9. Polarizable force fields for scoring protein-ligand interactions

Jiajing Zhang[+], Yue Shi[+] and Pengyu Ren[*]

Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas, 78712, USA
*Corresponding Author, Tel:1-512-232-1832, Fax:1-512-471-0616, Email:pren@mail.utexas.edu
[+]These authors contributed equally.

9.1 Introduction and overview

Molecular mechanics based protein-ligand binding free energy calculation has been an active research area and holds the promise for *in silico* drug discovery [1-8]. Despite the progress made on various fronts, such as the rigorous treatment of long-range electrostatic interactions and solvation models, challenges remain in order to make chemically accurate prediction of ligand binding affinities. Significant advancement in this field requires improvement on the effectiveness of configurational sampling and the accuracy of physics-based force field [1]. The current-generation force fields [9, 10] treat electrostatic contributions with fixed atomic partial charges and neglect electronic polarization due to local electrostatic environment. Such force fields have served as cornerstones in molecular dynamics simulations of biological systems over decades. However, the oversimplified fixed-atomic charge model is problematic and limits the further progress of molecular mechanics [9-12]. It has been long been noted that atomic charges are inadequate and higher order moments are needed for accurate representation of electrostatic potentials near molecules [13]. In addition, the charge distribution within a molecule can be significantly distorted due to induction. Such non-additive induction effect could account for 10-20% of the total interaction energy in typical molecular complexes at the vdW minimum [11], and more in highly charged systems [14]. Protein-ligand binding is a process where both molecules experience significant changes in the environment upon dehydration and association.

There has been growing effort to develop general force fields that explicitly treat the electronic polarization effect. There exist a few models for incorporating the polarization effect into the molecular mechanics framework, by means of point dipole induction [15], fluctuating charge [16] or Drude oscillator [17], each with its own advantages and drawbacks [10-12]. AMOEBA is the first general-purpose force field that has been utilized in molecular dynamics simulations of protein-ligand binding and calculation of absolute and relative binding free energy [18-20]. AMOEBA force field uses permanent atomic multipoles (up to quadrupoles) and accounts for polarization effect explicitly via atomic dipole induction [10, 20-22]. The development of AMOEBA model puts its emphasis on reproduction of molecular polarizability, electrostatic potential, as well as

interaction energy from gas to condensed-phase. These features enable AMOEBA to produce accurate electrostatic interaction and better transferability. In this chapter, we describe the effort to apply the AMOEBA polarizable force field to predict protein-ligand binding affinity with both explicit and implicit solvent methods. We begin by introducing the basic principles and formula underlying the AMOEBA force field. Along with small molecule and ion hydration free energy studies, we describe the calculation of binding free energy of benzamidine analogs to trypsin using AMOEBA-based molecular dynamics alchemical perturbation. Next, we review the theoretical development of continuum solvent models for polarizable solutes, including the polarizable multipole Poisson-Boltzmann surface area (PMPB/SA) [23] and Generalized Kirkwood surface area (GK/SA) [24] models. Both models are evaluated again in the calculation of binding affinities of the trypsin-benzamidine systems.

9.2 AMOEBA polarizable potential energy model

The interaction energy among atoms in the AMOEBA potential function is expressed as the sum of various contributions:

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \tag{9.1}$$

where the first five terms describe the short-range valence interactions: bond stretching, angle bending, bond-angle cross term, out-of-plane bending, and torsional rotation. The last three terms are the nonbonded van der Waals (vdW) and electrostatic contributions. The individual terms are elaborated in the following sections.

Bond, Angle and Cross Energy Terms

For atoms directly bonded and separated by two bonds, the covalent energies are represented by empirical functions of bond lengths and angles adopted from MM3 force field [25]. In addition, a Wilson-Decius-Cross function [26] is used at sp^2-hybridized trigonal centers to restrain the out-of-plane bending.

$$U_{bond} = K_b (b - b_0)^2 \left[1 - 2.55(b - b_0) + 3.793125(b - b_0)^2 \right] \tag{9.2}$$

$$\begin{aligned} U_{angle} = K_\theta (\theta - \theta_0)^2 \big[& 1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5} (\theta - \theta_0)^2 \\ & - 7.0 \times 10^{-7} (\theta - \theta_0)^3 + 2.2 \times 10^{-8} (\theta - \theta_0)^4 \big] \end{aligned} \tag{9.3}$$

$$U_{b\theta} = k_{b\theta} \left[(b - b_0) + (b' - b'_\theta) \right] (\theta - \theta_0) \tag{9.4}$$

$$U_{oop} = k_\chi \chi^2 \tag{9.5}$$

Torsional Energy Term

$$U_{torsion} = \sum_n K_{n\phi}[1 + \cos(n\phi \pm \delta)] \quad (9.6)$$

A traditional Fourier expansion (typically from 1 to 3) torsional functional is used to merge the short range "valence" with the long range "nonbonded" interactions. For dihedral angles involving two joined trigonal centers, such as the amide bond of the protein backbone, a Bell torsion [27] functional is applied in addition to the regular torsional terms to allow appropriately increased flexibility of atoms bonded to two trigonal centers (e.g. amide hydrogen atoms) [28]. In AMOEBA protein force field, the backbone torsional energy uses a bi-cubic interpolation function that depends on both phi and psi angles. This term essentially makes up the difference between AMOEBA and *ab initio* conformational energy on the dipeptide Ramachandran map, an approach first adopted by CHARMM [29]. Note the torsional contribution only accounts for partially the conformational energy, which is generally dominated by vdW and electrostatic interactions. In early development of AMOEBA, the LMP2/cc-pVTZ(-f) relative conformational energies on a 24 by 24 grid of the *phi-psi* surface are taken as target for AMOEBA to reproduce. Recently, we updated the QM energy map with extrapolated RI-TRIM MP2/aug-cc-pV(TQ)Z values [30].

Van der Waals Interactions

The pairwise additive vdW interaction in AMOEBA adopts the buffered 14-7 functional form [31]

$$U_{vdw}(ij) = \varepsilon_{ij}\left(\frac{1.07}{\rho_{ij}+0.07}\right)^7\left(\frac{1.12}{\rho_{ij}^7+0.12}-2\right) \quad (9.7)$$

where the potential well depth ε_{ij} is in kcal/mol, $\rho_{ij} = R_{ij}/R_{ij}^0$ where R_{ij} in Å, is the actual separation between *i* and *j*, and R_{ij}^0 is the minimum energy distance. For heterogeneous atom pairs, the combining rules are given by

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\varepsilon_{ii}^{1/2}+\varepsilon_{jj}^{1/2})^2} \text{ and } R_{ij}^0 = \frac{(R_{ii}^0)^3+(R_{jj}^0)^3}{(R_{ii}^0)^2+(R_{jj}^0)^2} \quad (9.8)$$

The buffered 14-7 form is considered superior as it provides a better fit to gas phase *ab initio* results and liquid properties of noble gases [31].

