The Dissertation Committee for Dian Jiao Certifies that this is the approved version of the following dissertation:

ION MODELING AND LIGAND-PROTEIN BINDING CALCULATION WITH A POLARIZABLE FORCE FIELD

Committee:

Pengyu Ren, Supervisor

Ron Elber

Wolfgang Frey

Christine Schmidt

Muhammad Zaman

ION MODELING AND LIGAND-PROTEIN BINDING CALCULATION WITH A POLARIZABLE FORCE FIELD

by

Dian Jiao B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2009

Dedication

For my wife, Lili.

I could not have done this without your endless love and support.

EPIGRAPH

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

Laplace's demon

Acknowledgements

First and foremost, I would like to thank my mentor, Dr. Pengyu Ren for countless opportunities to work on good projects. The depth of his patience and wisdom in life is unfathomable. He certainly taught me much more than science. I owe my sincere thanks to my willingly served dissertation committee members, Dr. Ron Elber, Dr. Wolfgang Frey, Dr. Christine Schmidt and Dr. Muhammad Zaman for their guidance in all my research endeavors. I am grateful to former and current members of the Ren group for their assistance and advices. Not only did I earn esteemed colleagues but also good friends. I would also like to express my gratitude to our collaborators, Thomas A. Darden and Robert E. Duke, at National Institute of Environmental Health Sciences. They provided us with AMOEBA-implemented AMBER and PMEMD packages. Thanks to the National Institutes of Health and Welch Foundation for the financial support. Besides, I would like to acknowledge all the faculties and staffs in the Department of Biomedical Engineering for help. Very special thanks to my wife Lili for her love and understanding. Lastly, I would like to say thanks to my family and friends back in China who always have trust and faith in me.

ION MODELING AND LIGAND-PROTEIN BINDING CALCULATION WITH A POLARIZABLE FORCE FIELD

Publication No._____

Dian Jiao, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Pengyu Ren

Specific recognition of ligands including metal ions by proteins is the key of many crucial biological functions and systems. Accurate prediction of the binding strength not only sheds light on the mechanism of the molecular recognition but also provides the most important prerequisite of drug discovery. Computational modeling of molecular binding has gained a great deal of attentions in the last few decades since the advancement of computer power and availability of high-resolution crystal structures. However there still exist two major challenges in the field of molecular modeling, i.e. sampling issue and accuracy of the models. In this work, I have dedicated to tackling these two problems with a noval polarizable force field which is believed to produce more accurate description of molecular interactions than classic non-polarizable force fields. We first developed the model for divalent cations Mg^{2+} and Ca^{2+}, deriving the

parameters from quantum mechanics. To understand the hydration thermodynamics of these ions we have performed molecular dynamics simulations using our AMOEBA force field. Both the water structures around ions and the solvation free energies were in great accordance with experiment data. We have also simulated and calculated the binding free energies of a series of benzamidine-like inhibitors to trypsin using explicit solvent approach by free energy perturbation. The calculated binding free energies are well within the accuracy of experimental measurement and the direction of change is predicted correctly in all cases. Finally, we computed the hydration free energies of a few organic molecules and automated the calculation procedure.

Table of Contents

Acknowledgements i

Table of Figures vii

Table of Tables ix

List of Abbreviations x

1 Introduction 1

1.1 Computational Chemistry and Molecular Modeling 1

1.2 Molecular Modeling and Drug Discovery 2

1.3 Molecular Modeling methods 3

1.4 Molecular Mechanics Force Fields 6

1.5 Computational Simulation 12

1.6 Overview of Thesis Work 16

2 Modeling Divalent Metal ions 17

2.1 Introduction 17

2.2 Method 19

2.3 Results and Discussion 24

2.4 Conclusion 36

3 Calculation of Absolute Protein-Ligand Binding Free Energy with a Polarizable Force Field 38

3.1 Introduction 38

3.2 Method 43

3.3 Results and discussion 53

3.4 Conclusion 68

4 Calculation of relative ligand-protein binding free energies 71

4.1 Introduction 71

4.2 Method 74

4.3 Results and Discussion 84

4.4 Conclusion 96

5 Hydration Free Energy of Small Organic Molecules 98

5.1 Introduction....98

5.2 Method 100

5.3 Results and Discussion 106

5.4 Conclusion 109

Appendix A. Tutorial on ligand parameterization in 110

AMOEBA force field....110

A.1 Build the molecule 110

A.2 Optimize the structure....111

A.3 Single point calculation....112

A.4 Multipole calculation 113

A.5 Average multipoles 116

A.6 Refit the electrostatics....117

A.7 Add other parameters....118

A.8 Final check 119

Appendix B. Structures and parameters of all the ligands....120

B.1 Ligand A: Benzamidine 120

B.2 Ligand B: 1,3-diazamidine....123

B.3 Ligand C: 1,4-diazamidine....125

B.4 Ligand D: 4-amino-benzamidine 129

B.5 Ligand E: Benzylamine....132

B.6 F. 4-amino-diazamidine 135

Appendix C. Softcore modification of vdW in AMBER....140

C.1 Introduction 140

C.2 Modification details....141

References.....152

VITA.....163

Table of Figures

Figure 1. Representation of force field models and potentials. 7

Figure 2. Ion-water dimer binding energy in gas phase..... 25

Figure 3. Radial distribution functions of ion and oxygen atom in water. 29

Figure 4. Comparison of RDFs of mono and divalent cations in water. 31

Figure 5. Oxygen-oxygen radial distribution functions of water molecules in the first solvation shell.. 32

Figure 6. The O-ion-O angle distribution in the first solvation shell..... 33

Figure 7. The dipole moment of water molecules in the first solvation shell..... 35

Figure 8. Structure of benzamidine-bound trypsin 43

Figure 9. Thermodynamic cycle of absolute binding free energy calculation..... 45

Figure 10. The restraint between the benzamidine and the trypsin 50

Figure 11. Thermodynamic cycle of double decoupling with restraint. 51

Figure 12. Electrostatic and van der Waals decoupling free energiesof ligand-water system. 55

Figure 13. Electrostatic and van der Waals decoupling free energies of ligand-protein system 57

Figure 14. Polarization effect upon binding. 63

Figure 15. Electrostatic free energy change in ligand-bound trypsin with difference components being turned off. 66

Figure 16. Electrostatic free energy of ligand-protein with zeroing out polarization or quadrupole of the outer region atoms. 67
Figure 17. Chemical structures of trypsin ligands studied..... 75
Figure 18. Thermodynamic scheme to calculate relative binding free energy 78
Figure 19. Superposition of hybrid ligand A/E and benzamidine in trypsin crystal structure..... 81
Figure 20. Intermolecular hydrogen bonding structures between ligands and trypsin at the binding site..... 83
Figure 21. Decomposition of binding free energies..... 87
Figure 22. Correlation between dipole/polarizability of the ligands and binding free energy..... 91
Figure 23. Evolution of the hydrogen bond distances between the ligand and the surroundings..... 94
Figure 24 Structures of NMe-Formamide and n-butane..... 100
Figure 25. Thermodynamic cycle of hydration free energy calculation 101
Figure 26. Flow chart of hydration free energy calculation..... 105

