A NOVEL ALGORITHM OF SOLVATION FREE ENERGY CALCULATION: THE KECSA-MOVABLE TYPE IMPLICIT SOLVATION MODEL (KMTISM)

By

Ting Wang

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry - Master of Science

2015

ABSTRACT

A NOVEL ALGORITHM OF SOLVATION FREE ENERGY CALCULATION: THE KECSA-MOVABLE TYPE IMPLICIT SOLVATION MODEL (KMTISM)

By

Ting Wang

A number of theoretical methods have been developed for calculating solvation free energies for biological and chemical processes. In this paper an implicit solvation model, KECSA-Movable Type Implicit Solvation Model (KMTISM) is created by utilizing an energy sampling approach termed the "Movable Type" (MT) method, and a statistical energy function for solvation modeling, "Knowledge-based and Empirical Combined Scoring Algorithm" (KECSA). The solvation free energies can be obtained from the NVT ensemble partition function generated by the MT method within the implicit solvent model approximation. Several subsets from the Minnesota Solvation Database v2012 are selected to use as validation sets. The solvation free energies getting from KMTISM are compared with several solvation free energy calculation methods, including MM-GBSA and MM-PBSA. Comparison against a quantum mechanics-based polarizable continuum model is also discussed (Cramer and Truhlar's Solvation Model 12).

ACKNOWLEDGMENTS

First of all I want to thank my adviser, Professor Kenneth M. Merz, Jr., for his unqualified support and guiding in my research during the three years in United States. I also want to thank his understanding and support for some of my decisions I made about my own life.

I also want to thank all of my committee members, Professor Katharine Hunt, Dr. Benjamin Levine, and Dr. Heedeok Hong, for their help and useful discussion in my project.

I want to give my sincere thanks to all of my group members, especially Dr. Zheng Zheng, Dr. Yipu Miao, Dr. Li-Li Pan, Dr. Nihan Ucisik, and Mr. Pengfei Li for all of the helpful discussion and support in my research and life. I also want to thank Dr. Mona Minkara, Zhuoqin Yu, Nupur Bansal, and Lin Song; they also give me a lot of support and advise. Sincere thanks are given to my friends in University of Florida, especially Shuai Wang, and Linna Hu, and also to all of my friends in Michigan State University.

I want to thank the Department of Chemistry of Michigan State University as well as the University of Florida, for their financial support when I am pursing my master's degree.

At last, I want to thank my family, from where I got selfless love, heartfelt care, and

emotionally support. I appreciate the understanding and supports for all of my decisions in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AMBER	Assisted Model Building with Energy Refinement
COSMO	Conductor-like Screening Model
CSD	Cambridge Structural Database
GB	Generalized Born Model
KECSA	Knowledge-based and Empirical Combined Scoring Algorithm
KMTISM	KECSA-Movable Type Implicit Solvation Model
MAE	Mean Absolute Errors
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MT	Movable Type
PB	Poisson-Boltzmann Model
PCM	Polarizable continuum Model
PDB	Protein Data Bank
PMF	Potential Mean Force
QM	Quantum Mechanics
RMSE	Root-Mean-Square Errors
SA	Solvent Accessible Surface Area
SMX	Cramer and Truhlar's Solvation Model

CHAPTER 1. INTRODUCTION

In the world of chemistry, chemical processes, which occur in aqueous environment including biological processes, always have interested researchers for many years. A rigorous simulation of those liquid phase reactions requires accurate methods to calculate solvation free energies, which are based on a good understanding of the solvation process of a solute, and an appropriate treatment of solvent. The existing methods for calculating solvation energies treat solvent molecules either explicitly[1-57] or implicitly[58-148]. When solvent molecules around solute molecules are treated explicitly, sampling methods such as molecular dynamics (MD) and Monte Carlo (MC) are widely used to calculate the solvation free energies. On the other hand, the solvation methods using implicit solvent models usually treat solvent as a continuum dielectric. To obtain electrostatic energies, either classical or quantum mechanics-based methods are used. Among the force field-based methods, which are based on fundamental molecular mechanics principles, the Poisson-Boltzmann and Generalized Born models combined with the hydrophobic solvent accessible area (MM-PBSA & MM-GBSA) are two of the most frequently used molecular mechanics implicit solvent models.[127-148] Reasonably accurate solvation free energies can be obtained by those models for both small and large molecules. For the calculation of solvation free energies of small molecules and ions using an implicit solvent model, quantum mechanics-based methods, such as the polarizable continuum models (PCM) coupled with different QM methods, [31,32,35,37,39,46] COSMO,[77] and Cramer and Truhlar's Solvation Models (SM) series[78-126] have yielded impressive performances. An accurate calculation of solvation free energy is crucial in the prediction of binding affinity of ligands to proteins, which is the key step in

application of drug design. In this thesis, a novel solvation free energy calculation approach, which employs statistical potential models together with basic statistical mechanical methodologies, will be described in detail.

In force field-based methods, pairwise potential energies are usually categorized into Lennard-Jones and electrostatic potentials. Unlike force field energy functions, statistical potential-based methods directly transform pairwise probabilities into a pairwise potential by using a hypothetical reference state.[149-186] The number density distributions of certain pairwise atoms display the impacts of chemical environments on the pairwise potential at the same time. Another critical aspect that affects the performance of statistical potentials is the availability of structure data for both proteins and small molecules. The success of statistical potentials in many applications, for example, protein folding[170-186] and protein-ligand binding,[149-169] substantially depends on protein and small molecule structural database. In this thesis, the Protein Databank (PDB) and the Cambridge Structural Database (CSD), which provide accurate position information of crystal waters, are used to construct atom pairwise information for solvation free energy calculations.

The core idea of statistical potentials is derived from the concept of potential of mean force (PMF). As shown in Equation 1, $\omega_{ij}^{(2)}(r_{12})$ is the mean potential between certain atom pairs, and $g^{(2)}$ is the correlation function. $\beta = \frac{1}{k_B T}$, where k_B is the Boltzmann constant and T is the temperature. $\rho_{ij}(r_{12})$ is the number density for atom type i and atom type j observed in molecules structures from selected

database, and $\rho_{ij}^{*}(r_{12})$ is the number density of the same atom pairs in the reference state.

$$\boldsymbol{\omega_{ij}^{(2)}(r_{12}) = -\frac{1}{\beta} ln\left(g^{(2)}(r_{12})\right) = -\frac{1}{\beta} ln\,(\frac{\rho_{ij}(r_{12})}{\rho_{ij}^{*}(r_{12})})} \quad (1)$$

Atom pairwise radial distribution functions are used to remove the background influence in application of statistical potentials. Atom pairwise radial distributions reflect all interactions in chemical space, and the conversion from radial distributions to energy functions is a challenge.[168,169]

Both accurate energy functions and extensive phase space sampling are required for the computation of solvation free energies. Equation 2 shows the Helmholtz solvation free energy of transferring a molecule (L) from vacuum to aqueous phase (S), which is obtained from the ratio of partition functions.

$$\Delta G_{solv}^{L} \approx \Delta A_{solv}^{L} = -RTln\left(\frac{Z_{LS}}{Z_{L}}\right) = -RTln(\frac{\int e^{-\beta E_{LS}(r)} dr}{\int e^{-\beta E_{L}(r)} dr}) \quad (2)$$

Movable Type (MT) method is a novel and efficient sampling method that was developed by Dr. Kenneth Merz Group. The MT method is able to estimate solvation free energies, binding free energies and even protein-ligand poses,[187] by sampling all atom pairwise energies at all achievable distances from molecular structure database.

In the following content, a detailed description of new statistical potential method for solvation free energy calculation will be discussed. The method combines the MT sampling method and a new reference state model, is going to be discussed. We selected 393 small molecules from Cramer and Truhlar's MNSol database[193] to validate our model, and the results from our model were compared with those results from MM-GBSA and MM-PBSA models, which are available in AMBER.[146-148]

CHAPTER 2. METHOD

2.1 Movable Type Continuum Solvation Model

Originated from the idea of ancient printing technique where a list of characters is created and then assembled using a movable type system, the "Movable Type" free energy calculation method[187] firstly needs a database that contains interaction energy information between all kinds of atom pairs found in the chemical space of interest, as "characters" in printing. The modified " Knowledge-based and Empirical Combined Scoring Algorithm" (KECSA) is used to obtain the atom pairwise energies. After the "characters" database is built, a Z-matrix is constructed to represent the Boltzmann-weighted energy ensemble, where atom pairwise energies of different distances are gathered to represent free energies of the chemical space of interest. As Equation 3 shows, Z_k^L matrix is a Boltzmann-weighted energy matrix for the kth atom pair in molecule L including energies ranging from r_1 to r_n of distance. The Z-matrix is composed of the Boltzmann-weighted energies of the observed atom pair at different distances, from the first atom pair in the molecular system. All of the energy terms are chosen from the energy database.

$$Z_k^L = \begin{bmatrix} e^{-\beta E_k^L(r_1)} & e^{-\beta E_k^L(r_{i+1})} & \cdots & e^{-\beta E_k^L(r_{n-i+1})} \\ e^{-\beta E_k^L(r_2)} & e^{-\beta E_k^L(r_{i+2})} & \cdots & e^{-\beta E_k^L(r_{n-i+2})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_k^L(r_i)} & e^{-\beta E_k^L(r_j)} & \cdots & e^{-\beta E_k^L(r_n)} \end{bmatrix} \quad (3)$$

As shown in Equation 4, the Boltzmann-weighted energy combinations between different atom pairs at different distances with a matrix size of n forms a pointwise product of Z-matrices, where Z_{12}^L represents the atom pairwise energy of atom pair

1 and atom pair 2 within molecule system L, and the sampled distance ranges are from r_1 to r_n and r_1' to r_n' respectively.

$$\mathrm{Z}_{12}^{L} = \mathrm{Z}_{1}^{L} \times \mathrm{Z}_{2}^{L} =$$

$$\begin{bmatrix} e^{-\beta(E_1^L(r_1)+E_2^L(r_1'))} & e^{-\beta(E_1^L(r_{i+1})+E_2^L(r_{i+1}'))} & \cdots & e^{-\beta(E_1^L(r_{n-i+1})+E_2^L(r_{n-i+1}'))} \\ e^{-\beta(E_1^L(r_2)+E_2^L(r_2'))} & e^{-\beta(E_1^L(r_{i+2})+E_2^L(r_{i+2}'))} & \cdots & e^{-\beta(E_1^L(r_{n-i+2})+E_2^L(r_{n-i+2}'))} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta(E_1^L(r_i)+E_2^L(r_i'))} & e^{-\beta(E_1^L(r_j)+E_2^L(r_j'))} & \cdots & e^{-\beta(E_1^L(r_n)+E_2^L(r_n'))} \end{bmatrix} \quad (4)$$

If random disordered permutations are applied to the Z-matrices, the diversity of energy combinations at different distances can be maximized. For the entire molecule system L, the final Z-matrix is a pointwise product of disordered matrices of all atom pairs within the molecule system, which represents the collection of Boltzmann-weighted energies with a size of n configurations. The final Z-matrix is created based on the assumption that the molecular energy is composed of all of the atom pairwise energies within the same molecular system.