Permanent Electrostatic Interactions

The electrostatic energy in AMOBEA consists of contributions from both permanent and induced multipole moments. The permanent atomic multipole moments at each atomic center are consisted of the monopole, dipole and quadrupole moments,

$$M_i = \left[q_i, d_{ix}, d_{iy}, d_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz}, Q_{iyx}, Q_{iyy}, Q_{iyz}, Q_{izx}, Q_{izy}, Q_{izz}\right]^T \quad (9.9)$$

where q_i is the point charge located at the atom i, d is the dipole and Q is the quadrupole moment. In the Cartesian Polytensor formalism [32], the interaction energy between atoms i and j separated by r_{ji} is represented as $U_{ele}^{perm}(ij) = M_i^T T_{ij} M_j$, or in expanded form as

$$U_{ele}^{perm}(ij) = \begin{bmatrix} q_i \\ \mu_{ix} \\ \mu_{iy} \\ \mu_{iz} \\ Q_{ixx} \\ \vdots \end{bmatrix}^t \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \left(\frac{1}{r_{ji}} \right) \begin{bmatrix} q_j \\ \mu_{jx} \\ \mu_{jy} \\ \mu_{jz} \\ Q_{jxx} \\ \vdots \end{bmatrix} \tag{9.10}$$

The dipole and quadrupole are defined with respect to a local reference frame set by neighboring atoms [21]. The permanent atomic multipole moments remain constant with respect to their local frames as the molecule moves over time and space. At the beginning of each calculation, multipole moments at each site are first rotated from the local frame as defined in the parameter file into the current global coordinate frame by

$$M_{global} = \Re M_{local} \tag{9.11}$$

where $\Re$ is the rotation matrix transforming the local into the global reference frame [32].

Electronic Polarization

Electronic polarization refers to the distortion of electron density in response to the external field. It represents a significant contribution to the many-body effect experienced in molecular clusters or condensed phases, even though situations exist where other contributions related to charge-transfer, dispersion and repulsion are significant [13]. In the current AMOEBA model, a classical point dipole moment is induced at each polarizable site (typically atomic center) and proportional to the electric field generated by other permanent and induced multipoles. Molecular polarization is achieved via an interactive induction strategy with distributed atomic polarizabilities based on Thole's method [33]. This interactive or mutual induction scheme requires that an induced dipole produced at any site will further polarize all the other sites, and such mutual induction will continue until the induced dipoles at all sites achieve convergence. One key feature of Thole's approach is the damping of the polarization interaction at very short range to avoid a well-known artifact of point polarizability models, namely the "polarization catastrophe". The damping is effectively achieved by smearing one of the

atomic multipole moments in each pair of interaction sites. The smearing function for charges adopted by AMOEBA has the functional form

$$\rho = \frac{3a}{4\pi}\exp(-au^3) \tag{9.12}$$

where $u = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the effective distance as a function of linear separation r_{ij} and atomic polarizabilities of sites *i* (α_i) and *j* (α_j). The factor "*a*" is a dimensionless width parameter of the smeared charge distribution to effectively control the damping strength, and therefore avoids polarization catastrophes at small separations.

Starting from the ab initio atomic multipoles distributed multipole analysis (DMA) [34] for an arbitrary conformer of a model compound, M_i^{DMA}, one can derive the intrinsic "permanent" atomic multipole moments, M_i , that satisfy

$$M_i^{DMA} = M_i + \mu_i \tag{9.13}$$

where μ_i is the dipole induced by intramolecular polarization by M_i [22]. In other words, the combination of one set of permanent $\{M_i\}$ and conformation-dependent induced dipole $\{\mu_i\}$ moments is expected to reproduce the QM electrostatic potential of flexible molecules at various conformations.

Polarization Energy

Formally, the induced dipole vector on any polarizable site *i* can be expressed as

$$\mu_i^{ind} = \alpha_i\left(\sum_{j\neq i} T_{ij}^1 M_j + \sum_{k\neq i} T_{ik}^{11}\mu_k\right) \tag{9.14}$$

and the energy is

$$U_{ele}^{ind} = -\frac{1}{2}\sum_i \left(\mu_i^{ind}\right)^T E_i \tag{9.15}$$

where $T_{ij}^1 = \left[\nabla^1, \nabla^2, \nabla^3, ..\right]$ is a 3 × 13 matrix with $\nabla^{l+m+n} = \frac{\partial^l}{\partial x^l}\frac{\partial^m}{\partial y^m}\frac{\partial^n}{\partial z^n}$, representing the second through fourth rows of the matrix T in equation (9.10). $T_{ik}^{11} = \nabla_{ik}^2$, a 3×3 sub matrix, consists of elements in T_{ij}^1 corresponding to the dipole moments. In equation (9.14), the first term inside the parenthesis on the right hand side is the "direct" electric field, E, due to permanent multipoles from outside the group (index *j*). The second term corresponds to "mutual" induction by other induced dipoles (index *k*). Direction induction only occurs among groups while mutual polarization involves every atom pair. The induced dipole in equation (9.14) is solved iteratively to obtain the final induced dipoles. The convergence is accelerated via a successive over-relaxation fashion. The particle-mesh Ewald (PME) for polarizable point multipoles developed by Sagui and Darden [35] has been implemented in TINKER [36] and AMBER programs [37], which significantly improved the computational efficiency.

9.3 AMOEBA explicit water simulation applications

Alchemical transformation, which gently perturbs one chemical system to another, allows computing the free energy difference reliably. Techniques such as free energy perturbation (FEP) [38] and thermodynamic integration (TI) [39], are most widely employed to fulfill this process. Compared with fixed-charge models, AMOEBA polarizable model shows its advantage by including the explicit treatment of dipole polarization in respond to changing or heterogeneous environments [20]. Here we highlight a number of applications in which AMOEBA is utilized in alchemical simulations to compute hydration free energies of small molecules and ion solvation, and protein-ligand binding thermodynamics. In this rigorous approach, hydration free energy is an important component in the protein-ligand binding thermodynamic cycle.

9.3.1 Small molecule hydration free energy calculations

In recent years there have been extensive studies of hydration of organic small molecules using explicit solvent MD simulations and alchemical transformation [40, 41]. The mean unsigned error (MUE) of predicted hydration free energy (HFE) is typically over 1 kcal/mol and worse for polar molecules [42]. With AMOEBA polarizable force field, it is possible to obtain HFE within an accuracy of 0.4 kcal/mol (MUE) using gas-phase atomic multipoles from MP2/aug-cc-pVTZ, with the polarization modeled via induced atomic dipoles. While the improvement is only a fraction of kcal/mol, as we will show in the protein-ligand binding section, neglecting polarization will lead to significant errors in the protein environment and likely in the opposite direction to that in hydration.