Table of Tables

Table 1. Parameters for ions.. 26

Table 2. Solvation free energy of calcium and magnesium ion in water..... 26

Table 3. The effective sizes of ions as indicated by Born theory and RDF..... 28

Table 4 Absolute free energy of benzamidine binding to trypsin..... 58

Table 5. Absolute solvation free energies of benzamidine in water and trypsin. 59

Table 6. Relative and absolute binding free energies 85

Table 7. Molecular dipole moments. 88

Table 8. Hydration free energies of NMA and butane..... 106

Table 9. Comparison of hydration free energy with different perturbation paths 107

Table 10. Van der Waals long-range correction with different cutoff..... 108

List of Abbreviations

MM	Molecular Mechanics
QM	Quantum Mechanics
MD	Molecular Dynamics
MC	Monte Carlo
BD	Brownian Dynamics
HF	Hartree–Fock
DFT	Density Functional Theory
AIMD	Ab Initio Molecular Dynamics
CPMD	Car-Parrinello Molecular Dynamics
DPD	Dissipative Particle Dynamics
FEM	Finite element method
FEP	Free Energy Perturbation
BAR	Bennett Acceptance Ratio
BSSE	Basis-set superimposition error
RDF	Radial Distribution Function
DDM	Double Decoupling Method
PMF	Potential of Mean Force
TI	Thermodynamic Integration
LRA	Linear Response Analysis
LIE	Linear Interaction Energy
PDLD	Protein Dipoles Langevin Dipole
PFF	Polarized Force Field

WHAM	Weighted-Histogram Analysis Method
REM	Replica-Exchange Method
REST	Replica Exchange with Solute Tempering
PREMD	Partial Replica Exchange Molecular Dynamics
LREMD	Local Replica Exchange Molecular Dynamics
GDMA	Graphic Distributed Multipole Analysis
PME	Particle Mesh Ewald
VDW	Van Der Waals
RMS	Root Mean Square
QSAR	Quantitative Structure-Activity Relationships
QSPR	Quantitative Structure-Property Relationships
NMA	NMe-Formamide

1 Introduction

1.1 Computational Chemistry and Molecular Modeling

Chemistry is the science that deals with the construction, transformation and properties of molecules. Theoretical chemistry is the description of chemistry with mathematical methods along with fundamental laws of physics (Jansen, 2006). Computational chemists have devised the mathematical equations and algorithms to quantitatively describe physical and chemical phenomena, for example, energy, structures, reactivity etc. These algorithms are then programmed in computer languages. Very few perspectives of chemistry can be computed exactly, but almost every aspect of chemistry has been computed qualitatively or quantitatively. Computational chemistry techniques such as *ab initio* quantum mechanics, mechanical mechanics, and simulation approaches like molecular dynamics and Monte Carlo, have been powerful tools for theoretical chemists to study molecular structure and function through model building and simulation.

The emergence of biomolecular modeling dates back to 1960s. However, molecular modeling did not develop rapidly until triggered by the advent of supercomputers in the early 1980s. In the last few decades, advances have been driven by the improvements in NMR and X-ray crstallization, increasing computer power in hardware and software, and development of force fields and algorithms. Molecular modeling has become one of the hottest areas in many sciences and a widely used tool for research as an indispensible partner to experiment.

The use of molecular modeling is not restricted to researchers who are solely interested in fundamental aspects, such as molecular structure, kinetics, reaction mechanisms, and thermodynamics. Molecular modeling has also become a useful weapon in other scientific disciplines, such as materials science (Jörg-Rüdiger Hill, 2005), environmental sciences (James R. Rabinowitz, 2008) and life sciences. Currently, the largest discipline that benefits from computational chemistry and molecular modeling is probably pharmaceutical sciences.

1.2 Molecular Modeling and Drug Discovery

The discovery and development of a new chemical compound is an arduous and costly process. Statistics shows that, for every drug that is approved for medical use, up to 10,000 compounds are synthesized and tested; up to 100 compounds are assessed for safety; and up to 10 compounds are tested clinically in humans (Cohen, 1996). Even if a drug is approved for marketing, it does not mean success is garanteed; many drugs fail because they are not sufficiently efficacious practically or undesirable side effects are found from studies. Hundreds of millions of dollars are invested in fundamental research and clinical studies which hopefully lead to approval from the FDA. Traditionally, drugs have been "discovered" predominately through random or targeted screening, followed by distinct structural changes in the molecule to optimize the properties responsible for the desired activity.

The increasing power of computers and the fast visualization of 3-dimensional structures have allowed the emergence of sophisticated computer programs specifically

designed to help medicinal chemists in drug design. The combined application of molecular graphics, computational modeling, as well as chemical and biological informatics promise to fulfill a long-coveted goal of medicinal chemists: the prediction of biological activity prior to extensive laboratory synthesis and biological testing. Computational modeling based on biological information can be used to extend the limits of our understanding, thereby increasing the accuracy of prediction. Molecular modeling methods are still in their infancy, but they have already had a significant influence on drug discovery and development. Incorporation of molecular modeling has the potential to save millions of dollars based on increased efficiency (DiMasi, Hansen, & Grabowski, 2003).

The role of molecular modeling in drug discovery is to predict if a chemical molecule can bind to a target molecule and if so how tightly. The conformational changes of the substrate and receptor upon binding together with binding affinity and other thermodynamic properties can be calculated by molecular mechanics methods. *Ab initio* quantum chemistry, semi-empirical methods or experimental measurement usually offer optimized parameters for the MM calculations.

1.3 MOLECULAR MODELING METHODS

1.3.1 Quantum Mechanics

Molecular modeling methods range from highly accurate to very approximate; highly accurate methods are typically feasible only for small systems. Although *ab initio* quantum mechanics (QM) came along many years before the invention of the first

computer, many of the methods in common use today for molecular modeling are based on quantum mechanics theories. The development of computational chemistry techniques that are implemented on a computer allows quantum mechanics now to be used to perform calculations on molecular systems of practical interest.

QM methods are based on the solution of the Schrödinger equation (Levine, 1991) which describes the motions of the electrons and nuclei in a molecular system from first principle.

$$\hat{H}\psi_n = E_n\psi_n \qquad (1)$$

where the Hamiltonian operator $\hat{H}$ is the sum of kinetic energy and potential energy. E_n and ψ_n are quantum states (eigenvalues) and eigenfunctions respectively. In the Born-Oppenheimer approximation to the Schrödinger Equation, the motions of the molecule are separated into electrons and nuclei. There are two basic classes of QM techniques which rely on the Born-Oppenheimer approximation, *ab initio* and semi-empirical.

The term *ab initio* indicates that the calculation is from first principles and that no empirical data is used. Molecular orbitals are approximated by a summation of atomic orbitals. These are defined for a certain basis set, usually Gaussian functions. The most fundamental theories of *ab initio* is the Hartree–Fock (HF) theory, in which the correlation between electrons is not taken into account. The distribution of multi-electron system is described as the linear combination of single-electron distributions. Although Hartree-Fock calculations are faster than other QM methods due to this approximation, they inevitably overestimate the energies. The accuracy of HF calculation is dependent

on the number and quality of the basis sets to a great extent. By including correlation effects, for example, Møller–Plesset perturbation theory can improve the accuracy of the calculations (Cramer, 2002).

