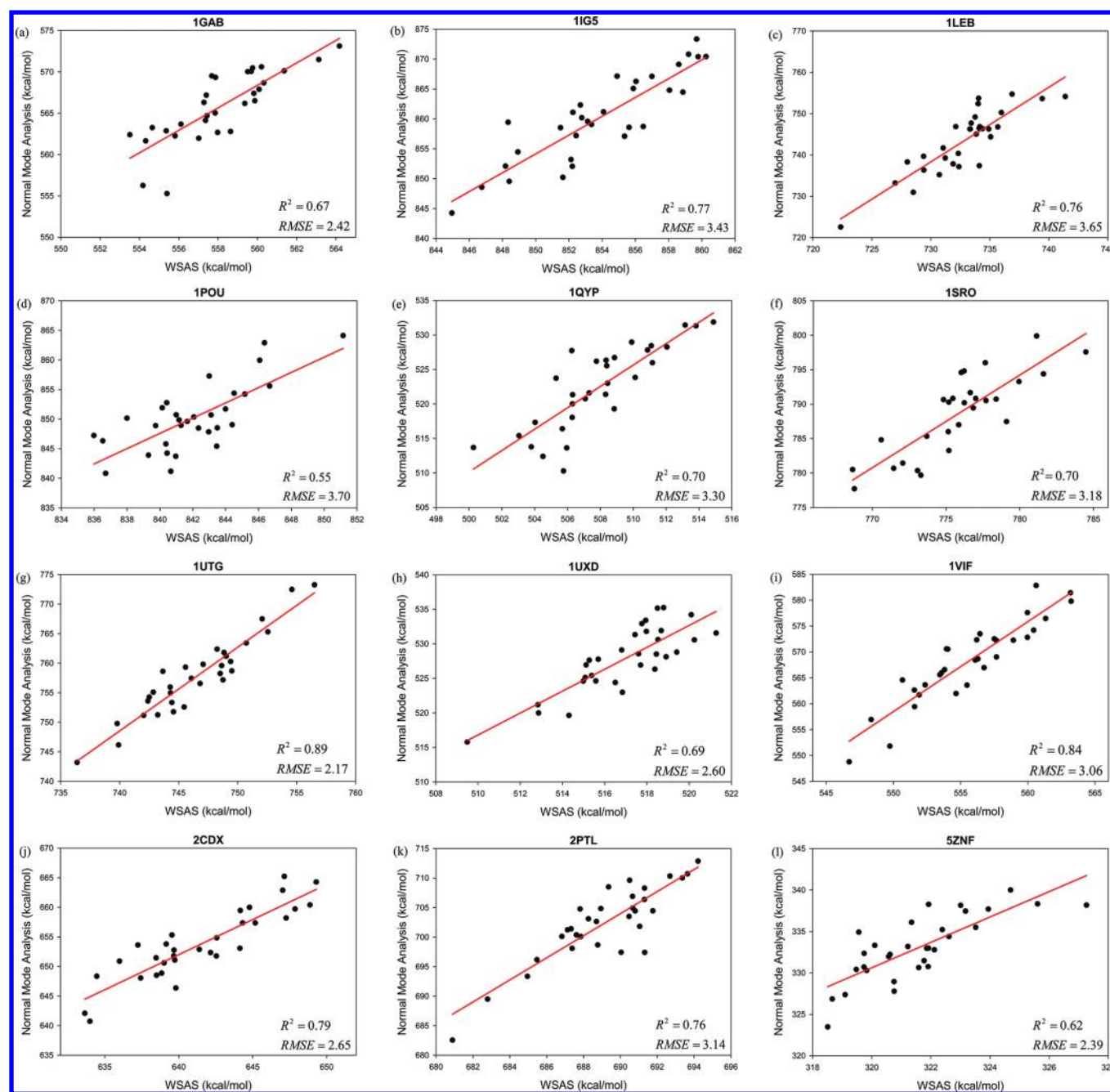


Figure 5. Performance of the WSAS entropy model in reproducing the TS by normal mode analysis for 12 protein decoys in data set III.