The final Z-matrix of the molecular system is shown in Equation 5, where Z_{total}^{L} represents the pointwise product of disordered matrices of atom pairs from 1 to k of molecule L.

$$Z_{total}^{L} = disordered(Z_1^L) \times disordered(Z_2^L) \times \cdots \times disordered(Z_k^L)$$

$$= \begin{bmatrix} e^{-\beta(E_1^L(r_5)+E_2^L(r_3)+\cdots E_k^L(r_l))} & e^{-\beta(E_1^L(r_i)+E_2^L(r_{n-1})+\cdots E_k^L(r_{i+2}))} & \cdots & e^{-\beta(E_1^L(r_3)+E_2^L(r_{i+1})+\cdots E_k^L(r_n))} \\ e^{-\beta(E_1^L(r_{l+1})+E_2^L(r_i)+\cdots E_k^L(r_1))} & e^{-\beta(E_1^L(r_{i+1})+E_2^L(r_2)+\cdots E_k^L(r_{l+1}))} & \cdots & e^{-\beta(E_1^L(r_{i-1})+E_2^L(r_{l-2})+\cdots E_k^L(r_3))} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta(E_1^L(r_n)+E_2^L(r_l)+\cdots E_k^L(r_i))} & e^{-\beta(E_1^L(r_l)+E_2^L(r_{i-1})+\cdots E_k^L(r_2))} & \cdots & e^{-\beta(E_1^L(r_{i-2})+E_2^L(r_{n-2})+\cdots E_k^L(r_l))} \end{bmatrix}$$

(5)

The atom pairwise information collected within the molecular system is from certain bond lengths, angels and torsions. This will make the final Z-matrix contains some unreasonable distance combinations between different atom pairs. That is why a Q-matrix of atom pairwise radial distribution probabilities is needed to prevent the appearance of physically unreasonable combinations between different atom pairs. In order to obtain a reliable Q-matrix, a large structural database, which contains 8256 protein crystal structures from PDBBind v2013[190-192] database and 44766 small molecules from both PDBBind v2013 and CSD small molecule database, was used to get the elements of the Q-matrix. For each element in an atom pairwise Z-matrix, there is a corresponding distance – dependent probability value selected from the radial distribution probability database of the same atom pair type. The Q-matrices are constructed in a similar manner to that of the corresponding Z-matrix, namely, the Q-matrices are also formed by using pointwise product. To make sure the overall probability is 1, the final Q-matrix for the same molecular system is going to be normalized first. After that, the final matrix, C_{total}^{L}, is generated by multiplying the final Q-matrix by the final Z-matrix (Z_{total}^{L}). As shown in Equation 6, the sum of

the final matrix (C^L_{total}) provides the Boltzmann factors average of a matrix size (sampling size) of n.

$$\langle e^{-\beta E_L} \rangle = sum(C^L_{total}) = sum(\overline{Q^L_{total}}\, Z^L_{total}) \quad (6)$$

Hence, the energies of different molecular conformations can be created simultaneously by matrix products over all atom pairs by using the Z and Q matrices with a sampling size of n. By combining the ensemble average of Boltzmann factors, the solvation free energy can be calculated as Equation 7 shown.

$$\Delta G^L_{solv} \approx -RT \ln\left[\frac{Z_{LS}}{Z_L}\right] = -RT \ln\left[\frac{\int e^{-\beta E_{LS}(r)}\, dr}{\int e^{-\beta E_L(r)}\, dr}\right] = -RT \ln\left[\frac{DOF_{LS}\langle e^{-\beta E_{LS}(r)}\rangle}{DOF_L\langle e^{-\beta E_L(r)}\rangle}\right] \quad (7)$$

In Equation 7, the energy of the molecule in solution (E_{LS}) is expressed as Equation 8, and DOF_{LS} and DOF_L represents the degrees of freedom of the molecule in solution and in the gas phase respectively. For simplicity, the values of DOF_{LS} and DOF_L are set to be equal in the current implicit solvation model.

$$E_{LS}(r) = E_L(r) + E_{L-S\ interaction}(r) \quad (8)$$

In order to avoid the issues that are related to the additivity of the free energy, the solvation free energy is computed directly from the NVT ensemble. It is revealed both theoretically[194,195] and experimentally,[196] that energy can be broken down, but the entropy and free energies cannot. Hence, we obtain the interaction energies by

using Equation 8, and substitute it into Equation 7 to get the solvation free energy directly, so that the issues that are related to the decomposition of free energies can be avoided. Avoiding the free energy issues described above is also a big advantage of our MT method.

The MT energy sampling method can be used to calculate the solvation free energy involving both an explicit and implicit solvent model. Previously, we apply the MT method to a simple continuum ligand-solvent interaction model.[186] In this document, a new semi-continuum water model is developed, where the solute-solvent interaction is calculated by placing water molecules around the solute. In our simulation, the water molecules were modeled as isotropic rigid balls, whose Van der Waals radii are set as 1.6 Å.[197,198] The solvent layers were set to start at 8 Å away from the solute's van der Waals surface; water molecules were evenly put into those solvent layers with an increment of 0.005 Å per layer. The number of water molecules at each layer was related to their maximum cross-sectional areas, and the solvent accessible surface area at each solvent layer for each atom in the solute molecules. Figure 1 shows the modeling of the implicit solute-solvent model using the movable type method. The number of water molecules (N_w) accessible to each atom at distance R away from the atomic center of mass is rounded down via filtering using the maximum cross-sectional area (S_w) of water with the atomic solvent accessible surface area (S_a) in the solvent layer at distance R.

$$N_w(R) = floor\left(\frac{S_a(R)}{S_w}\right) \quad (9)$$

According to Figure 1, the maximum cross-sectional areas (S_w) of a water molecule is calculated as:

$$S_w = \int_{\pi/2-\theta}^{\pi/2} 2\pi(R_a + R_w)R_w \sin\left(\frac{\pi-\theta}{2}\right) d\left(\frac{\pi}{2}-\theta\right)$$

$$= 2\pi(R_a + R_w)R_w \cos\left(\frac{\pi-\theta}{2}\right) \quad (10)$$

where R_w and R_a are the van der Waals radii for water and the atom in the solute molecule respectively.

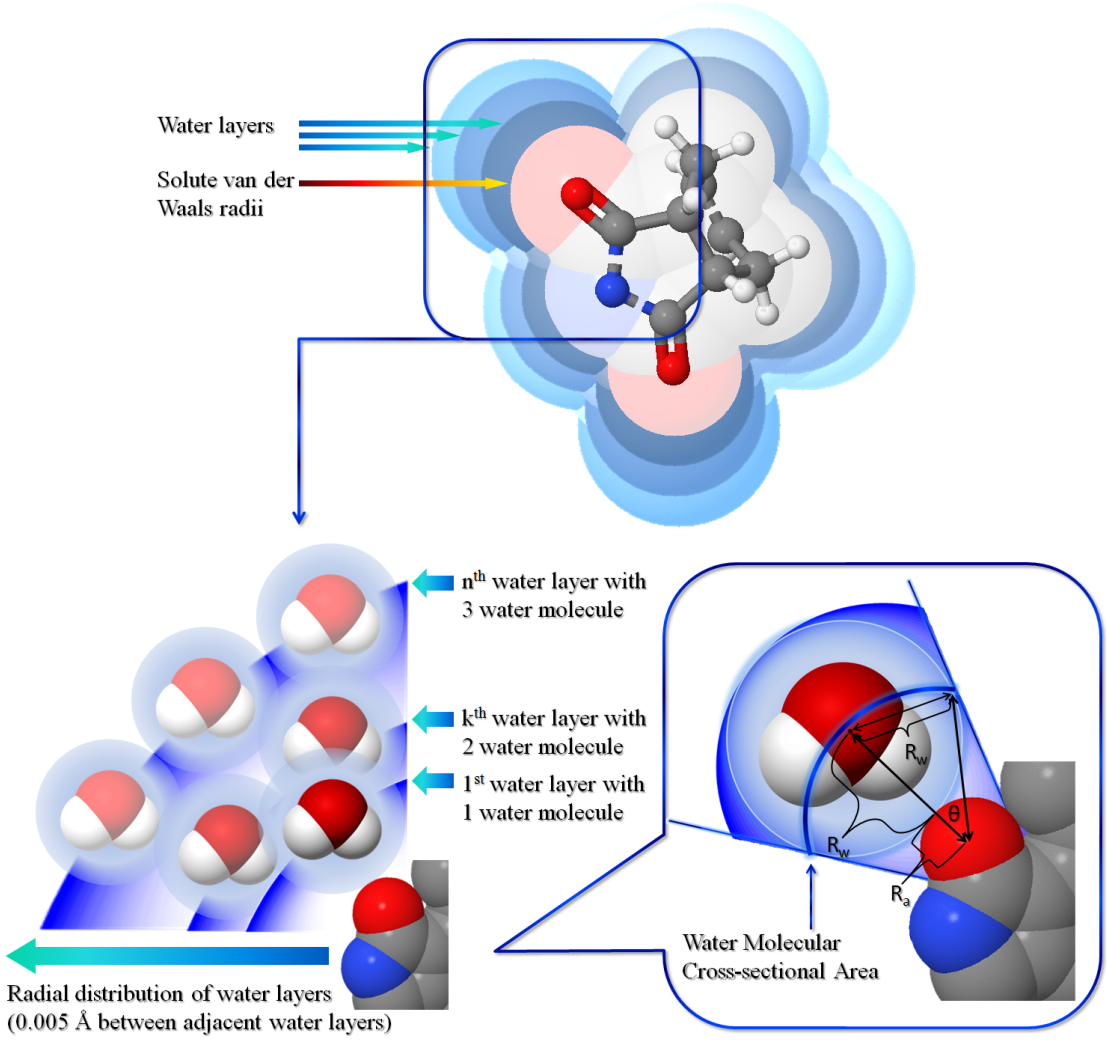

Figure 1. Modeling of the implicit solute – solvent model using the movable type method.