To calculate the HFE using AMOEBA, the small molecule is decoupled from bulk water in several steps. First, the electrostatic interactions between the small molecule and water are turned off by scaling down the electrostatic parameters of the small molecules linearly in 10 steps. Subsequently the vdW interactions are turned off in 12 steps. During the decoupling of the vdW interactions, soft-core modifications [43] are introduced to the buffered 14-7 vdW potential function, to avoid the singularity problem at the end point:

$$U_{ij} = \lambda^n \varepsilon_{ij} \frac{1.07^7}{[\alpha(1-\lambda)^2 + (\rho + 0.07)^7]} \left(\frac{1.12}{\alpha(1-\lambda)^2 + \rho^7 + 0.12} - 2 \right) \tag{9.16}$$

where ε is the well depth and λ is the scaling factor. MD simulations are performed for each step along the decoupling pathways for 500 ps. Bennett Acceptance Ratio estimator (BAR) [44] allows computing free energy differences between the properly defined adjacent intermediate states based on configurations sampling

$$\Delta A(j)_{\lambda i \to \lambda i+1} = \ln \frac{\left\langle 1/\left[1+\exp((E_{\lambda i} - E_{\lambda i+1} + C)/RT)\right]\right\rangle_{\lambda i+1}}{\left\langle 1/\left[1+\exp((E_{\lambda i+1} - E_{\lambda i} + C)/RT)\right]\right\rangle_{\lambda i}} + C \tag{9.17}$$

where C is given by $C = \Delta A(j-1)_{\lambda i \to \lambda i+1}$, and *j* is the iteration index. Here, $E_{\lambda i}$ is the total energy of the system evaluated using the simulation snapshots at λ_i, with a dipole convergence of 10^{-6} D. ΔA is solved iteratively until the value of ($\Delta A(j) - \Delta A(j-1)$) is less than 0.01 kcal/mol. While a relatively looser induced dipole convergence (0.01 D) has been used for computational efficiency during the simulations, a tighter convergence of 10^{-6} D per atom is applied to re-analyze the saved snapshots and compute the results. The re-weighting can be incorporated rigorously into the BAR formula:

$$\Delta A(j)_{\lambda i \to \lambda i+1} = \ln \frac{\left\langle \left[1/(1+\exp((E_{\lambda i} - E_{\lambda i+1} + C)/RT))\right] \times \left[\exp((E'_{\lambda i+1} - E_{\lambda i+1})/RT)\right] \right\rangle_{\lambda' i+1}}{\left\langle \exp((E'_{\lambda i+1} - E_{\lambda i+1})/RT) \right\rangle_{\lambda' i+1}} + C \qquad (9.18)$$
$$\frac{\left\langle \left[1/(1+\exp((E_{\lambda i+1} - E_{\lambda i} + C)/RT))\right] \times \left[\exp((E'_{\lambda i} - E_{\lambda i})/RT)\right] \right\rangle_{\lambda' i}}{\left\langle \exp((E'_{\lambda i} - E_{\lambda i})/RT) \right\rangle_{\lambda' i}}$$

where $E'_{\lambda i}$ is the total energy of the system evaluated at λ_i using a dipole convergence of 0.01 D, while $E_{\lambda i}$ indicates the potential energy evaluated using full convergence. λ' indicates the ensemble obtained using the looser dipole convergence.

To verify the numerical reliability of the re-analysis method, we have used a tighter convergence of 10^{-6} D in both the MD simulation and the free energy analysis (Eq 9.17) and the resulting electrostatic decoupling free energy of the benzamidine-water system is -48.31 kcal/mol. With the re-analysis method based on Eq (9.18), we obtained -48.19 kcal/mol. The difference of 0.12 kcal/mol is well within the statistical error. The computational saving is almost a factor of three. With the re-analysis approach, the computational cost of AMOEBA is within an order of magnitude of that fixed-charged AMBER. It should be noted that we decoupled the electrostatic interaction by zeroing out both inter- and intra- molecular electrostatic interactions of the small molecule. The free energy contributions of intramolecular interaction cancel out between ligand-protein and ligand-water decoupling during the calculation of protein-ligand binding affinity. However, this contribution cannot be ignored in HFE calculations. The intramolecular interaction is therefore restored by gradually recharging the ligand in vacuum. A time step of 0.1 fs should be used.

Table 9.1. Comparison of hydration free energies of small molecules with four different basis sets. All units are kcal/mol. Statistical errors are from 0.04~0.22 kcal/mol [45].

	Ethylbenze	p-cresol	Isopropanol	Imidazole	Methylethyl sulfide	Acetic acid	Ethanol	MUE	RMSE
cc-pVTZ	-0.68	-6.5	-4.04	-9.07	-1.61	-4.57	-4.19	0.63	1.09
6-311g**	-0.53	-6.47	-4.34	-8.70	-1.53	-4.45	-4.51	0.61	1.15
6-311++g(2d,2p)	-0.70	-7.26	-5.58	-10.11	-1.87	-5.53	-5.67	0.64	0.77
aug-cc-pVTZ	-0.44	-7.05	-4.32	-9.72	-1.56	-5.48	-5.04	0.41	0.63
Experiment[a]	-0.70	-6.10, -6.60	-4.70, -4.80	-9.63	-1.50	-6.70	-4.90		

[a]Experimental values are reported in ref [46], except for imidazole taken from ref [47].

In Table 9.1, the HFEs of seven organic small molecules have been calculated using atomic multipoles obtained from different basis sets (6-311G**, 6-311++G(2d,2p), cc-pVTZ, and aug-cc-pVTZ) at the MP2 level. The overall comparisons show a basis-set "convergence" with the aug-cc-pVTZ set giving the best performance. Results from 6-311++G(2d,2p) are also comparable although it seems to produce too favorable hydration free energy systematically. The cc-pVTZ and 6-311G** basis sets give similar results with each other. The RMSD in the HFEs between these two basis sets is 0.23 kcal/mol. This study suggests that it is important to use basis sets with diffusion functions to derive electrostatic parameters in order to capture intermolecular interactions accurately. This is however not necessarily applicable to fixed-charge parameters which are also scaled empirically from their gas-phase value to account for "polarization" in condense-phase.

9.3.2 Ion solvation thermodynamics

Nearly half of all proteins contain metal ions, and ions play essential roles in many fundamental biological functions. Metal ions can act as signaling molecules, like Na^+, K^+ and Ca^{2+}, and can be involved in specific bindings with proteins as well, especially divalent ions such as Mg^{2+}, Zn^{2+}. Not only do these metal ions help modify and stabilize the protein structures, they also take part in catalytic processes. In enzymes, the ions can bind and orient the substrates through electrostatic interactions at the active sites, and thus control the enzyme catalytic reaction, of which the contribution is a key ingredient. In addition, Ion solvation thermodynamics are of great interest and significance, as the interplay between the ion-water and ion-protein interactions could provide the basis for unraveling the ion selection mystery. However, modeling these ions with classical models has always been a challenging task because of the high electron density and softness of the metal ions.