biotin/avidin complexes were also extensively studied by Genheden et al. using eight solvation models (two PBSA, four GBSA, and two MM/3D-RISM solvation models).⁸¹ Their finding is similar to ours: the variation of solvation free energies of the eight models is very extensive and the differences are up to 49.7 kcal/mol; therefore, the absolute binding free energies of this particular protein system are difficult to predict, while the relative binding free energies can be well-predicted (R^2 ranges from 0.59 to 0.93). In drug design, it is more meaningful to correctly rank the binding affinities of a set of ligands than to predict their absolute binding free energies. The poor correlation between the predicted and experimental binding free energies of cytochrome C is understandable: the range of the experimental binding free energies is very narrow and 80% of the data points fall between -6.2 and -4.2 kcal/mol.

The RMSE of prediction for this system are actually very good as shown in Tables 3 and 4.

The means and uncertainties of energy terms other than $T\Delta S$ by WSAS were listed in the work of Hou et al.¹⁹ The uncertainty of WSAS TS measured by the root-mean-square deviation of block averages are summarized as follows: the mean root-mean-square deviations are 0.05, 0.03, 0.02, 0.02, 0.02, and 0.06 kcal/mol for α -thrombin, avidin, cytochrome C peroxidase, neuraminidase, P450cam, and penicillopepsin, respectively.

It is encouraging that MM-GBSA-WSAS and MM-PBSA-WSAS achieve a comparable performance to that of two scoring functions employing NMA as demonstrated by Tables 3 and 4. Interestingly, the performance of the two PBSA schemes is inferior to that of the two GBSA schemes, which might suggest that the PBSA model Hou et al. used needs further parametrization.

Table 3. Performance of Two MM-GBSA Protocols in Binding Free Energy Calculations^a

protein systems	no. ligand	MM-GBSA-NMA			MM-GBSA-WSAS		
		AUE	RMSE	R ²	AUE	RMSE	R ²
α -thrombin	7	0.93	1.10	0.83	0.61	0.76	0.92
avidin	7	1.72	1.97	0.86	1.64	2.11	0.84
cytochrome C peroxidase	18	0.70	0.80	0.03	0.68	0.76	0.14
neuraminidase	8	1.95	2.10	0.35	1.78	2.16	0.31
P450cam	9	0.55	0.67	0.40	0.58	0.70	0.34
Penicillopepsin ^b	4	0.64	0.79	0.32	0.58	0.67	0.51
mean		1.08	1.24	0.47	0.98	1.19	0.51

^aAUE and RMSE are in kilocalories per mole. ^bThree ligands, f5, f6, and f7 in the paper of Hou et al.¹⁹ were eliminated because of apparently wrong molecular structures.

Table 4. Performance of Two MM-PBSA Protocols in Binding Free Energy Calculations^a

protein systems	no. ligand	MM-PBSA-NMA			MM-PBSA-WSAS		
		AUE	RMSE	R ²	AUE	RMSE	R ²
α -thrombin	7	1.40	1.57	0.64	1.21	1.42	0.70
avidin	7	1.62	1.92	0.87	2.43	2.58	0.77
cytochrome C peroxidase	18	0.66	0.77	0.09	0.65	0.76	0.11
neuraminidase	8	1.75	1.92	0.46	1.87	2.09	0.36
P450cam	9	0.54	0.64	0.46	0.66	0.75	0.25
Penicillopepsin ^b	4	0.89	0.94	0.03	0.66	0.86	0.18
mean		1.14	1.29	0.42	1.25	1.41	0.40

^aAUE and RMSE are in kilocalories per mole. ^bThree ligands, f5, f6, and f7 in the paper of Hou et al.¹⁹ were eliminated because of apparently wrong molecular structures.