Based on the Hohenberg-Kohn model (P. Hohenberg, 1964), Density Functional Theory (DFT) is another QM method. Without calculating the multi-electron wavefunctions, DFT relates the electronic energy to the total electron density. DFT methods in general offer a good combination of accuracy and computational requirements especially for large systems, and they are computationally more efficient than *ab initio* methods (Yang & Lee, 1995).

The most effective way to reduce the computational cost is by neglecting some of the electron integrals which take up majority of computational time by QM methods. Semiempirical QM methods only take into account the electrons in the outer shells which participate in valence bonding with other atoms. That is why semi-empirical method is parameterized to best agree with experimental values so as to correct the bias introduced by approximation. Although semiempirical QM methods are much faster than *ab initio* methods, the results can be unreliable because of the challenging task of a fine parametrization (Schlick, 2002).

1.3.2 Molecular Mechanics

Quantum mechanics deals with the electrons of each atom, so unavoidably it is computationally expensive. Many systems we would like to investigate in biological

applications are way too large for QM. Molecular mechanics is a more popular tool for handling systems with significant numbers of atoms.

Molecular mechanics is also based on Born-Oppenheimer approximation without which it would be impossible to represent the energy as a function of the nuclear coordinates at all (Leach, 2001). As opposed to QM, molecular mechanics is built upon a simple model of the interactions (e.g. Hooke's law) with contributions from processes such as the stretching of bonds, the bending of angles and the rotations about bonds. Transferability is a crucial element of a MM model, for it enables a set of parameters developed and tested on a relatively small number of cases to be applied to a much wider range of problems. I will discuss the details about MM models (force fields) in the section 1.4.

1.4 MOLECULAR MECHANICS FORCE FIELDS

1.4.1 Classical Force Field

At the heart of molecular mechanics are the force fields, i.e. potential functions along with parameters to describe the interactions among atoms. In a force field, a molecule is represented as a mechanical system in which balls (atoms) are linked by springs (bonds), with atoms having different sizes and "softness" and bonds different lengths and "stiffness" (Schlick, 2002).

First introduced by Lifson in late 1960's (Lifson, 1967; S. L. a. A. Warshel, 1968), molecular mechanics force fields can be interpreted in terms of a relatively simple equation with intra- and inter- molecular forces within the molecular system.

$$
\begin{aligned}
V &= V_{bond} + V_{angle} + V_{torsion} + V_{oop} + V_{ele} + V_{vdW} \\
&= \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\varphi - \gamma)) \\
&+ \sum_{oop} k_i(\omega_i - \omega_0) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)
\end{aligned}
\tag{2}
$$

Here V denotes the potential energies. The first term in Equation 2 is the interaction between pairs of bonded atoms, modeled by a harmonic spring that gives the increase in energy as the bond length l_i deviates from the standard value $l_{i,0}$. The second term is a summation over all angles θ_i in the molecule using a harmonic potential as well. The third and fourth term represent the torsional and out-of-plane potential which describes how the energy changes when a bond rotates. The fifth term is the nonbonded potential between two all pairs of separate atoms (i and j), including Lenard-Jones potential for Van der Waals interactions and a Coulomb potential for electrostatic interactions (Leach, 2001). Descriptive representations of these kinds of interactions are shown in Figure 1.

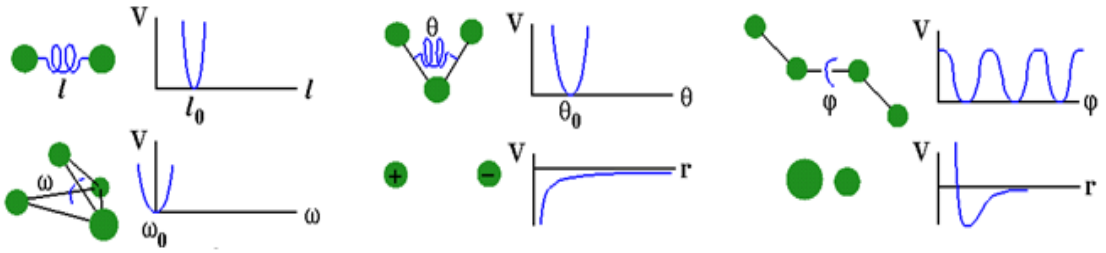


Figure 1. Representation of force field models and potentials of each term.

There have been a great number of force fields used over the years for simulation of biomolecules since the early 1980s. The current generation of classical force fields for biological systems include AMBER, CHARMM, OPLS, and MM, to name a few.

The AMBER force field (Assisted Model Building with Energy Refinement) was originally parameterized for a limited number of organic systems and then widely used for proteins and nucleic acids (Cornell et al., 1995; Weiner et al., 1984). Back then, AMBER included an explicit hydrogen-bond energy term in the potential function, which was different from most other force fields (K. I. Ramachandran, 2008). The MM family of force fields (MM2, MM3 and MM4) was introduced by Allinger et al. (Allinger, Chen, & Lii, 1996; N. L. Allinger, Y. H. Yuh, & J. H. Lii, 1989). MM force fields were parameterized to fit value obtained from high quality electron diffraction experiments; therefore, they are often considered the "gold standard". CHARMM is the abbreviation of Chemistry at Harvard Macromolecular Mechanics, developed by Karplus et al. (Brooks et al., 1983). OPLS (Optimized Potentials for Liquid Simulations) was initially designed by Jorgensen and his colleagues for water and organic liquids (Jorgensen, 1981).

1.4.2 Polarizable Force Fields

While the current-generation force fields are widely used in many areas of biological and materials sciences, several aspects require closer inspection, especially the fixed atomic-charge based electrostatic model. First, the restriction to only partial charges and to only the nuclear sites results in a model insufficiently flexible to describe certain features of molecular charge distributions. Second, the use of fixed-charges means that the model lacks transferability among different chemical and physical environments. In other words, they are unable to respond directly to the environment (Ponder & Case, 2003a). Third, fixed-charge models are difficult to parameterize in many ways. The most

common approach is to empirically scale the charges derived from gas-phase quantum mechanics calculations. As a result, the atomic charges vary significantly among force fields depending on the parameterization procedure (Hu, Elstner, & Hermans, 2003).

These classical additive force fields are still going strong in molecular modeling, but the need for explicit inclusion of polarization has caught a great deal of consideration in the past few years (Cieplak, Dupradeau, Duan, & Wang, 2009; Ponder & Case, 2003b; Roux & Berneche, 2002; van der Vaart, Bursulaya, Brooks, & Merz, 2000). Fixed-charge force fields typically employ effective pair potentials that include many-body effects in an average way. The increasing accuracy of current force fields raises the demand for the explicit inclusion of these non-additive properties, which is made possible by the growth in computer power (Ferenczy & Reynolds, 2001). It is believed that explicit polarization is required if a single set of parameters is to correctly describe the system regardless what environment is, be it gas-phase or bulk. Besides, the ability to transfer quantum-derived electrostatics to bulk-phase modeling is a major advantage of polarizable force fields over fixed-charge models (Cieplak, Dupradeau, Duan, & Wang, 2009).