AMOEBA potential energy function has been very successful in modeling ions, from monovalent to divalent [14, 48-51]. For Ca^{2+}, Mg^{2+} and Zn^{2+}, we discover that it is necessary to modify the polarizable damping factor, which to some extent reflects the size of the charge distribution of these ions. The vdW parameters and damping factors are typically obtained by fitting to the *ab initio* binding energies of ion-water dimers in gas phase. The contribution of the polarization effect to ion-water interaction energy in AMOEBA model has been verified by comparing to QM decomposition [14, 51]. Using QM-derived AMOEBA polarizable ion models, we have been able to accurately predict the hydration free energies of these ions. It is worth noting that experimental hydration energy exists only for whole salt, not for single ions, for which the "experimental" values were extracted based on certain extra thermodynamic assumptions and varied significantly from source to source. As seen in Table 9.2, the HFE of Ca^{2+}, Mg^{2+} [50]

and Zn^{2+} [14] are as large as ~-400 kcal/mol, which are about 4-5 time larger than the monovalent ions [48, 49]. The calculated hydration free energies of the three divalent ions using AMOEBA polarizable model are all in excellent agreement with both the best "experimental data" available and quasi-chemical approximation.

Table 9.2. Hydration free energy of divalent metal ions in water[a] (in kcal/mol). Statistical errors of the calculations are from < 4.4 kcal/mol.

ion	Mg^{2+}	Ca^{2+}	Zn^{2+}
ΔA	-431.1	-354.9	-458.9
Experimental	-435.4[d]	-357.2[d]	-467.7[c]
quasi-chemical[b]	-435.2	-356.6	-460

[a]A 1 mol L^{-1} solution is chosen as the standard state; [b]Ref. [52]; [c]Ref. [53]; [d]Ref. [54]

In addition to the thermodynamic energetics, the structure of solvent environment can also be well predicted by MD simulations with AMOEBA polarizable ion models. Based on the radial distribution function (RDF), simulations show that the Zn^{2+} interacts with water molecules in a similar fashion as Mg^{2+}, only a little tighter at the first water shell. The first peak of Ca^{2+}-O RDF is at 2.41 Å, indicating a looser binding of water molecules to Ca^{2+} than the other two ions (2.07 Å for Mg^{2+} and 1.98 Å for Zn^{2+} respectively). The born radius, characterized by the first minimum in the ion-O RDF shows an order of $Ca^{2+}>Mg^{2+}\cong Zn^{2+}$.

9.3.3 Binding free energy of Trypsin and benzamidine analogs

The first attempt to predict the absolute binding affinity using AMOEBA polarizable force field is focused on the trypsin-benzamidine system [18]. Trypsin is a serine protease characterized by the serine residue in the active site. S1 pocket is a common target for inhibition because the aspartic acid residue in S1 site can provide strong electrostatic interactions with counter-charged substrates. In this study, double decoupling method [55] was utilized to calculate the absolute and relative free energy of benzamidine analogs binding to trypsin. The nonbonded interactions between ligand and its environment were gradually "disappeared" in both bulk water and protein, using the similar approach as mentioned in section 9.3.1, except that the intramolecular electrostatic interaction energies do not have to be restored due to the cancelation between the ligand-protein and ligand-water systems.

To prevent the ligand drifting away from the binding pocket as the nonbonded interactions are getting weaker, a harmonic potential was used to restrain the ligand to the protein pocket during the decoupling process [55, 56]. This artificial bias should finally be removed from the binding free energy obtained from decoupling with a correction:

$$\Delta A_{bind} = \Delta A_{wat}(L \to 0) - \Delta A^{r}_{pro}(L \to 0) - \Delta A^{r}(L) - RT \ln\left[C^{o} \left(\frac{2\pi RT}{k} \right)^{3/2} \right] \qquad (9.19)$$

in which the third term is the free energy change via exerting the restraint when the interactions between the protein and the ligand are intact. The last term was calculated as $-RTln(C^0V)$, where C^0 is the standard concentration and V is the sampling volume of the ligand under the restraint [55]. In addition, the restraint between ligand and protein at fully interacting state needs to be removed using alchemical perturbation [19].

Table 9.3 Absolute free energy of benzamidine binding to trypsin (in kcal/mol) computed using different force constants (in kcal/mol Å^{-2}) and soft-core coefficients. The standard errors are from 0.2-0.4 kcal/mol.

Restraint constant	*Softcore*	$\Delta A_{wat}(L\rightarrow 0)$		$\Delta A_{pro}(L\rightarrow 0)$		*Restraint correction*	ΔA_{calc}	ΔA_{exp}
		ΔA_{ele}	ΔA_{vdw}	ΔA_{ele}	ΔA_{vdw}			
20	0.5/4	1.27	-2.27	7.78	3.42	6.26	-6.72	-6.3[b],
20	0.7/5	1.27[e]	-2.42	7.78[e]	3.72	6.26	-7.27	-7.3[c],
40	0.7/5	1.27[e]	-2.35[a]	7.57	4.56	7.03	-7.28	-6.4[d]

[a] Averaged from the two ligand-water vdW decoupling free energies in the rows above.
[b] Ref. [57]; [c] Ref. [58]; [d] Ref.[59]; [e] The value is taken from the row above.

Table 9.3 shows the absolute free energy of benzamidine binding to trypsin computed with different force constants and soft-core coefficients. The calculated absolute binding affinities range from -6.7 to -7.3 Kcal/mol, in reasonable agreement with experiments [57-59]. Results from different restraint constants show that the restraint affects mostly the vdW decoupling. After taking the corrections into account, the vdW contributions in water and protein environments mostly cancel and thus we conclude that electrostatic force is the main driving force for binding.

To investigate the effect of polarization, free energies change was calculated by disabling dipole induction between benzamidine and its environment. It is interesting to notice that the free energy change due to the removal of the polarization between benzamidine and water is 4.49 kcal/mol and -22.37 kcal/mol between benzamidine and trypsin. The negative sign in the latter indicates that polarization is weakening benzamidine-trypsin interaction. This observation may be counterintuitive at first sight however is easily understood as a dielectric effect of the protein itself, which acts to screen the charge-charge interaction within proteins. Even implicit solvent model can capture this and various arguments exist in the literature that a large “effective” dielectric constant (e.g. 4 or 10) is more appropriate [1, 3, 60]. The significant contribution by polarization to protein-ligand binding is consistent with early findings that electrostatics is sensitive to local environment [61, 62]. A fixed-charge force field that neglects polarization effect within protein environment would potentially overestimate the binding free energy.