The Boltzmann factor matrix for the kth solute atom-water interaction (Z_k^{A-S}) is defined as a Boltzmann weighted solute atom-water energy multiplied by the number of accessible water molecules at the different distances. The final solute molecule-water Z-matrix (Z_{total}^{L-S}) is composed of the multiplication of the Z-matrices for all solute atom-water interactions. The final Z-matrix for the solute-solvent complex system (Z_{total}^{L-S}) is derived by the multiplication of the Z-matrix for the intra-solute molecular interactions (Z_{total}^{L}). Multiplication of the final Z-matrix with its corresponding normalized Q-matrix constructs the Boltzmann-weighted energy ensemble (C_{total}^{LS}), and then the solvation free energy is calculated by Equation 14.

$$\mathrm{Z}_k^{A-S} = \begin{bmatrix} e^{-\beta E_k^{A-S}(r_1)N_w(r_1)} & e^{-\beta E_k^{A-S}(r_{i+1})N_w(r_{i+1})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+1})N_w(r_{n-i+1})} \\ e^{-\beta E_k^{A-S}(r_2)N_w(r_2)} & e^{-\beta E_k^{A-S}(r_{i+2})N_w(r_{i+2})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+2})N_w(r_{n-i+2})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_k^{A-S}(r_i)N_w(r_i)} & e^{-\beta E_k^{A-S}(r_j)N_w(r_j)} & \cdots & e^{-\beta E_k^{A-S}(r_n)N_w(r_n)} \end{bmatrix} \tag{11}$$

$$\mathrm{Z}_{total}^{L-S} = disorder(\mathrm{Z}_1^{A-S}) \times disorder(\mathrm{Z}_2^{A-S}) \times \cdots \times disorder(\mathrm{Z}_n^{A-S}) \tag{12}$$

$$\mathrm{C}_{total}^{LS} = \overline{Q_{total}^{LS}} \times \overline{\mathrm{Z}_{total}^{LS}} = \overline{Q_{total}^{L-S}} \times \overline{Q_{total}^{L}} \times \overline{\mathrm{Z}_{total}^{L-S}} \times \overline{\mathrm{Z}_{total}^{L}} \tag{13}$$

$$\Delta G_{solv}^{L} \approx -RT \ln\left[\frac{Z_{LS}}{Z_L}\right] = -RT \ln\left[\frac{\int e^{-\beta E_{LS}(r)}\, dr}{\int e^{-\beta E_L(r)}\, dr}\right] =$$

$$-RT \ln\left[\frac{sum(\mathrm{C}_{total}^{LS})}{sum(\mathrm{C}_{total}^{L})}\right] = -RT \ln\left[\frac{sum(\overline{Q_{total}^{LS}} \times \overline{\mathrm{Z}_{total}^{LS}})}{sum(\overline{Q_{total}^{L}} \times \overline{\mathrm{Z}_{total}^{L}})}\right] \tag{14}$$

2.2 KECSA Energy Function

2.2.1 Data Collection

The first requirement to assemble a statistical potential is the collection of structural information. The Cambridge Structural Database (CSD) provides many crystal structures of small molecules that are co-crystalized with water molecules.[188] Besides, the Protein Data Bank (PDB) also has plenty of protein-ligand complexes with water molecules that are at the interface between binding pockets and ligand molecules, even though the resolution of those structures are usually poor compared to the structures in CSD. The aim of this research is to build a solvation free energy model especially for small molecules; hence the main resource for structural data collection was the CSD small molecule database. The criteria for data mining in CSD were set as: (1) the structures with an R factor less than 0.1 were required; (2) all polymer structures and molecules with ions were excluded. 7085 small molecules with crystal water molecules were obtained as our data set at last.

2.2.2 Atom Type Recognition

Our energy function methods are “fixed-charge” models for the selected atom or residue pairs; in other words, the statistical potentials are obtained by converting the number density distributions between two atoms or residues to energies. Many same atoms posses different electron densities however; in order to differentiate them from each other, a detailed atom type categorization has been constructed as shown in Table 1. 21 atom types were used in our current solvation model based on the structure information of the data set from CSD.

Table 1. List of 21 atom types in the current solvation model

Atom Type	Description
C1	sp^1 carbon
C2	sp^2 carbon
C3	sp^3 carbon
Car	Aromatic carbon
N2	sp^2 nitrogen
N3	sp^3 nitrogen
N4	Positively charged nitrogen
Nam	Amide nitrogen
Nar	Aromatic nitrogen
Npl3	Trigonal planar nitrogen
Ow	Water oxygen
O2	sp^2 oxygen
O3	Hydroxyl oxygen
OE	Ether and ester sp^3 oxygen
Oco2	Carboxylate, sulfate, & phosphate oxygen
S2	sp^2 sulfur
S3	sp^3 sulfur
P3	sp^3 phosphorus
F	Fluorine
Cl	Chlorine
Br	Bromine

2.2.3 Energy Function Modeling

The direct pairwise contacts combined with the indirect environmental effects contribute to the results of pairwise radial distribution. This is generally described as a "mean force"-driven correlation function. For the atom pairwise radial distribution, the mean force potential is usually defined as the following equation, where P_k is the mean force potential, $E_k(r)$ is the energy between the kth particle and any particle $i \in \{a, b, \cdots n\}$ in this system at a distance r separating these two particles.

$$P_k = \frac{\int E_k(r) e^{-E_k(r)/RT} \, dr_{ak} \, dr_{bk} \cdots dr_{nk}}{\int e^{-E_k(r)/RT} \, dr_{ak} \, dr_{bk} \cdots dr_{nk}} \tag{15}$$

Basically, statistical potentials have issues with analyzing various chemical environment effects on the observed atoms, thus would create a primary source of error in the potential models.[169] As noted by Thomas and Dill,[184] overrepresented contacts in a structural database could mask the presence of other contacts. Quantitatively minor contacts are generally underestimated in statistical potentials, while the reference state presumes a uniform availability of all contacts. KECSA is a new statistical potential energy function that was developed by our group; it defines a new reference state to eliminate the contact masking due to quantitative preferences.[189]

KECSA defines a reference state energy or background energy as the energy contributed by all atoms surrounding the observed atom pairs, unlike the traditional

statistical potentials using a reference state mimicking the ideal gas state. It introduces a reference state number distribution modeled by a linear combination of the number distribution under mean force ($n_{ij}(r)$) and the number distribution of an ideal gas state ($\left(\frac{N_{ij}}{V}\right) 4\pi r^2 \Delta r$):

$$n_{ij}^{**}(r) = \left(\frac{N_{ij}}{V} 4\pi r^2 \Delta r\right) x + (n_{ij}(r))(1 - x) \quad (16)$$

where x represents the intensity of the observed atom pairwise interaction in the chemical space V. This definition puts the number distribution of a certain observed atom pair in the reference state somewhere between the ideal gas state and the mean force state, depending on its relative strength. Stronger interactions have background energies closer to an ideal gas state while weaker interactions have background energies approaching the mean force state energy contributed by all atoms in the chemical space.

It requires us to define an "x" term for each atom pairwise interaction to build a KECSA energy function for modeling solvent, solute and solvent-solute interactions. Several methods have been used to model x in our knowledge-based energy function. In the original KECSA, the number ratio of the chosen atom pair *i* and *j* over the total atom pairs in the chemical space was used to represent the intensity x for simplicity. At the same time, we assigned an identical x for every atom pair found in the given chemical space based on the assumption that all contacts are

uniformly available in the chemical space given by the selected database.[154] As shown in Equation 17, N_t is the total atom type number in the chemical space.

$$n_{ij}^{**}(r) = \left(\frac{N_{ij}}{V} 4\pi r^2 \Delta r\right) x + \left(n_{ij}(r)\right)(1-x)$$

$$= \left(\frac{N_{ij}}{V} 4\pi r^2 \Delta r\right) \frac{1}{N_t} + \left(n_{ij}(r)\right)\left(1 - \frac{1}{N_t}\right) \quad (17)$$

The original model of $n_{ij}^{**}(r)$ is based on the notion that every atom pair has an equal contact opportunity in a background energy contributed by the other atom pairs, while neglecting the fact that the background energies have different effects on atom pairwise distributions with different interaction strengths (say atom i and j under a covalent bond constraint compared to atom k and l under a non-bond interaction constraint).

In order to take every atom pairwise contact as an energy state distributed between an ideal gas state energy and mean force state energy following a Boltzmann distribution in the reference state, a more accurate $n_{ij}^{**}(r)$ model is introduced. Based on that, the x factor is defined as Equation 18, where $e^{-\beta E_{ij}(r)}$ is the Boltzmann factor and $N_{ij}(r)$ is the degeneracy factor (contact number) for atom type i and j.

$$x = \frac{N_{ij}(r) e^{-\beta E_{ij}(r)}}{\sum_i^n \sum_j^n N_{ij}(r) e^{-\beta E_{ij}(r)}} \quad (18)$$

The number distribution of the observed atom pair in the background state $n_{ij}^{**}(r)$ is modeled as Equation 19, with the x term built up as a probability of all contacts.

$$n_{ij}^{**}(r) = \left(\frac{N_{ij}}{V} 4\pi r^2 \Delta r\right) \frac{N_{ij}(r) e^{-\beta E_{ij}(r)}}{\sum_i^n \sum_j^n N_{ij}(r) e^{-\beta E_{ij}(r)}} + (n_{ij}(r)) \left(1 - \frac{N_{ij}(r) e^{-\beta E_{ij}(r)}}{\sum_i^n \sum_j^n N_{ij}(r) e^{-\beta E_{ij}(r)}}\right) \quad (19)$$

Finally, the energy function for each atom type pair is built as Equation 20 shows.

$$E_{ij}(r) = -\frac{1}{\beta} \ln\left(\frac{n_{ij}(r)}{n_{ij}^*(r)}\right) = -\frac{1}{\beta} \ln\left[\frac{n_{ij}(r)}{\left(\frac{N_{ij}}{V} 4\pi r^2 \Delta r\right) x + \left(n_{ij}(r)\right)(1-x)}\right] = \frac{1}{\beta} \ln\left[x \left(\frac{N_{ij} 3 r^2 \Delta r}{R^3 n_{ij}(r)}\right) + (1-x)\right] = \frac{1}{\beta} \ln\left[\frac{N_{ij}(r) e^{-\beta E_{ij}(r)}}{\sum_i^n \sum_j^n N_{ij}(r) e^{-\beta E_{ij}(r)}} \left(\frac{N_{ij} 3 r^2 \Delta r}{R^3 n_{ij}(r)}\right) + \left(1 - \frac{N_{ij}(r) e^{-\beta E_{ij}(r)}}{\sum_i^n \sum_j^n N_{ij}(r) e^{-\beta E_{ij}(r)}}\right)\right] \quad (20)$$

In Equation 20, each $E_{ij}(r)$ can be derived iteratively at discrete distance point, with the energy functions built up in the chosen chemical space. By using this model, every $E_{ij}(r)$ derived using the KECSA energy function is never a mean force potential between atom type i and j as found in traditional statistical potentials.[189] Instead, $E_{ij}(r)$ represents pure atom pairwise interaction energy between atom type i and j, since the reference state energy defined in KECSA is a background energy contributed by all other atom pairs and not just the ideal gas state energy.