The development of polarizable force fields is a popular area of current research. There are three basic methods for including polarization: fluctuating charge, Drude oscillator and induced dipole models (Cieplak et al., 2009; Lopes, Roux, & MacKerell, 2009, Ponder & Case, 2003a).

The fluctuating charge model is based on the principle of electronegativity equalization, i.e. a charge flows between atoms until electronegativity of the atoms

equalizes. This method has been used in several force fields (Banks et al., 1999; Rappe, Casewit, Colwell, Goddard, & Skiff, 1992) and was employed by the polarizable water model TIP4P/FQ. The drawback of the fluctuating charge idea is that it allows the magnitude of charge to change but lacks electronic directionality. Thus it does not reproduce out-of-plane polarization for planar molecules like benzene (Cieplak et al., 2009).

Drude oscillator methods are sometimes referred to as shell models or charge-on-spring models in which a mobile charge is tethered to an atom by a harmonic restraint. The atom carries a charge at the nucleus and a restrained charge at a variable position. The electronic polarization is mimicked by separation of both charges due to a peripheral electrostatic field. Since Jacucci reported the first implementation of the Drude oscillator method in 1974 (Jaccuci G, 1974), a great deal of work has been carried out for the improvement and application of this method (Geerke, van Gunsteren, & Sk, 2007; Lamoureux, Harder, Vorobyov, Roux, & MacKerell, 2006; Mitchell & Fincham, 1993). While the Drude model is more flexible in handling polarization than the fluctuating charge model, the additional off-nucleus charge sometimes poses an issue in parameterization (Illingworth & Domene, 2009).

The induced dipole model is another well-studied approach for molecular polarization. The polarization energy is described as the interaction between static point charge and the dipole moment they induce. Despite the biggest downside of this approach, i.e., computational demand due to the iterative calculation of induced dipoles,

the superiority over classical models is promising. The AMOEBA force field has expanded on this model by including interaction between induced dipoles and higher order multipoles up to quadrupoles (P. Y. Ren & Ponder, 2003; P. Y. Ren, Ponder, 2002). The NEMO force field explores the possibility of including interactions between permanent electrostatics and higher-order induced multipoles (Holt & Karlstrom, 2008). In this work, we focus on the AMOEBA force field invented by Ponder and Ren.

The AMOEBA force field (Atomic Multipole Optimized Energetics for Biomolecular Applications) was first developed for water molecule in 2002 (P. Y. Ren, Ponder, 2002). Now, it has a complete set of parameters for protein and a good number of small molecules. Permanent atomic charges, dipoles and quadrupoles are placed on each atom in form of $M=[q,\mu_1,\ \mu_2,\ \mu_3,Q_{11},Q_{12},...,Q_{33}]^T$. Polarization effects are explicitly treated in the AMOEBA force field via mutual induction of dipoles at atomic centers. Induced dipole for each atom can be computed as $\mu_i^{ind}=\alpha_i E_{i,\alpha}$, where α_i is the atomic polarizability and $E_{i,\alpha}$ is the sum of the fields generated by both permanent multipoles and induced dipoles:

$$\mu_{i,\alpha}^{ind} = \alpha_i(\sum\nolimits_{\{j\}} T_\alpha^{ij} M_j + \sum\nolimits_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind} \qquad (3)$$

M_j is the permanent multipole components and T is the interaction tensor between site i and j. The solution for the induced dipole can be written as:

$$\mu_{i,\alpha}^{ind}(n+1) = \mu_{i,\alpha}^{ind}(n) + \alpha_i \sum\nolimits_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind} \qquad (4)$$

The induced dipole can then be solved by iteration or matrix solution.

Dipole interactions are damped by replacing point charges with a smeared charge distribution in order to avoid "polarization catastrophe" according to Thole's model (Thole & Md, 1981). As a result, the dipole interaction energy approaches a finite value instead of becoming infinite due to mutual induction at short separation distance.

$$\rho = \frac{3a}{4\pi}\exp(-au^3) \qquad (5)$$

In this equation, $u=R/(\alpha_i\alpha_j)^{1/6}$ is the effective distance as a function of polarizabilities for the pair of atoms *i* and *j*. The factor *a* is a dimensionless width parameter of the smeared charge distribution which controls the strength of damping.

1.5 COMPUTATIONAL SIMULATION

Molecular modelers seek to understand and to predict the properties of liquids, solutions and solids, and to study complex processes. In such systems, experimental measurements are made on macroscopic samples that contain extremely large numbers of atoms or molecules. Computer simulation methods facilitate the study of such systems and predict their properties through the use of techniques that consider small replications of the macroscopic system with manageable numbers of atoms or molecules. A simulation generates representative configurations (ensemble) of these small replications in such a way that accurate values of structural and thermodynamic properties can be obtained with a feasible amount of computation. Simulation techniques also enable the time-dependent behavior, e.g. kinetics, of atomic and molecular systems to be determined.

1.5.1 *Ab initio* molecular dynamics simulation

Due to the cost of treating the electronic degrees of freedom, the computational cost of QM simulation is much higher than any other simulation methods. However, beginning at the lowest level of description, quantum mechanics is the most accurate simulation method of all. Above this level, *ab initio* molecular dynamics (AIMD) computes the forces acting on the nuclei from electronic structure calculations that are performed "on-the-fly" as the molecular dynamics trajectory is generated (Iftimie, Minary, & Tuckerman, 2005). In this fashion, the electronic variables are not integrated out in advance, but are considered as active degrees of freedom. AIMD simulation has been used for systems called "chemically complex" where many different atom or molecule types give rise to a myriad of diffrent interatomic interactions (Hutter, 2000). The Car-Parrinello method (Car & Parrinello, 1985) is a representative of *ab initio* molecular dynamics techniques which utilizes density functional theory to calculate the forces. This CPMD method has played an important role in deepening our fundamental understanding of water and aqueous solutions (Boero, Terakura, Ikeshoji, Liew, & Parrinello, 2000).

1.5.2 Molecular Dynamics Simulation

As probably the most popular simulation method, molecular dynamics calculates the "real" dynamics of the system, from which time averages of properties can be calculated. Atomic positions are derived by applying Newton's equations of motion. Molecular dynamics is a deterministic method, i.e., the state of the system at any future time can be predicted from its present state. Early attempt of simulations were performed

using hard-sphere potential (Alder, 1957). The particles move in straight lines at constant velocity between collisions. After a collision, the new velocities of the colliding spheres are calculated based on conservation of linear momentum. In more realistic potentials, the force between two atoms changes continuously with their separation. That requires the equations of motion to be integrated by breaking the calculation into a series of very short time steps. At each step, the forces on the atoms are first computed. Then new positions and velocities of the atoms can be predicted based on the acceleration rate together with old positions and velocities. The atoms then traveled to the new sites and the new forces exerted on the atoms are computed. In this way, a MD simulation creates a trajectory that describes how the thermodynamic variables change with time (Leach, 2001).