Using alchemical transformation, one can also obtain the relative binding affinities of other benzamidine like ligands. In Figure 9.1, ligands B through E were perturbed from

ligand A, and ligand D was transformed into ligand F. Compared with absolute binding free energy, the relative binding free energy is more likely to be predicted accurately due to systematic error cancellation. Table 9.4 shows the relative and absolute free energy of six benzamidine analogs binding to trypsin [18]. The calculated binding affinities using explicit-solvent alchemical method are in excellent agreement with experimental measurements.

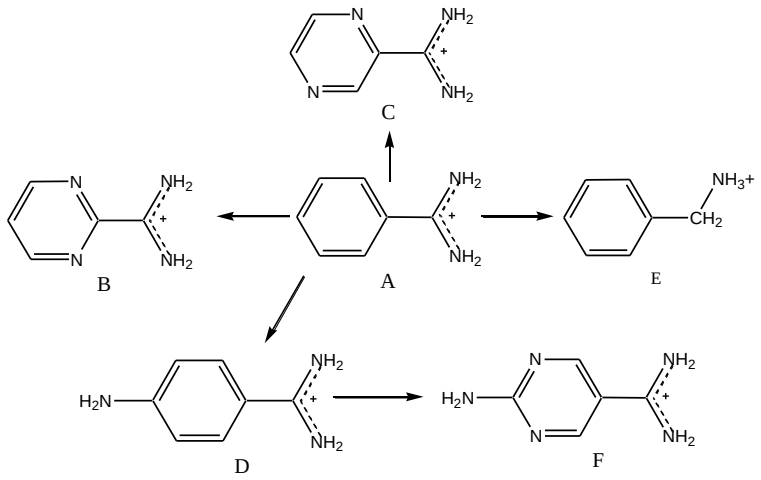


Figure 9.1 Chemical structure of trypsin ligands studied (A-F).

Table 9.4. Binding free energies from the explicit solvent alchemical simulations (section 9.3) and implicit methods (section 9.4).

	A	B	C	D	E	F
ΔA_{bind} (explicit-water alchemical)[a]	-6.7	-5.0	-4.9	-7.0	-5.4	-4.8
ΔA_{bind} (implicit GK alchemical)[b]	-6.7	-5.3	-5.5	-7.5	-5.3	-6.4
ΔA_{bind} (implicit MM-PMPB/SA)[c]	-8.2	-6.3	-2.7	-9.4	-6.6	-7.7
Experiment	-6.3[d,g], -6.4[e], -7.3[f]	-4.7[d]	-4.8[d]	-7.0[e], -7.2[g]	-3.8[h], -4.7[i]	-5.0[g]
Other computation	-6.4[d], -6.8[g]	-7.0[d]	-6.5[d]	-6.1[g]	-4.2[h], -2.4[i]	-4.7[g]

[a]Absolute binding free energy (BFE) calculated for ligand A; for ligand B~F, the absolute BFEs are listed for comparison purpose based on their relative BFEs computed after perturbed to ligand A during the simulations. [b] Same as above, except the average value of experimental BFE of ligand A is used as the reference point (-6.7kcal/mol). [c]Absolute binding free energies; [d]Ref. [63]. Calculation using PB/SA combined with QM/MM; [e]Ref [59]; [f]Ref [58]; [g]Ref [57]. Calculation using PB/SA; [h]Ref [64]. Non-Boltzmann thermodynamic integration (NBTI) MD simulations; [i]Ref [65]. Linear interaction energy (LIE).

9.4. Implicit solvent calculation using AMOEBA polarizable force field

An implicit description of solvent using a continuum dielectric model has always been of great interest towards the study of molecular recognition due to the lower computational cost. Recently, Schnieders *et al.* developed two continuum methods for solutes containing polarizable point multipoles [23, 24]. The first couples the polarizable multipole solute with Poisson Boltzmann continuum solvent, which we refer to as a "PMPB" method. The second approach extends the Generalized Born model for monopoles to the Generalized Kirkwood (GK) model for multipoles. Both the PMPB and

GK continuum methods have been compared with explicit solvent simulations for several small proteins [23, 24]. The solvent effects on the protein electrostatic moments produced by the two methods are in good agreement. Here we will introduce the two implicit solvent methods to investigate protein-ligand interactions for their computational efficiency and simplicity, i.e. the molecular mechanics-based polarizable multipole Poisson-Boltzmann surface area (MM-PMPB/SA) approach [19] and alchemical perturbation using implicit solvent GK [60].

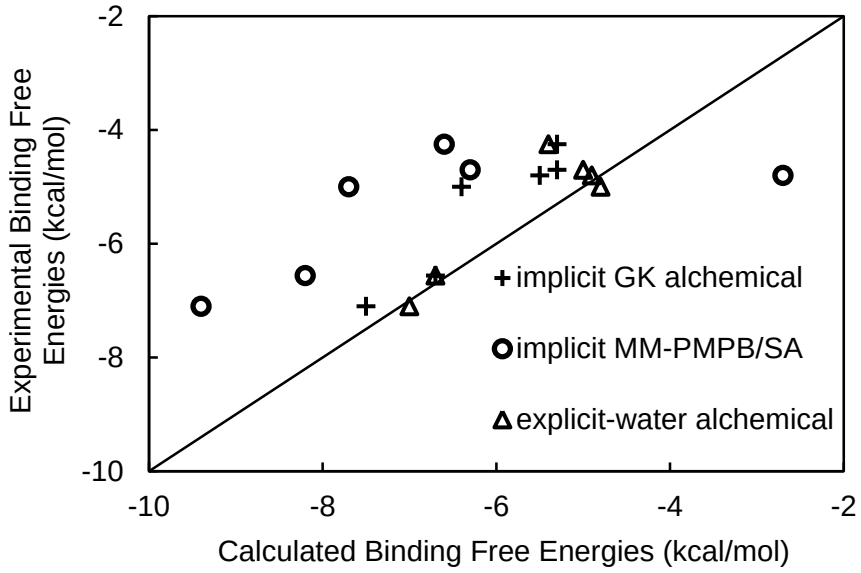


Figure 9.2 Comparison of calculated vs. experimental binding free energies. The experimental values from top to bottom are for ligand E, B, C, F, A and D.