2.3 Test Set Selection

There are two main differences between KMTISM and other continuum solvation models, which are (1) unlike the traditional continuum solvation models, which separate the Gibbs free energies into linear enthalpy and entropy components, the KMTISM calculates the free energy change using a ratio of partition functions in the NVT ensemble.[139] (2) Electrostatic interactions are calculated explicitly using classical or QM based energy calculation approaches while they are implicit via the categorization of pairwise atom-types in the KECSA model. In this manner, KMTISM can be viewed as the null hypothesis for the addition of explicit electrostatic interactions. If electrostatic interactions are added to a model, it should outperform the knowledge-based approach; if not, the explicit electrostatic model is not an improvement over the implicit inclusion of this key interaction. The validity of using pre-constructed atom-type pairwise energy data in free energy calculation for molecules with similar atoms, which differ in their chemical environment is a major concern for the KMTISM method. So, some compounds that contain C, O, N, S, P and halogen atoms with different functional groups were tested for the KMTISM examination. Since that KECSA energy function was parameterized by using organic structure data, the validation of KMTISM mainly focused on reproducing the aqueous solvation free energy of drug-like molecules. The Minnesota Solvation Database, which includes aqueous solvation free energies for 391 neutral molecules and 144 ions, is a well-constructed data set, and also meet our requirement quite well. Among the 391 neutral molecules and 144 ions in the database, 372 neutral

molecules and 21 ions were selected as our test set after excluding those inorganic molecules, and those molecules with atom types not represented in the KECSA potential. Various hydrocarbons, mono- and poly-functional molecules with solvation free energies ranging from -85 to 4 kcal/mole were included in this test set. This test set was further categorized into various subsets based on the functional group within the molecules. Because of the poly-functional nature of some molecules, some of the molecules were included in several subsets.

Carbon, nitrogen, and oxygen are key elements in organic molecules. More than one-half of the compounds in the neutral test set (219 out of 372 compounds) were composed exclusively of carbon, nitrogen and oxygen atoms. Four subsets from those 219 molecules from the Minnesota Solvation Database were created, which included 41 hydrocarbons, 91 molecules with oxygen-based functional groups, 44 molecules with nitrogen-based functional groups, and 43 molecules with mixed nitrogen and oxygen functional groups. Molecules with sulfur, phosphorus and halogen atoms were also tested in validation. In order to avoid interference from other polar atoms, a test set with only halocarbons was also created. On the other hand, sulfur and phosphorus are often contained in oxyacid groups in organic molecules. A test set with molecules that contain sulfur or phosphorus was selected from the neutral data set. Heterocyclic compounds, amides, and their analogs are pervasive in drug-like molecules and are well represented in the Minnesota Solvation Database. 37 heterocyclic compounds and 33 amides and their analogs were classified into two subsets. In addition, a challenging test with complex and

highly polar molecules was also constructed, which included 28 selected molecules containing three or more different functional groups. The ion test set in this research was mainly positively charged nitrogen and negatively charged carboxylate oxygen subsets, which was limited to biologically relevance of the ions. In this way, 11 cations and 10 anions, namely 21 ions in total, were selected from Minnesota Solvation Database. Alkoxide ions among others present in the Minnesota Solvation Database were excluded herein, but will be examined with the aid of molecular dynamics simulation of ion-water interactions for those ions in the future.

CHAPTER 3. RESULTS AND DISCUSSION

3.1 Comparison with MM-GBSA and MM-PBSA Results

The Solvation free energies were calculated by KMTISM and the results were compared with MM-GBSA and MM-PBSA for all subsets. Both MM-GBSA and MM-PBSA calculation were carried out by using AMBER with the General AMBER force field (GAFF). The MM-GBSA parameters were set as igb = 2 and saltcon = 0.100. For the MM-PBSA part, istrng was set as 0.100.

Against the neutral molecule test set, KMTISM and MM-PBSA gave comparable correlation coefficients (R^2) and both had a better correlation than MM-GBSA. Based on the values of Kendall's τ, MM-PBSA outperformed the other two methods in ranking ability, with KMTISM as the second best. In terms of accuracy of the models, KMTISM has the lowest root mean square error (RMSE), while the RMSE values for MM-GBSA and MM-PBSA were almost twice as large. The experimental data versus calculated data is shown in Figure 2, and the statistical analyses are given in Table 2 and Table 3.

A linear scaling model was applied to all three models using Equation 21 in order to bring their respective regression lines closer to $y = x$. Linear scaling provided a way to examine the deviation of the calculated results from their regression lines, but did not improve the performance of the methods. In Equation 21, a and b are the slope and the intercept of the regression line between experimental data and computed data, respectively.

$$y_{corrected} = \frac{y_{raw} - b}{a} \quad (21)$$

Table 2. Performance of KMTISM, MM-GBSA, and MM-PBSA for the prediction of the solvation free energies of neutral molecules.

	Total neutral molecule set			Amide set		
	KMTISM	MM-GBSA	MM-PBSA	KMTISM	MM-GBSA	MM-PBSA
R^2	0.792	0.734	0.804	0.660	0.493	0.509
Kendall's τ	0.755	0.708	0.793	0.568	0.484	0.465
Raw RMSE (kcal/mole)	2.597	4.629	4.647	4.368	8.666	9.717
Scaled RMSE (kcal/mole)	2.248	2.634	2.160	3.852	4.885	4.663
	Hydrocarbon set			Halocarbon set		
	KMTISM	MM-GBSA	MM-PBSA	KMTISM	MM-GBSA	MM-PBSA
R^2	0.699	0.906	0.954	0.648	0.004	0.594
Kendall's τ	0.663	0.748	0.887	0.656	0.091	0.625
Raw RMSE (kcal/mole)	0.858	1.179	0.925	1.052	2.768	1.148
Scaled RMSE (kcal/mole)	0.845	0.498	0.332	1.030	2.063	1.109
	Oxygenated molecule set			Organo-sulfur and -phosphorus set		
	KMTISM	MM-GBSA	MM-PBSA	KMTISM	MM-GBSA	MM-PBSA
R^2	0.829	0.881	0.916	0.762	0.751	0.777
Kendall's τ	0.657	0.723	0.754	0.680	0.626	0.618
Raw RMSE (kcal/mole)	2.104	4.232	3.868	4.337	8.297	9.179
Scaled RMSE (kcal/mole)	1.578	1.613	1.186	3.500	4.316	3.992

Table 2. (Cont'd)

	Nitrogenous molecule set			Heterocycle set		
	KMTISM	MM-GBSA	MM-PBSA	KMTISM	MM-GBSA	MM-PBSA
R^2	0.615	0.485	0.795	0.604	0.528	0.552
Kendall's τ	0.420	0.412	0.592	0.652	0.622	0.646
Raw RMSE (kcal/mole)	2.384	2.416	1.690	4.314	7.584	8.722
Scaled RMSE (kcal/mole)	2.276	2.555	1.797	3.721	4.413	4.217
	Oxygenated and nitrogenous			Polyfunctional molecule set		
	KMTISM	MM-GBSA	MM-PBSA	KMTISM	MM-GBSA	MM-PBSA
R^2	0.545	0.747	0.694	0.736	0.615	0.650
Kendall's τ	0.565	0.663	0.621	0.726	0.577	0.609
Raw RMSE (kcal/mole)	3.259	4.282	5.043	4.688	10.138	11.132
Scaled RMSE (kcal/mole)	2.991	2.794	2.484	3.597	5.335	4.804

Table 3. Performance of KMTISM, MM-GBSA and MM-PBSA for the prediction of the solvation free energies of ions

	Ion set		
	KMTISM	MM-GBSA	MM-PBSA
R^2	0.352	0.000	0.003
Kendall's τ	0.258	-0.057	-0.067
RMSE (kcal/mole)	5.777	11.736	10.481
	Carboxylate set		
	KMTISM	MM-GBSA	MM-PBSA
R^2	0.239	0.161	0.166
Kendall's τ	-0.090	-0.180	-0.180
RMSE (kcal/mole)	5.337	11.918	11.252
	Charged amine set		
	KMTISM	MM-GBSA	MM-PBSA
R^2	0.557	0.008	0.009
Kendall's τ	0.491	-0.127	-0.127
RMSE (kcal/mole)	6.149	11.569	9.727

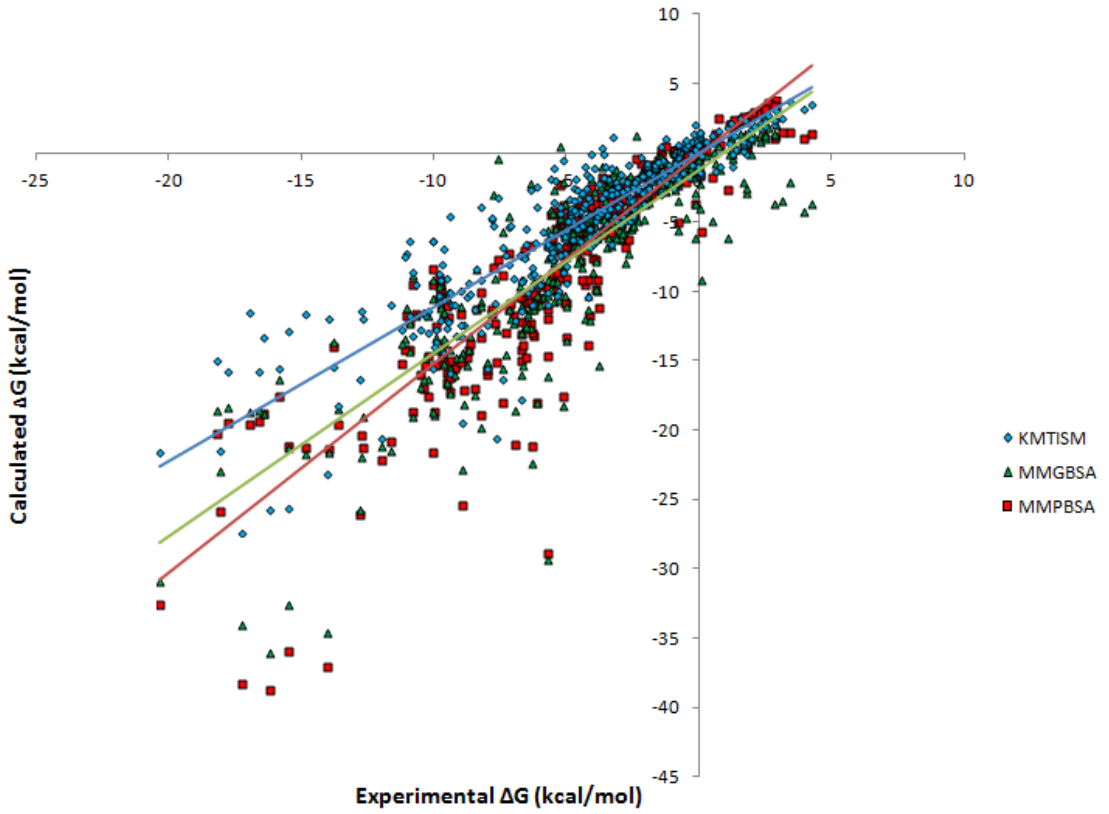

Figure 2. KMTISM, MM-GBSA, and MM-PBSA calculated versus experimental solvation free energies (kcal/mole) for 372 neutral molecules (kcal/mole).

The MM-GBSA and MM-PBSA results suggested a biased regression against the experimental solvation energies ($y = 1.3186x - 1.2902$ for MM-GBSA and $y = 1.5095x - 0.1556$ for MM-PBSA, where x and y represent the experimental and calculated solvation free energies). The slopes of their regression lines indicated an overestimation of the solvation free energies using these two methods. The significant improvement in RMSE values for MM-GBSA and MM-PBSA after the linear scaling as well as their correlation coefficient (R^2 and Kendall's τ) values show that they have a better ranking ability than free energy prediction. On the

other hand, KMTISM's regression function ($y = 1.1078x - 0.0811$) affected the RMSE to a lesser extent.