Since the first biomolecular MD simulation was done by McCammon in 1977 (McCammon, Gelin, & Karplus, 1977), MD simulation has been widely used to study proteins and nucleic acids. Several reviews have covered the progress in MD simulations (Karplus, 2003; Levitt, 2001; Norberg & Nilsson, 2002).

1.5.3 Monte Carlo Simulations

In a MD simulation, the successive configurations of the system are connected in time. This is not the case in a Monte Carlo simulation, where each configuration depends only upon its immediate precursor and not upon any other of the configurations previously visited. The Monte Carlo method generates positions of atoms randomly and uses a criterion to determine whether to accept this new configuration or not. Metropolis algorithm is the most common criterion used in MC simulations. The energy change of

the new and old configuration is evaluated. The new configuration will be accepted when new energy is lower or the Boltzmann factor of the energy difference is less than a random number between 0 and 1. If it is rejected, the old configuration will be preserved for the new move.

Unlike MD simulation, which is possible to predict the Cartesian coordinates of the system at any time in the future or in the past, in MC simulation the new configuration only depends on the neighboring predecessor. Additionally, MD simulation samples a 6N-dimensional including positions and momentums of particles, and thus a kinetic energy component in the total energy. However, MC simulation samples 3N-dimensional space with positions of particles only, such that the total energy is solely determined from the potential energy.

1.5.4 Brownian Dynamics simulation

BD simulation is used to simulate the dynamics of particles that undergo Brownian motion (Carmesin & Kremer, 1988). However, it introduced a few new approximations that allow one to perform simulations on the microsecond timescale. BD technique takes advantage of the fact that there is a big distinction in timescale between the rapid motion of solvent molecules and the much slower movement of large solute molecules. The ability to coarse-grain out these fast modes allows one to simulate much larger timescale than MD. Hence, BD is particularly useful to calculate the diffusion properties for systems where there is a large gap of timescale controling the motion of different components (Zeng, Yu, & Lu, 2008).

Other microscale simulation approaches include Dissipative particle dynamics (DPD), a particle-based method that simulate both Newtonian and non-Newtonian fluids (Hoogerbrugge & Koelman, 1992), Dynamic DFT and Lattice Boltzmann (Chen & Doolen, 1998), both modeling the dynamics of polymer solutions.

1.6 Overview of Thesis Work

In this work, there are several tasks to be accomplished. First, we build polarizable models for divalent metal ions which have hardly been successfully modeled before, and study the thermodynamics of ions in solvation, which is of fundamental importance in many biological and chemical processes. Second, we explore free energy simulation techniques for ligand-protein binding prediction based on a novel paolriazable atomic multipole based potential. We apply our model to the binding simulation of trypsin with charged inhibitors. Last, we calculate the hydration free energy of small organic molecules with AMOEBA and develop software tools that automate the simulation/calculation process.

2 Modeling Divalent Metal ions

2.1 INTRODUCTION

Ions play critical roles in fundamental biological functions including signal transduction, enzymatic activities, and organizing the structure of proteins and nucleic acids. Besides acting as nonspecific salt buffers, ions also interact with biomolecules in specific fashions (e.g. ion channels, metalloproteases.) In fact, many biological processes have been found to be ion specific. For example, many protein kinases require Mg^{2+} in coordination with ATP to facilitate phosphorylation, and the binding of Ca^{2+} to calmodulin is involved in DNA synthesis and cell division (Ivano Bertini, 2001). In addition, a recent review has underscored the importance of Zn^{2+} and other metal ions in survival and pathogenesis of many viruses including HIV, hepatitis, herpes simplex, Rubella and influenza virus (Chaturvedi & Shrivastava, 2005). Even though both are divalent ions of slightly different sizes, the ability of calcium and magnesium ions to coordinate with ligands differs between the two. Experimentally, it has been shown that Mg^{2+} binds six water molecules in an octahedral organization (Caminiti, Licheri, Piccaluga, & Pinna, 1977) , while the coordination number of Ca^{2+}, reported from various X-ray, neutron diffraction and EXAFS experiments, varies from 6 to 10 (Fulton, Heald, Badyal, & Simonson, 2003; Hewish, Neilson, & Enderby, 1982; Jalilehvand et al., 2001; Megyes, Grosz, Radnai, Bako, & Palinkas, 2004).

Ion solvation thermodynamics has been of great interest, as the interplay between the ion-water and ion-protein interactions may provide the basis for ion selection. *Ab*

initio molecular dynamic simulations of ion solvation have been reported (Lightstone, Schwegler, Allesch, Gygi, & Galli, 2005; Lightstone, Schwegler, Hood, Gygi, & Galli, 2001; Lyubartsev, Laasonen, & Laaksonen, 2001). However, most of the *ab initio* studies are limited to small systems of a few water molecules for a few picoseconds. Hybrid QM/MM approaches were also attempted in the investigation of ion solvation properties (Hofer, Tran, Schwenk, & Rode, 2004; W. B. Liu, Sakane, Wood, & Doren, 2002; Martinez, Pappalardo, & Marcos, 1999; Rempe & Pratt, 2001; Schwenk, Loeffler, & Rode, 2001). The quasi-chemical approach treats interactions between ions and immediate water molecules (inner-shell) quantum mechanically, along with applications of a dielectric continuum model for outer-shell medium (Martin, Hay, & Pratt, 1998). This approach has produced solvation thermodynamics consistent with experimental data for a wide variety of ions (Asthagiri, Pratt, Paulaitis, & Rempe, 2004; Rempe, Asthagiri, & Pratt, 2004a). On the other hand, classical molecular mechanical approaches are much more attractive computationally for larger molecular systems such as proteins. Detailed information at the atomic level pertaining to interactions, thermodynamics, structure and kinetics can be derived from such extensive simulations. However, quantitative ion solvation thermodynamic data, such as ion solvation free energies, is required a priori in order to arrive at sensible ion parameters in the classical additive model. Nonetheless, it has been shown previously that single ion solvation free energy is difficult to derive from experimental measurement, unlike the relative solvation free energy between ions or the solvation free energies of whole salt. In contrast, polarizable models that account for

many-body effect can be derived from high-level *ab initio* calculation in gas phase and extended to solution with confidence (Dang, 2002; Grossfield, Ren, & Ponder, 2003; Spangberg & Hermansson, 2004; Yague, Mohammed, Loeffler, & Rode, 2003). The Thole dipole induction model adopted by our model has been compared favorably with other approaches to describe polarization effects (Masia, Probst, & Rey, 2004).

Previously, a polarizable molecular mechanic model was successfully applied to the study of solvation of monovalent ion in water and other solvents. Electrostatics in this model is represented by atomic multipole moments with explicit atomic dipole induction. In this study, we report the extension of this polarizable model to divalent calcium and magnesium ions' interactions with water. *Ab initio* calculations of ion-water interaction in gas-phase are utilized to derive van der Waals parameters of the ion. Molecular dynamics simulation of ion solvation using Particle-Mesh Ewald is described. Ion solvation free energies are computed from both free energy perturbation (FEP) and Bennett acceptance ration (BAR) methods. Ion solvation structures and dynamics from the molecular dynamics simulations are compared with those of monovalent ions as well as experimental and other theoretical results.