Similar to the original MM-PB/SA [6, 66], the MM-PMPB/SA method represents the post-processing scheme to evaluate the absolute binding free energies by evaluating the binding energy of the snapshots from the MD simulation. In each snapshot, the free energy is calculated using the following equations:

$$\Delta A_{bind} = \Delta A_{gas} + \Delta A_{solv}^{PL} - \Delta A_{solv}^{P} - \Delta A_{solv}^{L} = \Delta U_{gas} - T\Delta S + \Delta A_{solv}^{PL} - \Delta A_{solv}^{P} - \Delta A_{solv}^{L} \quad (9.20)$$

$$\Delta U_{gas} \cong \Delta E_{gas} = \Delta E_{elec} + \Delta E_{vdw} \quad (9.21)$$

where ΔE_{gas} is the gas phase potential energy change upon ligand and protein binding, and ΔA_{solv}^{P}, ΔA_{solv}^{L} and ΔA_{solv}^{PL} are the solvation free energies of protein, ligand, and complex, respectively. Each solvation free energy is further decomposed into polar and nonpolar parts: the PMPB continuum electrostatic (polar) solvation and the nonpolar solvation approximated by solvent-accessible surface area [67].

$$\Delta A_{solv} = \Delta A_{polar} + \Delta A_{nonpolar} \quad (9.22)$$

The gas phase entropic changes ($T\Delta S$) upon binding were estimated by subtracting the entropies of the ligand-free protein and the ligand from that of the complex. For every molecule, the translational, rotational and vibrational entropies can be calculated based on normal mode analysis approach [68]. The protein atoms further than 8 Å away from the ligands were frozen in the calculation of vibration entropies. Note that the contribution due to conformational restriction requires additional treatment [69].

The calculated absolute binding free energies are compared with experimental data in Table 9.4 and Figure 9.2. Excluding ligand C (1,4-diazamidine), the calculated absolute binding free energies of all the other ligands are in good correlation with the experimental data (R=0.94). When ligand C is included, the correlation coefficient is 0.67. Given that the PMPB/SA model has not been extensively parameterized or tested, the results are quite promising. In the trypsin systems studied, the sum of gas phase binding energies and polar solvation energies determines the ranking of the binding affinities. In fact, the contributions of other energy terms, nonpolar solvation energy and entropy have little effects on the ranking order. While the nonpolar solvation contribution is favorable for all six ligands, their magnitudes only vary slightly, which may be a result of the oversimplified surface area term we adopted during the calculation. Although the entropy contribution (unfavorably) to binding is significant, it is almost invariant across the six ligands. Therefore, for the purpose of ranking ligand affinity, entropy and nonpolar solvation energy may be ignored. However, this may not be true especially when ligands differ significantly in size, shape or structure.

Continuum methods are expected to have limitations when specific water molecules are critically interacting with solutes or deep in the binding pocket [6, 70]. Even with the sophisticated electrostatic model in the PMPB method, the treatment of nonpolar solvation and entropy is challenging and can be computationally expensive [6, 71]. On the other hand, due to the elimination of intermediate perturbation steps, the PMPB method is orders of magnitude more efficient than the explicit solvent simulations.

In addition, we have also evaluated an implicit-solvent based alchemical perturbation approach. The electrostatic solvation energy is calculated using the polarizable GK model and nonpolar solvation contribution have both cavitation and dispersion terms [60]. The relative binding energy is computed as in the explicit solvent (section 9.3.3) approach by mutating one ligand into another in 10 steps. The difference is that here implicit solvent is used and the calculation is performed using TINKER program [36]. The same soft-core buffered 14-7 vdW potential is used to prevent energetic instabilities as those annihilated atoms could be penetrated by other atoms. Each simulation at the intermediate perturbation states is run for 150 ps and the time step is 1.5 fs.

Table 9.4 and Figure 9.2 compare the binding free energies calculated from MM-PMPB/SA, GK alchemical simulations with those from experiment and explicit-water

alchemical simulations. Note that we choose the calculated explicit-water alchemical simulation results of ligand A as the reference point (-6.7 kcal/mol) for GK results, as only the relative binding free energy has been computed. Although the GK based alchemical approach tends to predict the free energy change slightly more favorable, it is almost as accurate as the explicit water alchemical simulation. The results suggest that including intermediate states may be the best way to account for the entropy contribution although additional verification on a broad range of protein-ligand complexes is necessary. Given the computational efficiency and effectiveness, implicit-solvent based alchemical approach could be of practical use for virtual screening, particularly optimizing small molecules against macromolecular targets.

9.5. Conclusions and future directions

In this chapter, we have discussed recent progress in the development and application of AMOEBA polarizable force field to compute protein-ligand binding affinity. In the past AMOEBA has been successfully applied to modeling water [21, 72], mono- and divalent ion solvation [14, 48, 49], organic molecules and peptides [22], small molecule hydration free energies [45], trypsin-ligand binding prediction [18, 19, 60] and computational X-ray crystallography [73] with promising results. We hope that we have demonstrated that prediction of protein-ligand binding affinity can benefit significantly from rigorous electrostatic treatment and incorporation of electronic polarization effect. In addition to the quality of force field, configurational entropy calculation remains an enormous challenge, especially for protein-ligand binding that involves a large number of conformational degrees of freedom. Explicit but perhaps smart sampling of the configurational space, such as the implicit-solvent based alchemical approach we have discussed, seems promising.

Continuing effort has been made to refine the AMOEBA force field and its parameters. Electrostatic parameters are found to have better transferability when the permanent multipoles are derived using the original DMA method [34] at the level of MP2/6-311G** basis set and then optimized to reproduce the QM electrostatic potential from a higher level basis set (e.g. MP2/aug-cc-pVTZ). A new and more comprehensive AMOEBA protein force field is under development to enhance the parameter transferability and improve the accuracy. Parameterization for small molecules to score protein-ligand binding involves a substantial amount of work. A public-available software tool is available to automate the parameterization of arbitrary molecules (http://water.bme.utexas.edu/wiki/index.php/Software:Poltype) [74], in a way consistent with the existing AMOEBA protein and small molecule force fields.

The current computational efficiency of AMOEBA model is slower than that of a fixed point charge model by a factor of 8-10. Methods to accelerate the dipole convergence and to improve the parallelization are being actively pursued. To date, AMOEBA force

field has been implemented with a number of software programs including AMBER [37], OpenMM [75], Force Field X (http://kenai.com/projects/ffx) and TINKER [36]. We expect that AMOEBA, and other polarizable force fields, will see increasing applications in protein-ligand binding prediction.

References:

1. Gilson, M. K. and Zhou, H. X., *Calculation of protein-ligand binding affinities.* Annu. Rev. Biophys. Biomol. Struct., 2007. **36**: p. 21-42.
2. Gohlke, H. and Klebe, G., *Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors.* Angewandte Chemie-International Edition, 2002. **41**(15): p. 2645-2676.
3. Hou, T., Wang, J., Li, Y., and 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.* Journal of Chemical Information and Modeling, 2011. **51**(1): p. 69–82.
4. Jorgensen, W. L., *The many roles of computation in drug discovery.* Sci, 2004. **303**(5665): p. 1813-1818.
5. Knight, J. L. and Brooks, C. L., *lambda-Dynamics Free Energy Simulation Methods.* J. Comput. Chem., 2009. **30**(11): p. 1692-1700.
6. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models.* Acc. Chem. Res., 2000. **33**(12): p. 889-897.
7. Mobley, D. L. and Dill, K. A., *Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get".* Structure, 2009. **17**(4): p. 489-498.
8. Woo, H. J. and Roux, B., *Calculation of absolute protein-ligand binding free energy from computer simulations.* Proc. Natl. Acad. Sci. U. S. A., 2005. **102**(19): p. 6825-6830.
9. Mackerell, A. D., *Empirical force fields for biological macromolecules: Overview and issues.* J. Comput. Chem., 2004. **25**(13): p. 1584-1604.
10. Ponder, J. W. and Case, D. A., *Force fields for protein simulations.* Adv. Protein Chem., 2003. **66**: p. 27-85.
11. Cieplak, P., Dupradeau, F. Y., Duan, Y., and Wang, J. M., *Polarization effects in molecular mechanical force fields.* Journal of Physics-Condensed Matter, 2009. **21**(33).
12. Lopes, P. E. M., Roux, B., and MacKerell, A. D., *Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications.* Theor. Chem. Acc., 2009. **124**(1-2): p. 11-28.