Results for different test sets categorized by functional groups provide deeper insights into the prediction abilities of the three computational methods. Errors increased with the complexity of the functional groups fro all three models. As Figure 3 and Figure 4 shown, solvation free energies of hydrocarbons, and oxygen and nitrogen containing molecules were better reproduced than molecules with other functional groups, while amides and mixed polyfunctional groups resulted in the largest RMSEs.

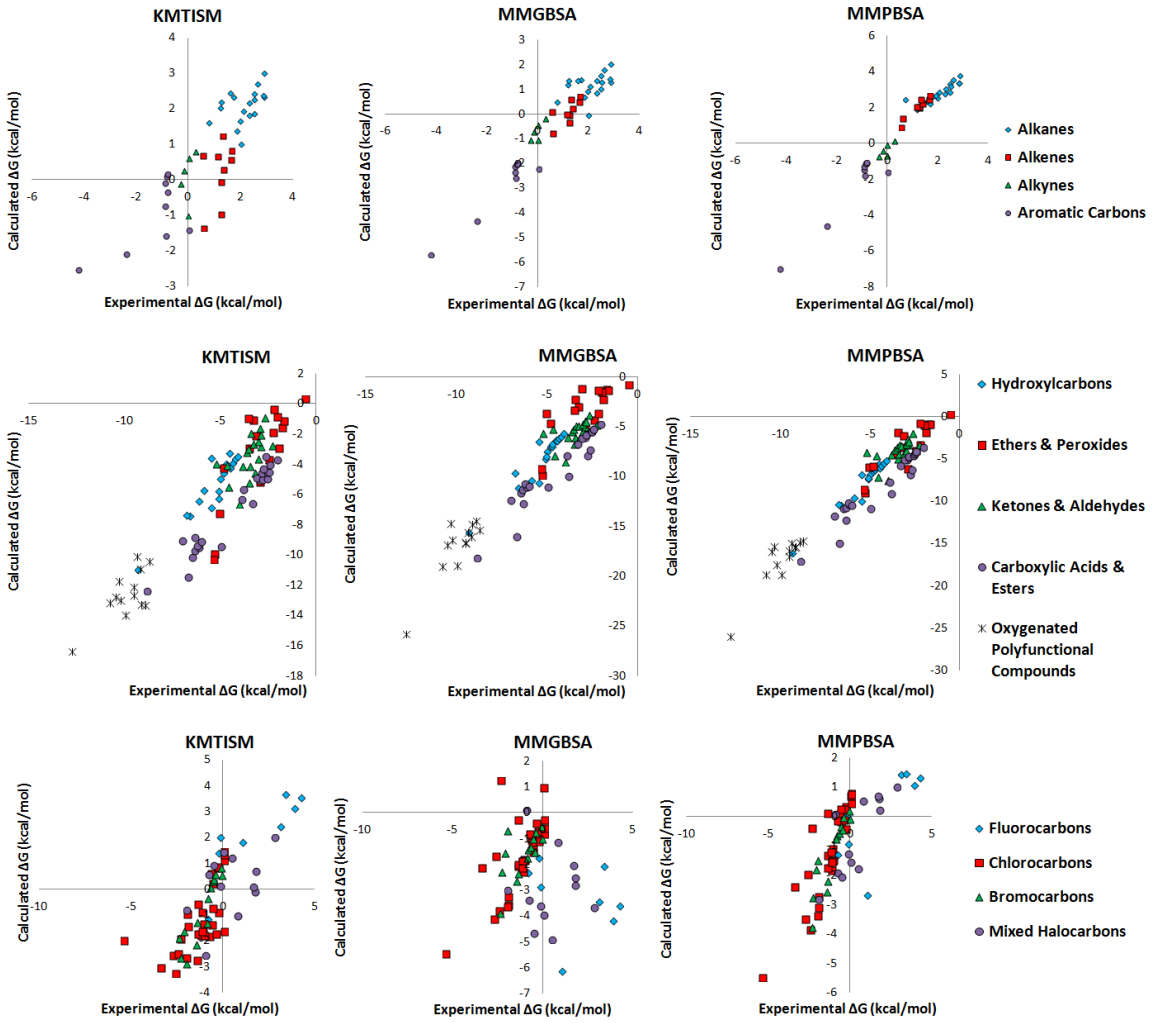

Figure 3. KMTISM's top three performing test sets according to RMSE. KMTISM, MM-GBSA, AND MM-PBSA calculated solvation free energies (kcal/mole) versus experimental data are listed from left to right. From the top to bottom the test sets are the hydrocarbon, oxygen containing, and halocarbon test sets.

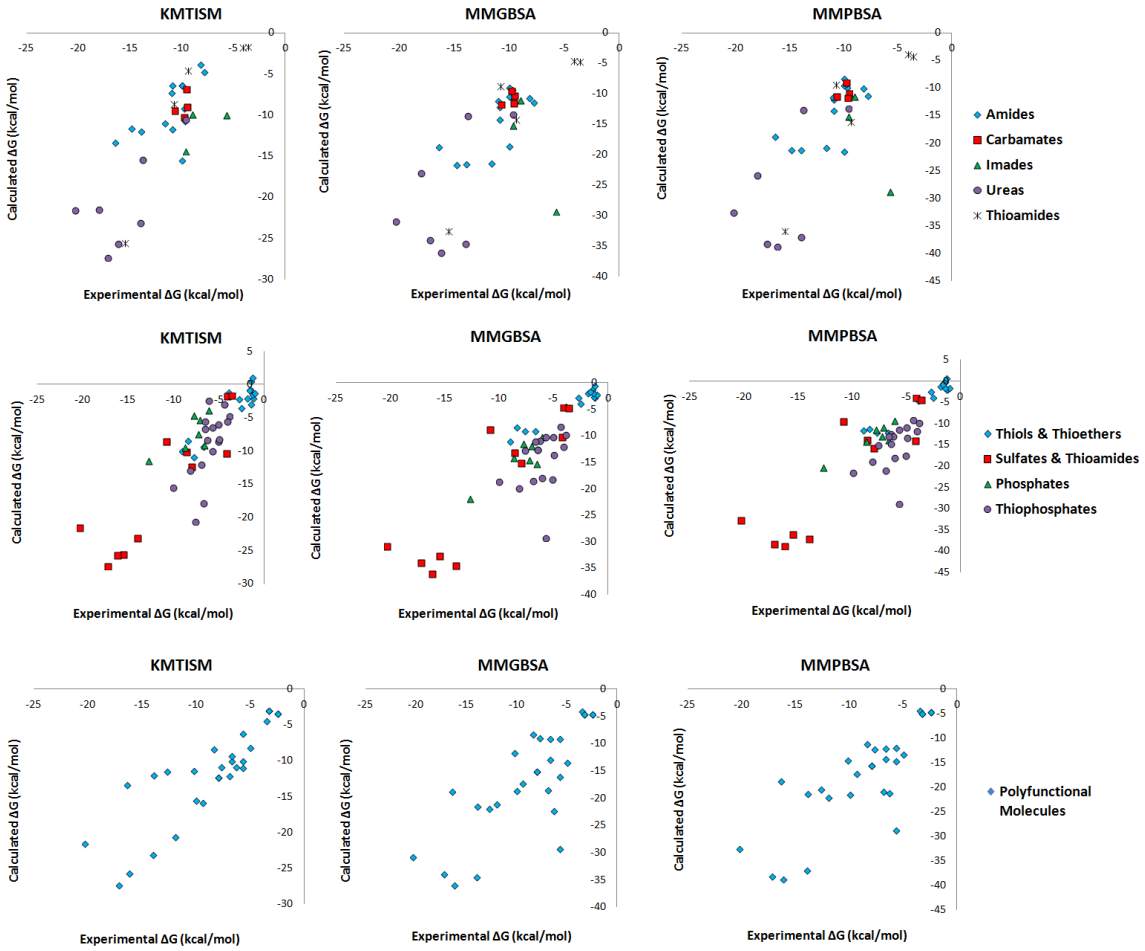

Figure 4. KMTISM's worst three performing test sets according to RMSE. KMTISM, MM-GBSA, and MM-PBSA calculated solvation free energies (kcal/mole) versus experimental data are listed from left to right. From the top to bottom the test sets are the amide, organosulfur and organophosphorus, and polyfunctional test sets.

Among all data sets, the hydrocarbon set was reproduced with the lowest RMSE values for all of the approaches, while unsaturated hydrocarbons proved more difficult for KMTISM than the other two models. The overestimation of the solvation free energies of unsaturated hydrocarbons causes the drop in R^2 for KMTISM, for

example, the two compounds with the largest error ($\sim 2\ kcal/mole$) were ethene and s-trans-1,3-butadiene, where all heavy atoms were sp^2 hybridized. In the training set of KECSA, which includes mostly drug-like molecules, different adjacent polar functional groups significantly altered the electron densities of adjacent unsaturated carbon atoms (via delocalization, for example) and this varies the electrostatic characteristics of these carbon atoms more than that seen in the case of sp^3 hybridized carbon.

On the other hand, polar atom types in the KECSA energy function were categorized according to their corresponding hydrophilic functional groups and were less affected by adjacent functional groups. Polar atom type-water radial probabilities were driven by a more fine grained atom pairwise set of interactions, thereby, improving the performance of the KECSA energy function for these groups. The top three test sets based on KMTISM's performance according to RMSE included the oxygenated molecule set and halocarbon set. Against the oxygen-containing molecule set, KMTISM gave a correlation coefficient comparable to MM-PBSA, while its RMSE was better than MM-GBSA. For the halocarbon set, the results of KMTISM were better than those of the MM-PBSA and MM-GBSA methods. For the data set of eight fluorocarbons, the RMSE for KMTISM was 1.1 kcal/mole, while the RMSE values for MM-GBSA and MM-PBSA were 5.8 kcal/mole and 2.2 kcal/mole, respectively.

The atom type classification process was an essential source of error for any statistical energy function, due to the variety of atom types in different molecules with different chemical environments. The use of atom types in classical potentials is also an issue, however it is typically mitigated by an explicit electrostatic model, which takes into account environmental effects. Collecting similar atom types into the same category can reduce the predictive ability of a statistical potential. For instance, with KECSA, the sp^3 oxygen atom in ethers and peroxides were modeled using the same atom type. This resulted in the solvation free energies for the two peroxides to be overestimated by KMTISM, that is, the ΔG_{sol} for methylperoxide was -9.90 kcal/mole or -8.86 kcal/mole (scaled) versus the experimental value of -5.28 kcal/mole and the ΔG_{sol} for ethylperoxide was -10.27 kcal/mole or -9.20 kcal/mole (scaled) versus the experimental value of -5.32 kcal/mole. In comparison with the MM-GBSA and MM-PBSA methods, the solvation free energy ΔG_{sol} for methylperoxide was -9.89 kcal/mole or -6.51 kcal/mole (scaled) using MM-GBSA and -9.07 kcal/mole or -5.90 kcal/mole (scaled) using MM-PBSA. The solvation free energy ΔG_{sol} for ethylperoxide was -9.21 kcal/mole or -6.00 kcal/mole (scaled) using MM-GBSA and -8.59 kcal/mole or -5.59 kcal/mole (scaled) using MM-PBSA. Therefore, none of the methods examined particularly did well modeling the solvation free energy of peroxides.