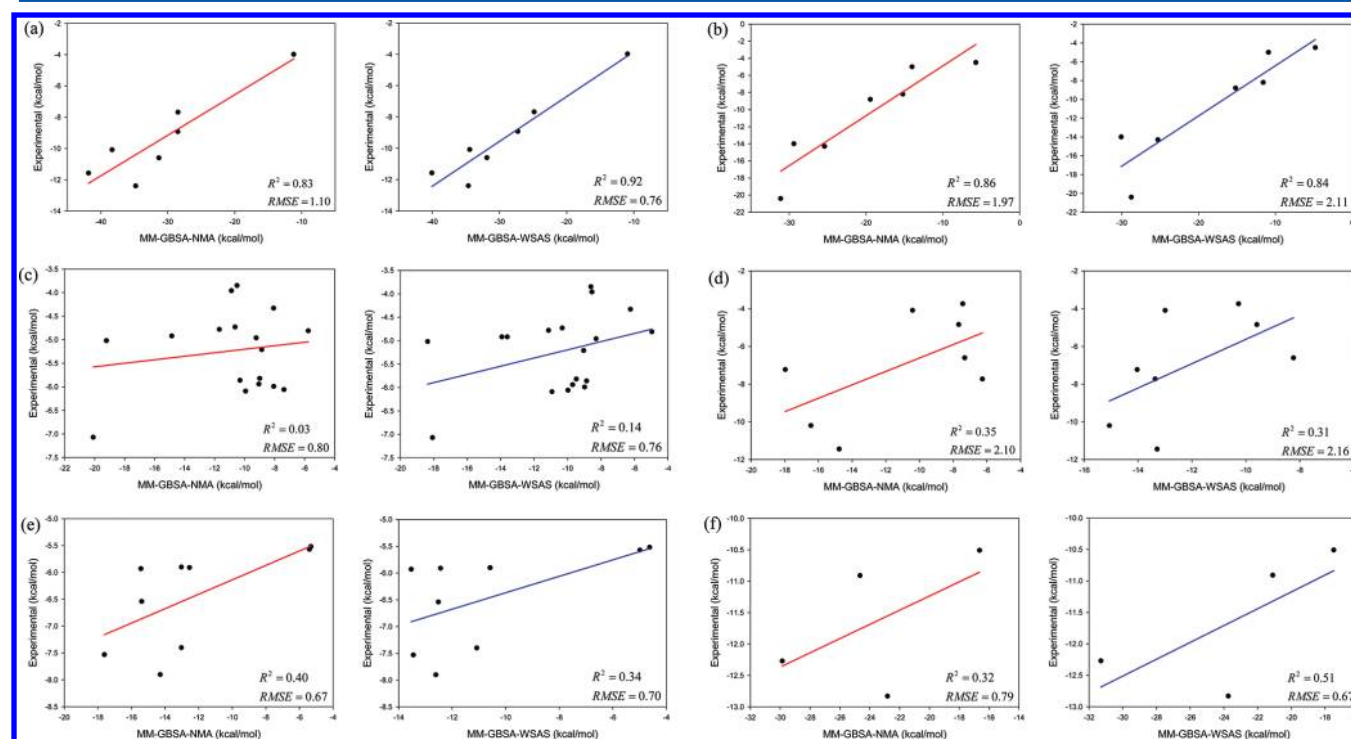


Figure 6. Comparison of two MM-GBSA protocols, MM-GBSA-NMA (left) and MM-GBSA-WSAS (right) in binding free energy calculations for six protein–ligand systems: (a) α -thrombin, (b) avidin, (c) cytochrome C peroxidase, (d) neuraminidase, (e) P450cam, and (f) penicillopepsin.

As this surprising result is unlikely related to conformational entropy, further discussion on this topic is beyond the scope of this paper.