13. Stone, A. J., *The Theory of Intermolecular Forces*. International Series of Monographs on Chemistry. 1996, Oxford: Oxford University Press.
14. Wu, J. C., Piquemal, J.-P., Chaudret, R., Reinhardt, P., and Ren, P., *Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field.* Journal of Chemical Theory and Computation, 2010. **6**(7): p. 2059-2070.
15. Bernardo, D. N., Ding, Y. B., Kroghjespersen, K., and Levy, R. M., *An Anisotropic Polarizable Water Model - Incorporation of All-Atom Polarizabilities into Molecular Mechanics Force-Fields.* J. Phys. Chem., 1994. **98**(15): p. 4180-4187.
16. Patel, S. and Brooks, C. L., *CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations.* J. Comput. Chem., 2004. **25**(1): p. 1-15.
17. Lopes, P. E. M., Lamoureux, G., Roux, B., and MacKerell, A. D., *Polarizable empirical force field for aromatic compounds based on the classical drude oscillator.* J. Phys. Chem. B, 2007. **111**(11): p. 2873-2885.
18. Jiao, D., Golubkov, P. A., Darden, T. A., and Ren, P., *Calculation of protein-ligand binding free energy by using a polarizable potential.* P Natl Acad Sci USA, 2008. **105**(17): p. 6290-6295
19. Jiao, D., Zhang, J. J., Duke, R. E., Li, G. H., Schnieders, M. J., and Ren, P. Y., *Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential.* J. Comput. Chem., 2009. **30**(11): p. 1701-1711.
20. Ponder, J. W., Wu, C. J., Ren, P. Y., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio, R. A., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., and Head-Gordon, T., *Current Status of the AMOEBA Polarizable Force Field.* J. Phys. Chem. B, 2010. **114**(8): p. 2549-2564.
21. Ren, P. Y. and Ponder, J. W., *Polarizable atomic multipole water model for molecular mechanics simulation.* J. Phys. Chem. B, 2003. **107**(24): p. 5933-5947.
22. Ren, P. Y., Ponder, J. W., and Mz, *Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations.* J. Comput. Chem., 2002. **23**(16): p. 1497-1506.
23. Schnieders, M. J., Baker, N. A., Ren, P. Y., and Ponder, J. W., *Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum.* JChPh, 2007. **126**(12): p. 124114.
24. Schnieders, M. J. and Ponder, J. W., *Polarizable atomic multipole solutes in a generalized Kirkwood continuum.* Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2082-2097.
25. Allinger, N. L., Yuh, Y. H., and Lii, J. H., *Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .1.* J. Am. Chem. Soc., 1989. **111**(23): p. 8551-8566.
26. Wilson, E. B., Decius, J. C., and Cross, P. C., *Molecular Vibrations. McGraw-Hill: New York.* 1955.

27. Bell, R. P., Faraday Society, 1945. **41**: p. 293.
28. Mannfors, B., Sundius, T., Palmo, K., Pietila, L. O., and Krimm, S., *Spectroscopically determined force fields for macromolecules. Part 3. Alkene chains.* J. Mol. Struct., 2000. **521**: p. 49-75.
29. MacKerell, A. D., Feig, M., and Brooks, C. L., *Improved treatment of the protein backbone in empirical force fields.* J. Am. Chem. Soc., 2004. **126**(3): p. 698-699.
30. DiStasio, R. A., Jung, Y. S., and Head-Gordon, M., *A resolution-of-the-identity implementation of the local triatomics-in-molecules model for second-order Moller-Plesset perturbation theory with application to alanine tetrapeptide conformational energies.* Journal of Chemical Theory and Computation, 2005. **1**(5): p. 862-876.
31. Halgren, T. A., *Representation of Vanderwaals (Vdw) Interactions in Molecular Mechanics Force-Fields - Potential Form, Combination Rules, and Vdw Parameters.* J. Am. Chem. Soc., 1992. **114**(20): p. 7827-7843.
32. Kong, Y., *Multipole Electrostatic Methods for Protein Modeling with Reaction Field Treatment*. 1997, Washington University St. Louis Dissertation.
33. Thole, B. T., *Molecular Polarizabilities Calculated with a Modified Dipole Interaction.* ChPh, 1981. **59**: p. 341-350.
34. Stone, A. J. and Alderton, M., *Distributed multipole analysis - Methods and applications (Reprinted from Molecular Physics, vol 56, pg 1047-1064, 1985).* MolPh, 2002. **100**(1): p. 221-233.
35. Sagui, C., Pedersen, L. G., and Darden, T. A., *Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations.* JChPh, 2004. **120**(1): p. 73-87.
36. Ponder, J. W., *TINKER: Software Tools for Molecular Design*. 2006, Washington University School of Medicine Saint Louis.
37. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J., *The Amber biomolecular simulation programs.* J. Comput. Chem., 2005. **26**(16): p. 1668-1688.
38. Frenkel, D. and Ladd, A. J. C., *New Monte-Carlo Method to Compute the Free-Energy of Arbitrary Solids - Application to the Fcc and Hcp Phases of Hard-Spheres.* JChPh, 1984. **81**(7): p. 3188-3193.
39. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L., *Comparison of Simple Potential Functions for Simulating Liquid Water.* JChPh, 1983. **79**(2): p. 926-935.
40. Shirts, M. R., Pitera, J. W., Swope, W. C., and Pande, V. S., *Extremely precise free energy calculations of amino acid side chain analogs: Comparison of*

common molecular mechanics force fields for proteins. JChPh, 2003. **119**(11): p. 5740-5761.