As the structural complexity of a molecule increased, the computed RMSE increased as well. Possible long range polar interactions add additional difficulty to accurate solvation free energy calculations using the methods described herein. The largest

errors were found in the amide set, organosulfur and organophosphorus set, and polyfunctional molecule set for all of the three methods. With lower errors for most individual polar functional groups based on the analysis of the monofunctional test set results, KMTISM had less cumulative error against these three test sets when compared with the MM-GBSA and MM-PBSA methods for both the raw RMSE values (see Table 2). This results shows that KMTISM has an advantage over the MM-GBSA and MM-PBSA methods for the prediction of the solvation free energy of polyfunctional molecules. This advantage will have an essential effect on the ability of this model to predict, for example, protein-ligand binding affinities, where the solvation free energy of the ligand can have an important impact on binding affinity prediction.

As shown in Figure 5, the magnitude of the errors in the solvation free energies for the ion test sets were relatively poor for all three methods, however, KMTISM still showed better correlations and RMSE than the other two implicit water models, especially for the charged amine test set (see Table 3). While the error magnitude was large all methods, the percentage error is comparable to the neutral set.

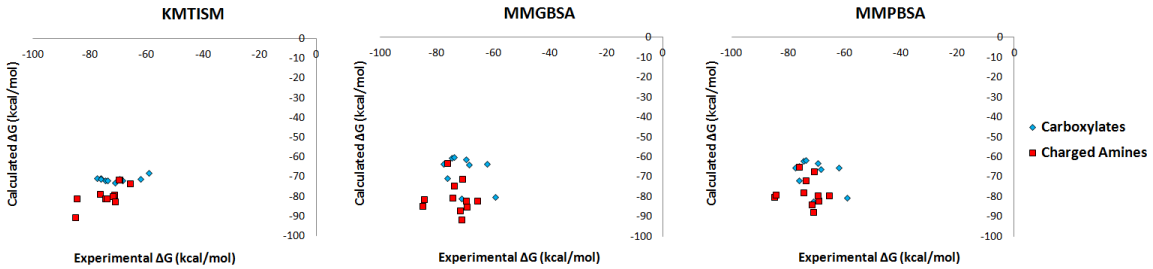

Figure 5. Solvation free energy results for KMTISM, MM-GBSA, and MM-PBSA against the ion test sets.

The carboxylate functional group, which is conjugated, lowered the accuracy of KMTISM's calculation, while charged amines, on the other hand, whose electron densities are more localized, were better reproduced by the KMTISM method.

3.2 Comparison with SMX Results

Overall, due to its higher computational expense, QM based solvation models have had limited application to the study of macromolecular systems, but are normally used to understand the effect of solvation on small molecules. A thorough analysis of KMTISM against QM based solvation models was not the focus of the present research, but a general comparison helps put the present work in perspective relative to more advanced models.

Cramer and Truhlar's Solvation Model (SMX) series has been developed over several decades and is considered to be one of the best methods available to calculate solvation free energies of small molecules.[78-126] The mean absolute errors (MAE) for solvation free energy prediction was reported as ranging from 0.57 to 0.84 kcal/mole by their most up-to-date models, for 274 neutral molecules using different QM methods. For 112 ions, the calculated solvation free energy MAEs ranged from 2.7 to 3.8 kcal/mole.[126] As for our KMTISM, a calculated solvation free energy MAEs of 1.8 kcal/mole for 372 neutral molecules was reproduced, while a MAE of 5.1 kcal/mole for 21 ions was obtained. The trend is quite clear that the latest SMX models are more accurate than KMTISM, as well as MM-GBSA and MM-PBSA, by ~1 kcal/mole for both the neutral molecules and ions as measured by MAE, even though the data sets tested were not the same. Nonetheless, the performance of our first generation KMTISM model is quite impressive, and we are quite confident that future versions of KMTISM can provide more accurate results.

CHAPTER 4. SUMMARY AND CONCLUSION REMARKS

MM-GBSA and MM-PBSA are two widely used implicit solvation models. KECSA-Movable Type Implicit Solvation Model (KMTISM), which uses Movable Type sampling method and combines with a statistical energy function developed by our group, the KECSA energy function, is a novel method to predict solvation free energies for small molecules and ions. It has shown to have a comparable or even better ability to predict the solvation free energy for several test sets chosen from the Minnesota Solvation Database. However all of these methods perform worse than the most recent SMX model reported by Cramer and Truhlar. The advantages of KMTISM is that it uses computed energies to directly determine free energies, rather than using the approximation that the free energy of solvation is a collection of linearly combined free energies, as is employed in many traditional continuum solvent models. Therefore, the Helmholtz free energy is calculated by the construction of the relevant partition functions. A novel sampling method, the MT method, which is able to estimate free energy, enthalpy, as well as entropy changes very fast, was employed to assemble those partition functions. The use of a statistical energy function, whose parameterization can have weak spots for atom types with high polarizabilities and uncommon atom types whose lack of available experimental data can produce issues, is the disadvantages of the KMTISM model. In the future, several aspects should be worked on. First, a detailed study of enthalpy changes and entropy changes using the MT method is going to be carried out. Then, the statistical energy terms derived from data collection of MD simulations of atom types with high polarizability and uncommon atom types in structural databases are going to be improved. Also, replacing the statistical energy

function with different force field based energy functions and combine them with the MT sampling method in order to affect the fast evaluation of thermodynamic quantities, is another aspect that we will focus on.

REFERENCES

REFERENCES

(1) 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.

(2) Schaefer, M.; van Vlijmen, H. W.; Karplus, M. Electrostatic Contributions to Molecular Free Energies In Solution. *Adv. Protein Chem.* **1998**, 51, 1.

(3) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical Prediction of Hydration Free Energies for SAMPL. *J. Comput.-Aided Mol. Des.* **2012**, 26, 551.

(4) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed–Charge Force Fields: Small–Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B.* **2007**, 111, 2242.

(5) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head–Gordon, T. Development of an Improved Four–Site Water Model for Biomolecular Simulations: TIP4P–Ew. *J. Chem. Phys.* **2004**, 120, 9665.

(6) García, A. E.; Sanbonmatsu, K. Y. Exploring the Energy Landscape of a Beta Hairpin in Explicit Solvent. *Proteins.* **2001**, 42, 345.

(7) Gibas, C. J.; Subramaniam, S. Explicit Solvent Models in Protein pKa Calculations. *Biophys. J.* **1996**, 71, 138.

(8) Gonçalves, P. F. B.; Stassen, H. Calculation of the Free Energy of Solvation from Molecular Dynamics Simulations. *Pure Appl. Chem.* **2004**, 76, 231.

(9) Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P. Multipole Electrostatics in Hydration Free Energy Calculations. *J. Comput. Chem.* **2011**, 32, 967.

(10) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies Using All–Atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput.-Aided Mol. Des.* **2010**, 24, 307.

(11) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, 6, 1509.

(12) Wang, L.; Voorhis, T. V. A Polarizable QM/MM Explicit Solvent Model for Computational Electrochemistry in Water. *J. Chem. Theory Comput.* **2012**, 8, 610.

(13) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed–Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, 5, 350.

(14) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, 122, 134508.

(15) Zhou, R.; Berne, B. J.; Germain, R. The Free Energy Landscape for β Hairpin Folding in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 14931.

(16) Scarsi, M.; Apostolakis, J.; Caflisch, A. Comparison of a GB Solvation Model with Explicit Solvent Simulations: Potentials of Mean Force and Conformational Preferences of Alanine Dipeptide and 1,2–Dichloroethane. *J. Phys. Chem. B.* **1998**, 102, 3637.

(17) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. Comparison of Continuum and Explicit Models of Solvation: Potentials of Mean Force for Alanine Dipeptide. *J. Phys. Chem.* **1996**, 100, 1439.

(18) Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009**, 5, 919.

(19) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. Solvent Models for Protein–Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations. *J. Comput. Chem.* **2000**, 22, 591.

(20) Vorobjev, Y. N.; Hermans, J. ES/IS: Estimation of Conformational Free Energy by Combining Dynamics Simulations with Explicit Solvent with an Implicit Solvent Continuum Model. *Biophys. Chem.* **1999**, 78, 195.

(21) Huißmann, S.; Likos, C. N.; Blaak, R. Explicit vs. Implicit Water Simulations of Charged Dendrimers. *Macromolecules.* **2012**, 45, 2562.

(22) Zhou, R. Free Energy Landscape of Protein Folding in Water: Explicit vs. Implicit Solvent. *Proteins.* **2003**, 53, 148.

(23) Levy, R. M.; Gallicchio, E. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* **1998**, 49, 531.

(24) Nymeyer, H.; García, A. E. Simulation of the folding equilibrium of alpha–helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 13934.

(25) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small–Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, 51, 769.

(26) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, 55, 117.

(27) Miertuš, S.; Tomasi, J. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* **1982**, 65, 239.

(28) Pascual–Ahuir, J. L.; Silla, E.; Tuñón, I. GEPOL: An improved description of molecular–surfaces. 3. A new algorithm for the computation of a solvent–excluding surface. *J. Comput. Chem.* **1994**, 15, 1127.

(29) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chem. Phys. Lett.* **1996**, 255, 327.

(30) Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, 107, 3210.

(31) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anistropic dielectrics. *J. Chem. Phys.* **1997**, 107, 3032.

(32) Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **1997**, 106, 5151.

(33) Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics, and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation and Numerical Applications. *J. Phys. Chem. B.* **1997**, 101, 10506.

(34) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A.* **1998**, 102, 1995.

(35) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chem. Phys. Lett.* **1998**, 286, 253.

(36) Barone, V.; Cossi, M.; Tomasi, J. Geometry optimization of molecular structures in solution by the polarizable continuum model. *J. Comput. Chem.* **1998**, 19, 404.

(37) Cammi, R.; Mennucci, B.; Tomasi, J. Second–order Møller–Plesset analytical derivatives for the polarizable continuum model using the relaxed density approach. *J. Phys. Chem. A.* **1999**, 103, 9100.

(38) Cossi, M.; Barone, V.; Robb, M. A. A direct procedure for the evaluation of solvent effects in MC–SCF calculations. *J. Chem. Phys.* **1999**, 111, 5295.

(39) Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct.* **1999**, 464, 211.

(40) Cammi, R.; Mennucci, B.; Tomasi, J. Fast evaluation of geometries and properties of excited molecules in solution: A Tamm–Dancoff model with application to 4–dimethylaminobenzonitrile. *J. Phys. Chem. A.* **2000**, 104, 5631.

(41) Cossi, M.; Barone, V. Solvent effect on vertical electronic transitions by the polarizable continuum model. *J. Chem. Phys.* **2000**, 112, 2427.