3.6. Further Development. We have pointed out in the Methodology section that the weight parameters in eq 4 were derived solely using the B3LYP/6-31G* entropies of 2756 small molecules in data set I, while the k parameter was

determined using both data sets I and II macromolecular data. Because of this, our WSAS model is somewhat independent of molecular mechanical force fields. Thus, we anticipate that the k parameter does not deviate from 0.461 too much for a harmonic force field other than AMBER Parm99SB.

We must point out that the WSAS model inherits some limitations of normal mode analysis, for example the anharmonic

effect is totally ignored. However, the quality of the weight parameters may not be reduced as the anharmonicity is not significant for small molecules. The ideal k parameter may change when the conformational entropies of biological molecules in data sets II and III include the anharmonicity and high-order correlations. We plan to calculate the conformational entropies of biological molecules using the nonparametric kNN approach and to reoptimize the k parameter.

Although WASA has achieved an encouraging performance in several stringent tests, it is still improvable as the q^2 values of some cross validation runs are lower than the mean by up to 0.1 (Figure 3C). We plan to add more model compounds into data set I and oversample the populations of the building blocks of biomolecules, bioactive compounds as well as approved and experimental drugs. Second, atom type definition schemes will be explored by adding and/or modifying some atom types to improve the regression performance. We expect that the redeveloped WSAS model will be much more reliable and suitable to calculate the entropies of biomolecules as well as the entropic changes of protein–ligand and nucleic acid–ligand binding.

4. CONCLUSIONS

In this work, we proposed an approach of calculating conformational entropies using both the solvent accessible surface areas (SAS) and the buried SAS. The introduction of the tunable parameter k in eq 4, which balances the contribution of the buried SAS to the conformational entropy of a molecule, facilitates us to develop a general SAS-based entropy model for both small molecules and macromolecules. The ideal value of k , 0.461, was determined by a systematic search. This mode is very efficient since only SAS calculation is involved and geometric minimization is not needed prior to the entropy calculation.

The WSAS model has been extensively validated using both the small molecular and macromolecular data sets. The applications of WSAS in protein structural prediction and binding free energy calculations have been discussed. The overall performance of the WSAS model is very encouraging: the mean correlation coefficients between the WSAS and NMA are 0.56 and 0.73 for conformational decoys of biological molecules sampled by MD and Rosetta, respectively; in combination with MM-PBSA and MM-GBSA, the WSAS entropy model achieves a comparable performance to that of NMA in reproducing the experimental binding affinities of six protein–ligand systems. How to further improve the WSAS model has also been discussed. Given the fact that the WSAS is computationally very efficient, we expect the WSAS model to have great applications in both protein structural modeling and structure-based drug design.

■ ASSOCIATED CONTENT

■ Supporting Information

SMILES and SLN notions of 2756 small molecules in data set I are listed in Table S1. This table also lists the ab initio TS at the B3LYP/6-31G* level and the predicted TS by the WSAS model. The performance of the two MM-PBSA protocols (MM-PBSA-NMA and MM-PBSA-WSAS) is shown in Figure S1. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: junmei.wang@utsouthwestern.edu. Tel.: (214)-648-4146.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants R21GM097617 (J.W., P.I.) and R01GM79383 (Y. Duan, P.I.) and the National Science Foundation of China 21173156 (T. Hou, P.I.). We are grateful to TeraGrid for the computational time (TG-CHE090098, J.W., P.I.) and the research support from OpenEye Scientific Software, Santa Fe, NM.