2.2 METHOD

2.2.1 *Ab initio* calculations and parameterization

The potential model used in this work was based on the one previously reported for water, K^+, Na^+, and Cl^- (Grossfield et al., 2003). Only the nonbonded parameters for electrostatics and van der Waals interactions were obtained, because ions do not bind to

any other atoms. Apparently, charge of magnesium and calcium is 2+. Besides, neither of them has dipole or quadrupole values. The atomic polarizabilities of both ions were determined from B3LYP/6-31G* calculation which was performed using the Gaussian 03 package. Since the unit of polarizability from Gaussian calculation is $Bohr^3$, we needed to convert it to $Å^3$ which is compatible with AMOEBA force field (Bohr Radius = 0.529 Å).

The damping factor of each ion was determined together with van der Waals parameters by fitting to the QM calculation of ion-water dimer energy profile. The geometry of ion-water dimer was first fully optimized followed by single point calculations with these geometries. Ion-oxygen separation distance was varied between 1.5 Å and 5 Å with water geometry fixed at the optimized one, and the binding energy was obtained for each distance. The binding energies were computed as the total energy minus the isolated water and ion energies as if they were separated by an infinite distance. Basis-set superimposition error (BSSE) was removed in all calculations. Both MP2/6-311++G(3df,3pd) and MP2/aug-cc-pVTZ basis sets were used for Mg^{2+}, and MP2/6-311++G(3df,3pd) for Ca^{2+}. AMOEBA calculation was carried out using the same geometry as QM by TINKER package. The distance dependence of dimer binding energies was used to adjust vdW parameters including the radius and well depth (R and ε) and damping factors (*a*) of Ca^{2+} and Mg^{2+}.

2.2.2 Free energy simulations

The ion solvation free energies of Ca^{2+} and Mg^{2+} were computed from molecular dynamics simulations. First, we computed the solvation free energy for a neutral vdW

particle by running 12 independent simulations which scaled the calcium parameters according to

$$R(\lambda) = 1 + \lambda(R_{final} - 1)$$
$$\varepsilon(\lambda) = \lambda(\varepsilon_{final}) \qquad (6)$$

for λ = (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). This was followed by 12 simulations during which the ion's charge and polarizability were set to

$$q(\lambda) = \lambda(q_{final})$$
$$\alpha(\lambda) = \lambda(\alpha_{final}) \qquad (7)$$

for λ = (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). The Ca^{2+} ion was then perturbed to Mg^{2+} by changing both the vdW parameters and damping factor in three steps.

Molecular dynamics simulations were performed using the TINKER package (Ponder, 2001). The long range electrostatics was treated using Particle Mesh Ewald summation for atomic multipoles (Sagui, Pedersen, & Darden, 2004) with a cutoff of 7 Å in the real space and 0.75 Å spacing and 5th-order spline in the reciprocal space. Induced dipoles were iterated until the changes in atomic induced dipoles were less than 0.01 Debye. Molecular dynamics simulations were performed with a 1 fs time step for 300 ps. Coordinates of all atoms were saved every 0.1 ps, with the first 50 ps discarded as equilibration. The temperature was maintained at 298 K using the Berendsen weak coupling method (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984b). To investigate the effect of system size, a single ion was placed in a periodic cubic box of either 216 or 512 water molecules, with 18.64 Å or 25 Å on a side.

The Helmholtz free energy changes between adjacent simulations were calculated two different ways, first using free energy perturbation (FEP) methods (Jorgensen & Ravimohan, 1985), then using the Bennett acceptance ratio method (Bennett, 1976). In the standard FEP approach, the free energy changes between adjacent steps were computed as the average of the forward and the backward perturbations, and the error for each step was estimated from the difference between the average and the forward or the backward perturbation result. The overall error was computed as the sum of errors of all constituent steps. Using the Bennett formation, the free energy change between simulations λ_i and λ_{i+1} was computed iteratively using

$$\Delta A(j)_{\lambda_i \to \lambda_{i+1}} = \ln \frac{\left\langle 1/\left[1+\exp((E_{\lambda_i} - E_{\lambda_{i+1}} + C)/RT)\right]\right\rangle_{\lambda_{i+1}}}{\left\langle 1/\left[1+\exp((E_{\lambda_{i+1}} - E_{\lambda_i} + C)/RT)\right]\right\rangle_{\lambda_i}} + C - \ln \frac{n_{\lambda_{i+1}}}{n_{\lambda_i}} \qquad (8)$$

where C is given by:

$$C = \Delta A(j-1)_{\lambda_0 \to \lambda_1} \qquad (9)$$

and j is the current iteration. Here, $E_{\lambda i}$ is the total energy of the system evaluated using the parameters from λ_i. The subscripts outside the averaging brackets denote the MD trajectory used for evaluation of E. The variable n is the number of trajectory snapshots in each simulation. For j=1, the initial value of $C = \Delta A(0)$ was given an arbitrary value as a rough estimate of the free energy change. Iterations continued until the value of ($\Delta A(j) - \Delta A(j\text{-}1)$) < 0.01 kcal/mol. The final values calculated for ΔA were independent of the

initial values given for C. The statistical error of BAR method was estimated as the sum of the square root of the variance of ΔA between successive simulations according to:

$$\sigma^2_{\lambda i \to \lambda i+1} = \frac{\left\langle f^2 \right\rangle_{\lambda i} - \left\langle f \right\rangle^2_{\lambda i}}{n_{\lambda i} \left\langle f \right\rangle^2_{\lambda i}} + \frac{\left\langle f^2 \right\rangle_{\lambda i+1} - \left\langle f \right\rangle^2_{\lambda i+1}}{n_{\lambda i} \left\langle f \right\rangle^2_{\lambda i}}$$

$$\left\langle f \right\rangle_{\lambda i} = \left\langle 1/1+\exp((E_{\lambda i+1} - E_{\lambda i} + C)/RT) \right\rangle_{\lambda i} \quad (10)$$

$$\left\langle f \right\rangle_{\lambda i+1} = \left\langle 1/1+\exp((E_{\lambda i+1} - E_{\lambda i} + C)/RT) \right\rangle_{\lambda i+1}$$

where C is the final free energy value calculated from BAR method. In the above formula, n refers to the number of random samples that are independent of each other. In our calculations, we estimated the error during the particle growth by counting every 0.5 ps as one independent data point, and every 1 ps during charging based on the relaxation time scale of bulk water and water molecules in the first solvation shell. The error for the free energy change from Ca^{2+} to Mg^{2+} could be underestimated as the water molecules around Mg^{2+} relax at a much slower scale as will be discussed below.