(42) Cossi M.; Barone, V. Time–dependent density functional theory for molecules in liquid solutions. *J. Chem. Phys.* **2001**, 115, 4708.

(43) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **2001**, 114, 5691.

(44) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, 117, 43.

(45) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C–PCM solvation model. *J. Comput. Chem.* **2003**, 24, 669.

(46) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, 105, 2999.

(47) Chipman, D. M. Reaction field treatment of charge penetration. *J. Chem. Phys.* **2000**, 112, 5558.

(48) Cancès, E.; Mennucci, B. Comment on 'Reaction field treatment of charge penetration'. *J. Chem. Phys.* **2001**, 114, 4744.

(49) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. Solvent Effects 5. The Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation on ab initio Reaction Field Calculations. *J. Phys. Chem.* **1996**, 100, 16098.

(50) Kirkwood, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.* **1934**, 2, 351.

(51) Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, 58, 1486.

(52) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. Solvent Effects 1. The Mediation of Electrostatic Effects by Solvents. *J. Am. Chem. Soc.* **1991**, 113, 4776.

(53) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. Hartree–Fock Second Derivatives and Electric Field Properties in a Solvent Reaction Field – Theory and Application. *J. Chem. Phys.* **1991**, 95, 8991.

(54) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. Solvent Effects 2. Medium Effect on the Structure, Energy, Charge Density, and Vibrational Frequencies of Sulfamic Acid. *J. Am. Chem. Soc.* **1992**, 114, 523.

(55) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. Solvent Effects 3. Tautomeric Equilibria of Formamide and 2–Pyridone in the Gas Phase and Solution: An ab initio SCRF Study. *J. Am. Chem. Soc.* **1992**, 114, 1645.

(56) Mobley, D. L.; Ii, A. E.; Fennell, C. J.; Dill, K. A. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B.* **2008**, 112, 2405.

(57) Su, Y.; Gallicchio, E. The non-polar solvent potential of mean force for the dimerization of alanine dipeptide: the role of solute-solvent van der Waals interactions. *Biophys. Chem.* **2004**, 109, 251.

(58) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, 98, 1978.

(59) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science.* **1995**, 268, 1144.

(60) Beroza, P.; Case, D. A. Calculation of Proton Binding Thermodynamics in Proteins. *Methods Enzymol.* **1998**, 295, 170.

(61) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. Biological Applications of Electrostatic Calculations and Brownian Dynamics. *Rev. Comput. Chem.* **1994**, 5, 229.

(62) Gilson, M. K. Theory of Electrostatic Interactions in Macromolecules. *Curr. Opin. Struct. Biol.* **1995**, 5, 216.

(63) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A.* **1997**, 101, 8098.

(64) Pei, J.; Wang, Q.; Zhou, J.; Lai, L. Estimating Protein–Ligand Binding Free Energy: Atomic Solvation Parameters for Partition Coefficient and Solvation Free Energy Calculation. *Proteins.* **2004**, 57, 651.

(65) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a β–hairpin folding in water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 12777.

(66) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, 94, 2027.

(67) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V.; Energies, Structures, and Electronic Properties of Molecules in Solution with the C–PCM Solvation Model. *J. Comput. Chem.* **2003**, 24, 669.

(68) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, 2, 128.

(69) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. Free Energy Surfaces of β–Hairpin and α–Helical Peptides Generated by Replica Exchange Molecular Dynamics with the AGBNP Implicit Solvent Model. *Proteins.* **2004**, 56, 310.

(70) Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, 78, 1.

(71) Kolář, M.; Fanfrlík, J.; Hobza, P. Ligand Conformational and Solvation/Desolvation Free Energy in Protein–Ligand Complex Formation. *J. Phys. Chem. B.* **2011**, 115, 4718.

(72) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand Solvation in Molecular Docking. *Proteins.* **1999**, 34, 4.

(73) Rees, D. C.; Wolfe, G. M. Macromolecular Solvation Energies Derived from Small Molecule Crystal Morphology. *Protein Sci.* **1993**, 2, 1882.

(74) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature*. **1986**, 319, 199.

(75) Lazaridis, T.; Karplus, M. Effective energy function for proteins in solution. *Proteins.* **1999**, 35, 133.

(76) Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1992**, 1, 227.

(77) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2.* **1993**, 5, 799.

(78) Cramer, C. J.; Truhlar, D. G. General Parameterized SCF Model for Free Energies of Solvation in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, 113, 8305.

(79) Cramer, C. J.; Truhlar, D. G. Molecular Orbital Theory Calculations of Aqueous Solvation Effects on Chemical Equilibria. *J. Am. Chem. Soc.* **1991**, 113, 9901.

(80) Cramer, C. J.; Truhlar, D. G. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. *Science.* **1992**, 256, 213.

(81) Cramer, C. J.; Truhlar, D. G. PM3-SM3: A General Parameterization for Including Aqueous Solvation Effects in the PM3 Molecular Orbital Model. *J. Comput. Chem.* **1992**, 13, 1089.

(82) Cramer, C. J.; Truhlar, D. G. AM1-SM2 and PM3-SM3 Parameterized SCF Solvation Models for Free Energies in Aqueous Solution. *J. Comput.- Aided Mol. Des.* **1992**, 6, 629.

(83) Cramer, C. J.; Truhlar, D. G. Polarization of the Nucleic Acid Bases in Aqueous Solution. *Chem. Phys. Lett.* **1992**, 198, 74.

(84) Cramer, C. J.; Truhlar, D. G. Quantum Chemical Conformational Analysis of Glucose in Aqueous Solution. *J. Am. Chem. Soc.* **1993**, 115, 5745.

(85) Cramer, C. J.; Truhlar, D. G. Correlation and Solvation Effects on Heterocyclic Equilibria in Aqueous Solution. *J. Am. Chem. Soc.* **1993**, 115, 8810.

(86) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Entropic Contributions to Free Energies of Solvation. *J. Phys. Chem.* **1994**, 98, 4141.

(87) Cramer, C. J.; Truhlar, D. G. Quantum Chemical Conformational Analysis of 1,2-Ethanediol: Correlation and Solvation Effects on the Tendency to Form Internal Hydrogen Bonds in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1994**, 116, 3892.

(88) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved Methods for Semiempirical Solvation Models. *J. Comput. Chem.* **1995**, 16, 422.

(89) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. A General Semiempirical Quantum Mechanical Solvation Model for Nonpolar Solvation Energies. n-Hexadecane. *J. Am. Chem. Soc.* **1995**, 117, 1057.

(90) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. A Semiempirical Quantum Mechanical Solvation Model for Solvation Free Energies in All Alkane Solvents. *J. Phys. Chem.* **1995**, 99, 7137.

(91) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Screening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, 246, 122.

(92) Barrows, S. E.; Cramer, C. J.; Truhlar, D. G.; Elovitz, M. S.; Weber, E. J. Factors Controlling Regioselectivity in the Reduction of Polynitroaromatics in Aqueous Solution. *Environ. Sci. Technol.* **1996**, 30, 3028.

(93) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First-Solvation Shell Effects. *J. Phys. Chem.* **1996**, 100, 16385.

(94) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parameterized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, 100, 19824.

(95) Cramer, C. J.; Truhlar, D. G.; French, A. D. Exo-anomeric Effects on Energies and Geometries of Different Conformations of Glucose and Related Systems in the Gas Phase and Aqueous Solution. *Cat. Rev.* **1997**, 298, 1.

(96) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. A Solvation Model for Chloroform Based on Class IV Atomic Charges. *J. Phys. Chem.* **1997**, 101, 2061.

(97) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. New Methods for Potential Functions for Simulating Biological Molecules. *J. Chim. Phys.* **1997**, 94, 1448.

(98) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. What Controls the Partitioning of Nucleic Acid Bases Between Chloroform and Water? *J. Phys. Chem. B.* **1997**, 101, 5084.

(99) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parameterized Model for Aqueous Free Energies of Solvation Using Geometry-Dependent Atomic Surface Tensions with Implicit Electrostatics. *J. Phys. Chem. B.* **1997**, 101, 7147.

(100) Cramer, C. J.; Truhlar, D. G.; Falvey, D. E. Singlet-Triplet Splittings and 1,2-Hydrogen Shift Barriers for Methylphenylborenide, Methylphenylcarbene, and Methylphenylnitrenium in the Gas Phase and Solution. What a Difference a Charge Makes. *J. Am. Chem. Soc.* **1997**, 119, 12338.

(101) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. A Universal Solvation Model for the Quantum Mechanical Calculation of Free Energies of Solvation in Non-Aqueous Solvents. *Theor. Chem. Acc.* **1997**, 98, 85.

(102) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Universal Quantum Mechanical Model for Solvation Free Energies Based on Gas-Phase Geometries. *J. Phys. Chem. B.* **1998**, 102, 3257.

(103) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. OMNISOL: Fast Prediction of Free Energies of Solvation and Partition Coefficients. *J. Org. Chem.* **1998**, 63, 4305.

(104) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Density Functional Solvation Model Based on CM2 Atomic Charges. *J. Chem. Phys.* **1998**, 109, 9117.

(105) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Universal Reaction Field Model Based on Ab Initio Hartree-Fock Theory. *Chem. Phys. Lett.* **1998**, 288, 293.

(106) Sullivan, M. B.; Brown, K.; Cramer, C. J.; Truhlar, D. G. Quantum Chemical Analysis of Para-Substitution Effects on the Electronic Structure of Phenylnitrenium Ions in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1998**, 120, 11778.

(107) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Analytical Energy Gradients of a Self-Consistent Reaction-Field Solvation Model Based on CM2 Charges. *J. Chem. Phys.* **1999**, 110, 5503.

(108) Li, J.; Cramer, C. J.; Truhlar, D. G. Application of a Universal Solvation Model to Nucleic Acid Bases. Comparison of Semiempirical Molecular Orbital Theory, Ab Initio Hartree-Fock Theory, and Density Functional theory. *Biophys. Chem.* **1999**, 103, 3802.

(109) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the Platform of Applicability of the SM5.42R Universal Solvation Model. *Theor. Chem. Acc.* **1999**, 103, 9.

(110) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, 99, 2161.

(111) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. A Universal Solvation Model Based on Class IV Charges and the Intermediate Neglect of Differential Overlap for Spectroscopy Molecular Orbital Method. *J. Phys. Chem. A.* **2000**, 104, 2178.

(112) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. A Universal Solvation Model Based on the Conductor-like Screening Model. *J. Comput. Chem.* **2000**, 21, 340.

(113) Winget, P.; Weber, E. J.; Cramer, C. J.; Truhlar, D. G. Computational Electrochemistry: Aqueous One-Electron Oxidation Potentials for Substituted Anilines. *Phys. Chem. Chem. Phys.* **2000**, 2, 1231.

(114) Winget, P.; Cramer, C. J.; Truhlar, D. G. Parameterization of a Universal Solvation Model for Molecules Containing Silicon. *J. Phys. Chem. A.* **2002**, 106, 5160.