■ ABBREVIATIONS

S = conformational entropy

TS = product of temperature T and conformation entropy S

SAS = solvent accessible surface area

BSAS = buried solvent accessible surface area

WSAS = weighted solvent accessible surface area

MM-PBSA = molecular mechanics Poisson–Boltzmann surface area

MM-GBSA = molecular mechanics generalized Born surface area

NMA = normal mode analysis

QHA = quasi-harmonic analysis

kNN = k -nearest neighbor

QSAR = quantitative structure–activity relationship

MLR = multiple linear regressions

LOO = leave-one-out

CV = cross validation

R^2 = square of regression coefficient

q^2 = square of cross-validation regression coefficient

AUE = average unsigned error

RMSE = root-mean-square error

■ REFERENCES

- (1) 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., III Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (2) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047–1069.
- (3) Wang, J. M.; Hou, T. J.; Xu, X. J. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 287–306.
- (4) Jorgensen, W. L. Free-Energy Calculations - A breakthrough for modeling organic-chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (5) Beveridge, D. L.; Dicapua, F. M. Free-energy via molecular simulation - applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431–492.
- (6) Meirovitch, H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181–186.
- (7) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
- (8) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* **2000**, *18*, 113–135.
- (9) Swanson, J. M. J.; 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*, 67–74.
- (10) 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.

(11) Lee, M. R.; Tsai, J.; Baker, D.; Kollman, P. A. Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Biol.* **2001**, *313*, 417–430.

(12) Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412–1419.

(13) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J. M.; Luo, R.; Bayly, C. T.; 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.

(14) Wang, J. M.; 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.

(15) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.

(16) Wang, J. M.; Kang, X. S.; Kuntz, I. D.; Kollman, P. A. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J. Med. Chem.* **2005**, *48*, 2432–2444.

(17) 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*, 4040–4048.

(18) Yang, T. Y.; Wu, J. C.; Yan, C. L.; Wang, Y. F.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Y. Virtual screening using molecular simulations. *Proteins* **2011**, *79*, 1940–1951.

(19) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model* **2011**, *51*, 69–82.

(20) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, *32*, 866–877.

(21) 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*, 6596–6606.

(22) Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009**, *17*, 489–498.

(23) McQuarrie, D. A.; Simon, J. D. *Molecular Thermodynamics*; University Science Books: Sausalito, CA, 1999.

(24) Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.

(25) Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929.

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

(27) Genheden, S.; Mikulskis, P.; Hu, L. H.; Kongsted, J.; Soderhjelm, P.; Ryde, U. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* **2011**, *133*, 13081–13092.

(28) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J. Comput. Chem.* **2008**, *29*, 1605–1614.

(29) Jusuf, S.; Loll, P. J.; Axelsen, P. H. Configurational entropy and cooperativity to estimate ligand binding and dimerization in glycopeptide antibiotics. *J. Am. Chem. Soc.* **2003**, *125*, 3988–3994.

(30) Harpole, K. W.; Sharp, K. A. Calculation of configurational entropy with a Boltzmann-Quasi-harmonic model: The origin of high-affinity protein-ligand binding. *J. Phys. Chem. B* **2011**, *115*, 9461–9472.

(31) Chang, C. E. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534–1539.

(32) Liu, L.; Yang, C.; Guo, Q. X. A study on the enthalpy-entropy compensation in protein unfolding. *Biophys. Chem.* **2000**, *84*, 239–251.

(33) Livesay, D. R.; Dallakyan, S.; Wood, G. G.; Jacobs, D. J. A flexible approach for understanding protein stability. *Febs Lett.* **2004**, *576*, 468–476.

(34) Li, D. W.; Khanlarzadeh, M.; Wang, J. B.; Huo, S. H.; Bruschweiler, R. Evaluation of configurational entropy methods from peptide folding-unfolding simulation. *J. Phys. Chem. B* **2007**, *111*, 13807–13813.

(35) 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*, 243–256.

(36) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.

(37) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(38) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(39) Giordanetto, F.; Cotesta, S.; Catana, C.; Trosset, J. Y.; Vulpetti, A.; Stouten, P. F.; Kroemer, R. T. Novel scoring functions comprising QXP, SASA, and protein side-chain entropy terms. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 882–893.

(40) Chang, C. E.; Gilson, M. K. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.

(41) Chen, W.; Chang, C. E.; Gilson, M. K. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J.* **2004**, *87*, 3035–3049.