41. Mobley, D. L., Bayly, C. I., Cooper, M. D., Shirts, M. R., and Dill, K. A., *Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations.* Journal of Chemical Theory and Computation, 2009. **5**(2): p. 350-358.
42. Mobley, D. L., Dumont, E., Chodera, J. D., and 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**(9): p. 2242-2254.
43. Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R., and Vangunsteren, W. F., *Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations.* CPL, 1994. **222**(6): p. 529-539.
44. Bennett, C. H., *Efficient Estimation of Free-Energy Differences from Monte-Carlo Data.* JCoPh, 1976. **22**(2): p. 245-268.
45. Shi, Y., Wu, C. J., Ponder, J. W., and Ren, P. Y., *Multipole electrostatics in hydration free energy calculations.* J. Comp. Chem., 2010. (in press).
46. Cabani, S., Gianni, P., Mollica, V., and Lepori, L., *Group Contribution to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution.* JSCh, 1981. **10**: p. 563-595.
47. Wolfenden, R., Liang, Y., M, M., and R, W., *Cooperativity and Anticooperativity in Solvation by Water: Imidazoles, Quinones, Nitrophenols, Nitrophenolate, and Nitrothiophenolate Ions.* J. Am. Chem. Soc., 1987. **109**: p. 463–466.
48. Grossfield, A., Ren, P. Y., and Ponder, J. W., *Single ion solvation thermodynamics from simulations.* Biophys. J., 2003. **84**(2): p. 94A-94A.
49. Grossfield, A., Ren, P. Y., and Ponder, J. W., *Ion solvation thermodynamics from simulation with a polarizable force field.* J. Am. Chem. Soc., 2003. **125**(50): p. 15671-15682.
50. Jiao, D., King, C., Grossfield, A., Darden, T. A., and Ren, P. Y., *Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential.* J. Phys. Chem. B, 2006. **110**(37): p. 18553-18559.
51. Piquemal, J. P., Perera, L., Cisneros, G. A., Ren, P. Y., Pedersen, L. G., and Darden, T. A., *Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure.* JChPh, 2006. **125**(5): p. 054511.
52. Asthagiri, D., Pratt, L. R., Paulaitis, M. E., and Rempe, S. B., *Hydration structure and free energy of biomolecularly specific aqueous dications, including Zn^{2+} and first transition row metals.* J. Am. Chem. Soc., 2004. **126**(4): p. 1285-1289.
53. Marcus, Y., *A Simple Empirical-Model Describing the Thermodynamics of Hydration of Ions of Widely Varying Charges, Sizes, and Shapes.* Biophys. Chem., 1994. **51**(2-3): p. 111-127.

54. Schmid, R., Miah, A. M., and Sapunov, V. N., *A new table of the thermodynamic quantities of ionic hydration: values and some applications (enthalpy-entropy compensation and Born radii).* PCCP, 2000. **2**(1): p. 97-102.
55. Hamelberg, D. and McCammon, J. A., *Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method.* J. Am. Chem. Soc., 2004. **126**(24): p. 7683-7689.
56. Boresch, S., Tettinger, F., Leitgeb, M., and Karplus, M., *Absolute binding free energies: A quantitative approach for their calculation.* J. Phys. Chem. B, 2003. **107**(35): p. 9535-9551.
57. Schwarzl, S. M., Tschopp, T. B., Smith, J. C., and Fischer, S., *Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction?* J. Comput. Chem., 2002. **23**(12): p. 1143-1149.
58. Katz, B. A., Elrod, K., Luong, C., Rice, M. J., Mackman, R. L., Sprengeler, P. A., Spencer, J., Hataye, J., Janc, J., Link, J., Litvak, J., Rai, R., Rice, K., Sideris, S., Verner, E., and Young, W., *A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site.* J. Mol. Biol., 2001. **307**(5): p. 1451-1486.
59. Talhout, R. and Engberts, J. B. F. N., *Thermodynamic analysis of binding of p-substituted benzamidines to trypsin.* Eur. J. Biochem., 2001. **268**(6): p. 1554-1560.
60. Yang, T., Wu, J. C., Yan, C., Luo, R., Gonzales, M. B., Dalby, K. N., and Ren, P., *Virtual Screening Using Molecular Simulations.* Proteins: Structure, Function, and Bioinformatics, (submitted).
61. Lee, F. S., Chu, Z. T., Bolger, M. B., and Warshel, A., *Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603.* Protein Eng., 1992. **5**(3): p. 215-228.
62. Hensen, C., Hermann, J. C., Nam, K. H., Ma, S. H., Gao, J. L., and Holtje, H. D., *A combined QM/MM approach to protein-ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors.* J. Med. Chem., 2004. **47**(27): p. 6673-6680.
63. Grater, F., Schwarzl, S. M., Dejaegere, A., Fischer, S., and Smith, J. C., *Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics.* J. Phys. Chem. B, 2005. **109**(20): p. 10474-10483.
64. Ota, N., Stroupe, C., Ferreira-da-Silva, J. M. S., Shah, S. A., Mares-Guia, M., and Brunger, A. T., *Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: Relative free energy of binding of trypsin to*

benzamidine and benzylamine. Proteins-Structure Function and Genetics, 1999. **37**(4): p. 641-653.
65. Leiros, H. K. S., Brandsdal, B. O., Andersen, O. A., Os, V., Leiros, I., Helland, R., Otlewski, J., Willassen, N. P., and Smalas, A. O., *Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements.* Protein Sci., 2004. **13**(4): p. 1056-1070.
66. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices.* J. Am. Chem. Soc., 1998. **120**(37): p. 9401-9409.
67. Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C., *The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii.* JPCA, 1997. **101**(16): p. 3005-3014.
68. Tidor, B. and Karplus, M., *The Contribution of Vibrational Entropy to Molecular Association - the Dimerization of Insulin.* J. Mol. Biol., 1994. **238**(3): p. 405-414.
69. Gao, C., Park, M. S., and Stern, H. A., *Accounting for ligand conformational restriction in calculations of protein-ligand binding affinities.* Biophys. J., 2010. **98**(5): p. 901-910.
70. Resat, H., Marrone, T. J., and McCammon, J. A., *Enzyme-inhibitor association thermodynamics: Explicit and continuum solvent studies.* Biophys. J., 1997. **72**(2): p. 522-532.
71. Chen, J. H., Brooks, C. L., and Khandogin, J., *Recent advances in implicit solvent-based methods for biomolecular simulations.* Curr. Opin. Struct. Biol., 2008. **18**(2): p. 140-148.
72. Ren, P. Y. and Ponder, J. W., *Temperature and pressure dependence of the AMOEBA water model.* J. Phys. Chem. B, 2004. **108**(35): p. 13427-13437.
73. Schnieders, M. J., Fenn, T. D., Pande, V. S., and Brunger, A. T., *Polarizable atomic multipole X-ray refinement: application to peptide crystals.* Acta Crystallographica Section D-Biological Crystallography, 2009. **65**: p. 952-965.
74. Wu, J. C. and Ren, P. Y., *http://water.bme.utexas.edu/wiki/index.php/Software:Poltype.* (25 January 2011).
75. Eastman, P. and Pande, V. S., *OpenMM: A Hardware-Independent Framework for Molecular Simulations.* Computing in Science & Engineering, 2010. **12**(4): p. 34-39.