(115) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C–PCM, D–PCM, and IEF–PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and for Vapor Pressures. *J. Phys. Chem. A.* **2004**, 108, 6532.

(116) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. Density-functional theory and hybrid density-functional theory continuum solvation models for aqueous and organic solvents: universal SM5.43 and SM5.43R solvation models for any fraction of Hartree-Fock exchange. *Theor. Chem. Acc.* **2005**, 113, 107.

(117) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *J. Chem. Theory Comput.* **2005**, 1, 1133.

(118) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. Predicting Aqueous Free Energies of Solvation as Functions of Temperature. *J. Phys. Chem. B*. **2006**, 110, 5665.

(119) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Adding Explicit Solvent Molecules to Continuum Solvent Calculations for the Calculation of Aqueous Acid Dissociation Constants. *J. Phys. Chem. A.* **2006**, 110, 2493.

(120) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory Comput.* **2007**, 3, 2011.

(121) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. Extension of a Temperature-Dependent Aqueous Solvation Model to Compounds Containing Nitrogen, Fluorine, Chlorine, Bromine, and Sulfur. *J. Phys. Chem. B.* **2008**, 112, 3024.

(122) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B.* **2009**, 113, 6378.

(123) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on the Generalized Born Approximation with Asymmetric Descreening. *J. Chem. Theory Comput.* **2009**, 5, 2447.

(124) Liu, J.; Kelly, C. P.; Goren, A. C.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Zhan, C.-G. Free Energies of Solvation with Surface, Volume, and Local Electrostatic Effects and Atomic Surface Tensions to Represent the First Solvation Shell. *J. Chem. Theory Comput.* **2010**, 6, 1109.

(125) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Sorting Out the Relative Contributions of Electrostatic Polarization, Dispersion, and Hydrogen Bonding to Solvatochromic Shifts on Vertical Electronic Excitation Energies. *J. Chem. Theory Comput.* **2010**, 6, 2829.

(126) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, 9, 609.

(127) Holst, M.; Baker, N.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson–Boltzmann Equation I. Algorithms and Examples. *J. Comput. Chem.* **2000**, 21, 1319.

(128) Baker, N.; Holst, M.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson–Boltzmann Equation II. Refinement at Solvent–Accessible Surfaces in Biomolecular Systems. *J. Comput. Chem.* **2000**, 21, 1343.

(129) Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A. Recent Progress in Numerical Methods for the Poisson−Boltzmann Equation in Biophysical Applications. *Commun. Comput. Phys.* **2008**, 3, 973.

(130) Nicholls, A.; Honig, B. A Rapid Finite Difference Algorithm, Utilizing Successive Over−Relaxation to Solve the Poisson−Boltzmann Equation. *J. Comput. Chem.* **1990**, 12, 435.

(131) Baker, N. A. Improving implicit solvent simulations: a Poisson–centric view. *Curr. Opin. Struct. Biol.* **2005**, 15, 137.

(132) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson–Boltzmann Equation for Biomolecular Electrostatics: a Tool for Structural Biology. *J. Mol. Recognit.* **2002**, 15, 377.

(133) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J. McCammon J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 10037.

(134) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson–Boltzmann Calculations for Static and Dynamic Systems. *J. Comput. Chem.* **2002**, 23, 1244.

(135) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, 112, 6127.

(136) Schaefer, M.; Froemmel, C. A Precise Analytical Method for Calculating the Electrostatic Energy of Macromolecules in Aqueous Solution. *J. Mol. Biol.* **1990**, 216, 1045.

(137) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects. *Theor. Chem. Acc.* **1999**, 101, 426.

(138) Sagui, C.; Darden, T. A. Molecular Dynamics Simulations of Biomolecules: Long–Range Electrostatic Effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, 28, 155.

(139) 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.

(140) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers.* **2001**, 56, 275.

(141) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, 51, 129.

(142) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized–Born Model. *J. Am. Chem. Soc.* **1999**, 121, 8033.

(143) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B.* **2000**, 104, 3712.

(144) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large–Scale Conformational Changes with a Modified Generalized Born Model. *Proteins.* **2004**, 55, 383.

(145) Dzubiella, J.; Swanson, J. M.; and McCammon, J. A. Coupling Nonpolar and Polar Solvation Free Energies in Implicit Solvent Models. *J. Chem. Phys.* **2006**, 124, 084905.

(146) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, 51, 69.

(147) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, 32, 866.

(148) Bello, M. Binding Free Energy Calculations Between Bovine β-Lactoglobulin and Four Fatty Acids Using the MMGBSA Method. *Biopolymers.* **2014**, 101, 1010.

(149) Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. *J. Mol. Biol.* **1990**, 213, 859.

(150) Miyazawa S.; Jernigan R. L. Estimation of effective interresidue contact energies from protein crystal structures: quasi–chemical approximation. *Macromolecules*. **1985**, 18, 534.

(151) Hendlich, M.; Lackner, P.; Weitckus, S.; Floeckner, H.; Froschauer, R.; Gottsbacher, K.; Casari, G.; Sippl, M. J. Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J. Mol. Biol*. **1990**, 216, 167.

(152) Jones, D. T.; Taylor, W. R.; Thornton, J. M. A new approach to protein fold recognition. *Nature*. **1992**, 358, 86.

(153) Thomas, P. D.; Dill, K. A. An iterative method for extracting energy–like quantities from protein structures. *Proc. Natl. Acad. Sci. USA*. **1996**, 93, 11628.

(154) Thomas, P. D.; Dill, K. A. Statistical potentials extracted from protein structures: How accurate are they? *J. Mol. Biol.* **1996**, 257, 457.

(155) Lu, H.; Skolnick, J. A Distance–Dependent Atomic Knowledge–Based Potential for Improved Protein Structure Selection. *Proteins: Struct. Funct. Genet.* **2001**, 44, 223.

(156) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein–Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, 42, 791.

(157) Muegge, I. A knowledge–based scoring function for protein–ligand interactions: Probing the reference state. *Perspect. Drug Discovery Des.* **2000**, 20, 99.

(158) Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **2001**, 22, 418.

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

(160) Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore(CSD)–knowledge–based scoring function derived from small molecule crystal data with superior recognition rate of near–native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, 48 (20), 6296.

(161) DeWitte,R. S.; Shakhnovich, E. I. SMoG: de Novo design method based on simple, fast, and accutate free energy estimate. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.* **1996**, 118, 11733.

(162) Ishchenko, A. V.; Shakhnovich, E. I. Small molecule growth 2001 (SMoG2001): An improved knowledge–based scoring function for protein–ligand interactions. *J. Med. Chem.* **2002**, 45, 2770.

(163) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP—potential of mean force describing protein–ligand interactions: I. Generating potential. *J. Comput. Chem.* **1999**, 20, 1165.

(164) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Forster, M. J.; Thornton, J. M. BLEEP – Potential of mean force describing protein–ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.* **1999**, 20 (11), 1177.

(165) Huang, S.–Y.; Zou, X. An iterative knowledge–based scoring function to predict protein–ligand interactions: II. Validation of the scoring function. *J. Comput. Chem.* **2006**, 27, 1876.

(166) Huang, S.–Y.; Zou, X. Inclusion of Solvation and Entropy in the Knowledge–Based Scoring Function for Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2010**, 50, 262.

(167) Kirkwood, J. G. Statistical Mechanics of fluid Mixtures. *J. Chem. Phys.* 1935, 3, 300.

(168) Huang, S.; Grinter, S. Z.; Zou, X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, 12, 12899.

(169) Huang, S. Y.; Zou, X. Mean–Force Scoring Functions for Protein–Ligand Binding. *Annu. Rep. Comput. Chem.* **2010**, 6, 280.

(170) Sippl, M. J. Knowledge–based potentials for proteins. *Curr. Opin. Struct. Biol.* **1995**, 5, 229.

(171) Samudrala, R.; Moult, J. An All–atom Distance–dependent Conditional Probability Discriminatory Function for Protein Structure Prediction. *J. Mol. Biol.* **1998**, 275, 895.

(172) Zhang, J.; Zhang, Y. A Novel Side–Chain Orientation Dependent Potential Derived from Random–Walk Reference State for Protein Fold Selection and Structure Prediction. *PloS one* **2010**, 5, e15386.

(173) Xu, D.; Zhang, J.; Roy, A.; Zhang, Y. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins.* **2011**, 79, 147.

(174) Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge–based force field. *Proteins.* **2012**, 80, 1715.

(175) Zhou, H.; Zhou, Y. Single–body residue–level knowledge–based energy score combined with sequence–profile and secondary structure information for fold recognition. *Proteins.* **2004**, 55, 1005.

(176) Melo, F.; Sánchez, R.; Sali, A. Statistical potentials for fold assessment. *Protein Sci.* **2002**, 11, 430.

(177) Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* **2011**, 27, 343.

(178) Skolnick, J.; Jaroszewski, L.; Kolinski, A.; Godzik, A. Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Sci.* **1997**, 6, 676.

(179) Krishnamoorthy, B.; Tropsha, A. Development of a four–body statistical pseudo–potential to discriminate native from non–native protein conformations. *Bioinformatics.* **2003**, 19, 1540.

(180) Frishman, D.; Argos, P. Incorporation of non–local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* **1996**, 9, 133.

(181) Lu, H.; Skolnick, J. A distance–dependent atomic knowledge–based potential for improved protein structure selection. *Proteins.* **2001**, 44, 223.

(182) Jernigan, R. L.; Bahar, I. Structure–derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **1996**, 6, 195.

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

(184) Thomas, P. D.; Dill, K. A. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* **1996**, 257, 457.

(185) Liu, S.; Zhang, C.; Zhou, H.; Zhou, Y. A physical reference state unifies the structure–derived potential of mean force for protein folding and binding. *Proteins.* **2004**, 56, 93.

(186) Tobi, D.; Elber, R. Distance–dependent, pair potential for protein folding: results from linear optimization. *Proteins.* **2000**, 41, 40.

(187) Zheng, Z.; Ucisik, M. N.; Merz, K. M. Jr. The Movable Type Method Applied to Protein–Ligand Binding. *J. Chem. Theory Comput.* **2013**, 9, 5526.

(188) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.* **2002**, B58, 380.

(189) Zheng, Z.; Merz, K. M. Jr. Development of the knowledge-based and empirical combined scoring algorithm (KECSA) to score protein-ligand interactions. *J. Chem. Inf. Model.* **2013**, 53,1073.

(190) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, 49, 1079.

(191) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, 48, 4111.

(192) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, 47, 2977.

(193) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database – version 2012, University of Minnesota, Minneapolis, **2012**.

(194) Mark, A. E.; van Gunsteren, W. F. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J. Mol. Biol.* **1994**, 240, 167.

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

(196) Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G. Non-additivity of functional group contributions in protein-ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *J. Mol. Biol.* **2010**, 397, 1042.

(197) Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water, *J. Mol. Liq.* **2002**, 101, 219.

(198) Narten, A. H.; Danford M. D.; Levy, H. A. X-Ray diffraction study of liquid water in the temperature range 4-200 °C, *Faraday Discuss.* **1967**, 43, 97.