(42) Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.

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

(44) Chang, C. E.; Gilson, M. K. Tork: Conformational analysis method for molecules and complexes. *J. Comput. Chem.* **2003**, *24*, 1987–1998.

(45) Chen, W.; Chang, C. E.; Gilson, M. K. Concepts in receptor optimization: targeting the RGD peptide. *J. Am. Chem. Soc.* **2006**, *128*, 4675–4684.

(46) Chang, C. E.; Potter, M. J.; Gilson, M. K. Calculation of molecular configuration integrals. *J. Phys. Chem. B* **2003**, *107*, 1048–1055.

(47) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. *J. Chem. Phys.* **2004**, *120*, 2652–2661.

(48) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* **2007**, *127*, 024107.

(49) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J. Comput. Chem.* **2007**, *28*, 655–668.

(50) Cheluvajaya, S.; Meirovitch, H. Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the

hypothetical scanning molecular dynamics method. *J. Chem. Phys.* **2006**, *125*, 24905.

(51) Meirovitch, H. Methods for calculating the absolute entropy and free energy of biological systems based on ideas from polymer physics. *J. Mol. Recognit.* **2010**, *23*, 153–172.

(52) Hensen, U.; Grubmüller, H.; Lange, O. F. Adaptive anisotropic kernels for nonparametric estimation of absolute configurational entropies in high-dimensional configuration spaces. *Phys. Rev. E* **2009**, *80*, 011913.

(53) Hensen, U.; Lange, O. F.; Grubmüller, H. Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach. *PLoS one* **2010**, *5*, e9179.

(54) Włodek, S.; Skillman, A. G.; Nicholls, A. Ligand Entropy in Gas-Phase, Upon Solvation and Protein Complexation. Fast Estimation with Quasi-Newton Hessian. *J. Chem. Theory Comput.* **2010**, *6*, 2140–2152.

(55) Wang, J.; Hou, T.; Xu, X. Aqueous solubility prediction based on weighted atom type counts and solvent accessible surface areas. *J. Chem. Inf. Model.* **2009**, *49*, 571–581.

(56) Wang, J. M.; Krudy, G.; Hou, T. J.; Zhang, W.; Holland, G.; Xu, X. J. Development of reliable aqueous solubility models and their application in druglike analysis. *J. Chem. Inf. Model.* **2007**, *47*, 1395–1404.

(57) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305–320.

(58) Lide, D. R. E., *CRC Handbook of Chemistry and Physics*, 86 ed.; CRC Press: Boca Raton, FL, 2005; pp 4–523.

(59) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(60) Baron, R.; Bakowies, D.; van Gunsteren, W. F. Principles of carboxypeptid folding: a molecular dynamics simulation study. *J. Pept. Sci.* **2005**, *11*, 74–84.

(61) Scott, A. P.; Radom, L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

(62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Gaussian, Inc, Wallingford, CT, 2004.

(63) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.

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

(65) Case, D. A.; Darden, T. A.; Cheatham, L. T. E.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER11*; University of California, San Francisco, CA, 2010.

(66) 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*, 1049–1074.

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

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

(69) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Modell.* **2006**, *25*, 247–260.

(70) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7*, R55–60.

(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(72) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004**, *120*, 73–87.

(73) Miyamoto, S.; Kollman, P. A. Settle - an analytical version of the Shake and Rattle algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(74) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: observation, explanation, and exploitation. *J. Chem. Phys.* **2004**, *120*, 6363–6374.

(75) Izaguirre, J. A.; Catarella, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090–2098.

(76) Larini, L.; Mannella, R.; Leporini, D. Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* **2007**, *126*, 104101.

(77) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides - the frictional dependence of isomerization rates of N-acetylalanine-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.

(78) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized 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.

(79) Wang, Z. L.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Struct. Fold. Des.* **1998**, *6*, 1117–1128.

(80) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal-structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.

(81) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.