The structure and dynamics of water molecules in the first solvation shell were analyzed using the MD trajectory from the final charging stage, where the ion was fully charged. All results were based on the simulation of the 512 water system. In the remaining analyses, we defined the ion's first solvation shell to be all water molecules positioned within the first minimum of the radial distribution function (RDF) of the O-ion. For Mg^{2+}, additional 300 ps simulations (total 600 ps) have been performed to investigate the dynamics. Time correlation functions have been computed for the fluctuation of the first shell coordination number, from which the relaxation time of the first shell water is derived using an exponential decay model (Grossfield, 2005).

2.3 RESULTS AND DISCUSSION

2.3.1 Gas phase ion-water dimer interaction

A polarizable potential is capable of capturing the many-body effect in electrostatics when moving from one environment to another. As a result, parameters in the polarizable models can be conveniently determined and verified by comparison to the high-level *ab initio* results in gas-phase, as was previously demonstrated for monovalent ions (Grossfield et al., 2003). With the polarizabilities of Ca^{2+} and Mg^{2+} derived from DFT and the water model from previous work (P. Ren, Ponder, JW, 2003), the parameters remaining to be determined were primarily the van der Waals *R* and ε for each ion. We have chosen the vdW parameters to best match the *ab initio* binding energies of ion-water dimer in the gas-phase. The same approach has been shown to be effective in our previous study of K^+ and Na^+. An additional parameter, the damping coefficient, has been adjusted for the divalent ions to modify the polarization between the cation and other atoms at short distances. A recent investigation on dipole induction between cations and water showed that the Thole's induction model overestimates the induced dipole moments at short range when the original damping coefficient is used. As this was not the case for a point charge polarizing a water molecule, this effective reduction in polarizability has been attributed to the repulsion between the electron distributions of the ion and water. We indeed found it necessary to reduce the damping coefficient, i.e. enhance the damping, in order to match the *ab initio* equilibrium dimer binding energy and separation simultaneously. When our standard damping coefficient, 0.39, is used, the equilibrium dimer separation distance is shorter than the *ab initio* distance by 5% when

the binding energies agree. Figure 2 compares the distance dependence of binding energies given by the final model and *ab initio* calculations.

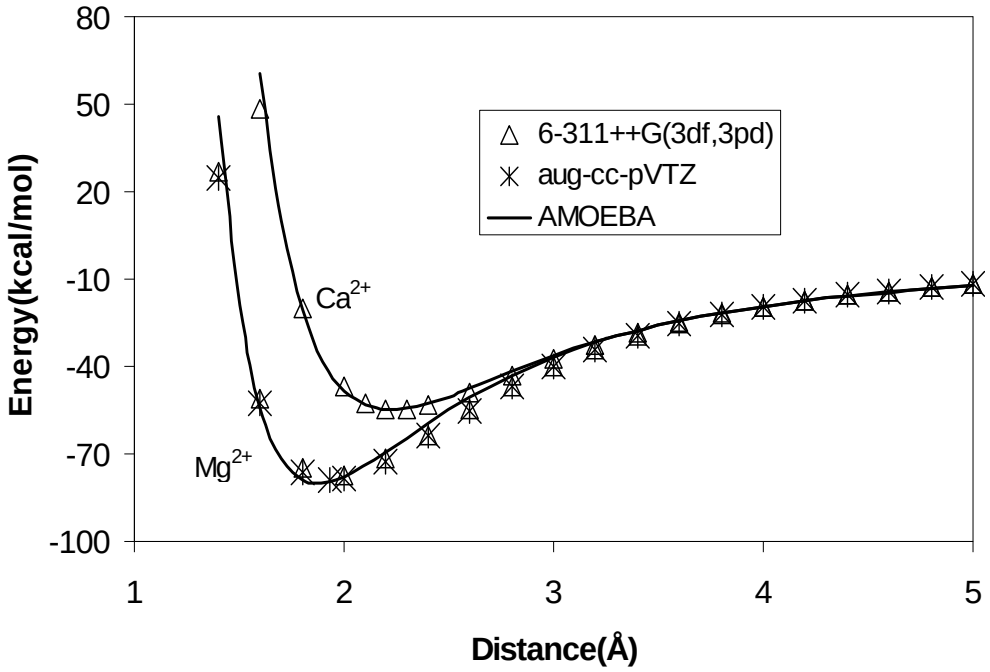


Figure 2. Ion-water dimer binding energy in gas phase as a function of ion and oxygen separation distance.

The final parameters of the two cations are listed in Table 1. The two basis sets used for Mg^{2+} gave consistent binding energies over a range of distances. The agreement between the final model and *ab initio* results is rather satisfactory. As expected, Mg^{2+} binds stronger than Ca^{2+} to water. The equilibrium distance between Mg^{2+} and water is 0.3 Å shorter than that of Ca^{2+}-water while the equilibrium binding energy is lower by 25 kcal/mol.

Table 1. Parameters for ions. R and ε are diameter and well depth for van der Waals potential in Å and kcal/mol, respectively. α is the polarizability in $Å^3$. a is the dimensionless damping coefficient in eq (2).

Ion	R	ε	α	A
Ca^{2+}	3.63	0.35	0.55	0.159
Mg^{2+}	3.21	0.28	0.08	0.095

2.3.2 Solvation thermodynamics

The solvation free energy is the key quantity describing the thermodynamic stability of an ion in solution. Solvation free energies of Ca^{2+} and Mg^{2+} ions in water have been computed from molecular dynamics simulations where a single ion is grown gradually in a water box by first turning on its vdW parameters, then the ionic charges and polarizabilities.

Table 2. Solvation free energy of calcium and magnesium ion in water. The number in parenthesis is the estimated error. 1 M in gas phase is chosen as the standard state.

		216 water	512 water	Schmid *et al.* [1]	Asthagiri *et al.* [2]
Ca^{2+}	FEP	-359.5 (7.0)	-360.3 (13.8)	-357.2	-354.7
	BAR	-357.4 (2.0)	-354.9 (1.7)		
Mg^{2+}	BAR		-431.1 (2.9)	-435.4	-433.3

1. Ref (Schmid, Miah, & Sapunov, 2000); 2. Ref (Asthagiri et al., 2004).

Table 2 lists the solvation free energies of Ca^{2+} and Mg^{2+}, which are about four to five times greater than those of monovalent K^+ and Na^+ (Grossfield et al., 2003). For the purpose of comparison, two different approaches, FEP and BAR, were used to obtain solvation free energies of Ca^{2+} based on the same set of simulations. Our results show

that using the BAR method significantly reduces the statistical uncertainty using the same amount of simulation data, in agreement with what others have reported (Shirts & Pande, 2005). Further, for the system under study, the difference between the free energies computed by FEP and BAR occurs for charge growth beyond 1 e, where the effective energy change between successive stages is largest. The solvation free energy of Mg^{2+} is obtained by turning Ca^{2+} into Mg^{2+} through the adjustment of the vdW and damping parameters. Increasing the system size from 216 to 512 waters leads to only negligible changes in the solvation free energy within the statistical uncertainty. The free energies from BAR method compare favorably to those from quasi-chemical method (Asthagiri et al., 2004) and the theoretical evaluation of Schmid (Schmid et al., 2000). In the quasi-chemical method, the ion and the immediately adjacent water molecules are treated quantum mechanically and kept fixed while the surrounding water is described by classical mechanics. Recently the same group has confirmed that there is indeed an "inner" shell of four water molecules around K^+ using *ab initio* molecular dynamics (Rempe, Asthagiri, & Pratt, 2004b). Due to the fact that experimentally it is only possible to measure the solvation free energy of whole salts, extrathermodynamic assumptions are used in order to determine the contributions from the cations and anions. By setting the proton hydration free energy, Schmid was able to estimate solvation free energies of other ions based on experimental free energies of whole salt.

The classic Born theory of ion solvation states that there exists an effective solvation radius, a_B, for each ion such that the solvation free energy of the ion in a dielectric medium is given by

$$\Delta A = \frac{q^2}{2a_B}\left(1 - \frac{1}{\varepsilon_d}\right) \qquad (11)$$

where q is the charge of the ion and ε_d is the dielectric constant of the medium. We have calculated the effective radius of Na^+, K^+, Ca^{2+} and Mg^{2+} based on the Born equation and the solvation free energy obtained from our simulations. As shown in Table 3, Mg^{2+} has the smallest radius while K^+ has the largest. Ca^{2+} and Na^+ have almost the same size according to the Born radius.

Table 3. The effective sizes of ions as indicated by Born theory and RDF.

	Effective Born Radius	First peak in ion-O RDF	First minimum in ion-O RDF
Mg^{2+}	1.56	2.07	2.95
Ca^{2+}	1.89	2.41	3.23
Na^{+} [1]	1.87	2.39	3.29
K^{+} [1]	2.30	2.76	3.53

1. The Born radii for Na^+ and K^+ are computed based on solvation free energies from ref. (Grossfield et al., 2003). The RDF values are taken from ref. (Grossfield, 2005).

2.3.3 Solvent structure and dynamics

To characterize the structure of water molecules around the ion, the radial distribution function (RDF) has been sampled from the dynamics trajectories. In Figure 3a and b, the RDF and their running integrations are shown for Ca-O and Mg-O respectively.

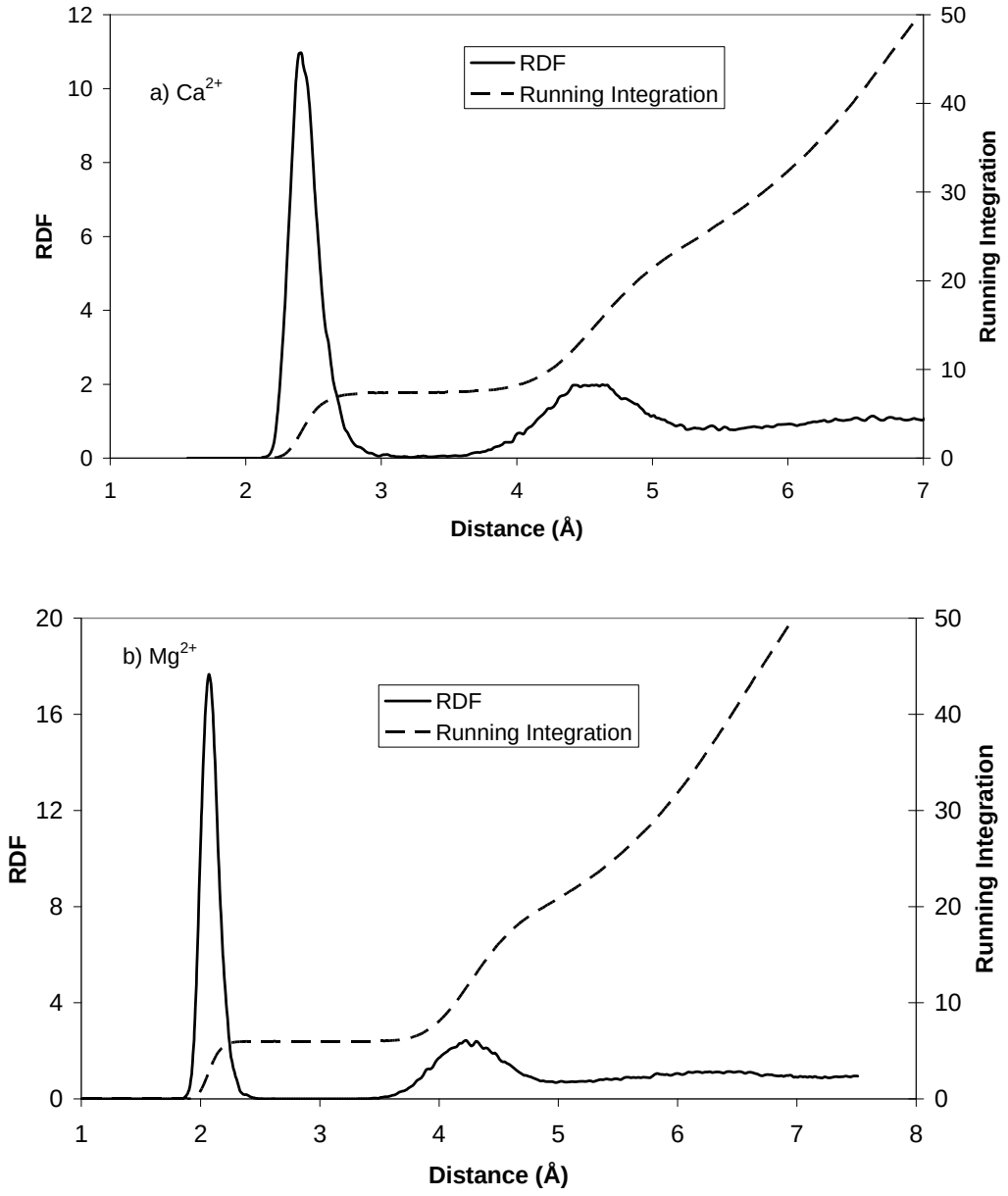


Figure 3. Radial distribution functions of ion and oxygen atom in water. a) Ca^{2+}, b) Mg^{2+}

The first peak of the RDF is located at 2.41 Å for Ca^{2+}, and 2.07 Å for Mg^{2+}. Previous work reported a first peak at 2.76 Å for K^+, and 2.39 Å for Na^+ (Grossfield, 2005). The order of the first peak location among these ions is consistent with the effective Born radius, i.e. $K^+ > Na^+ \cong Ca^{2+} > Mg^{2+}$. As shown in Table 3, the differences between the effective radii and the first peak positions are almost a constant of 0.5 Å for all four ions. The size of the ion plus the first shell of water molecules is related to the position of the first minimum in RDF. There also appears to be a constant offset of 1.3 Å between an ion's Born radius and the location of the first minimum, which can be considered as the "effective" size of the ion plus first shell water solvent. The height of the first peaks is much more prominent for Ca^{2+} and Mg^{2+} than for Na^+ and K^+ (Grossfield, 2005), correlating to the solvation free energies rather than the size of the ions. The sharp peaks indicate the highly ordered water structure around the divalent ions. Also, the first valleys of Mg-O and Ca-O RDF are wide and flat in contrast to those of Na^+ and K^+, signifying a clear separation between the first and second solvation shells (Figure 4).

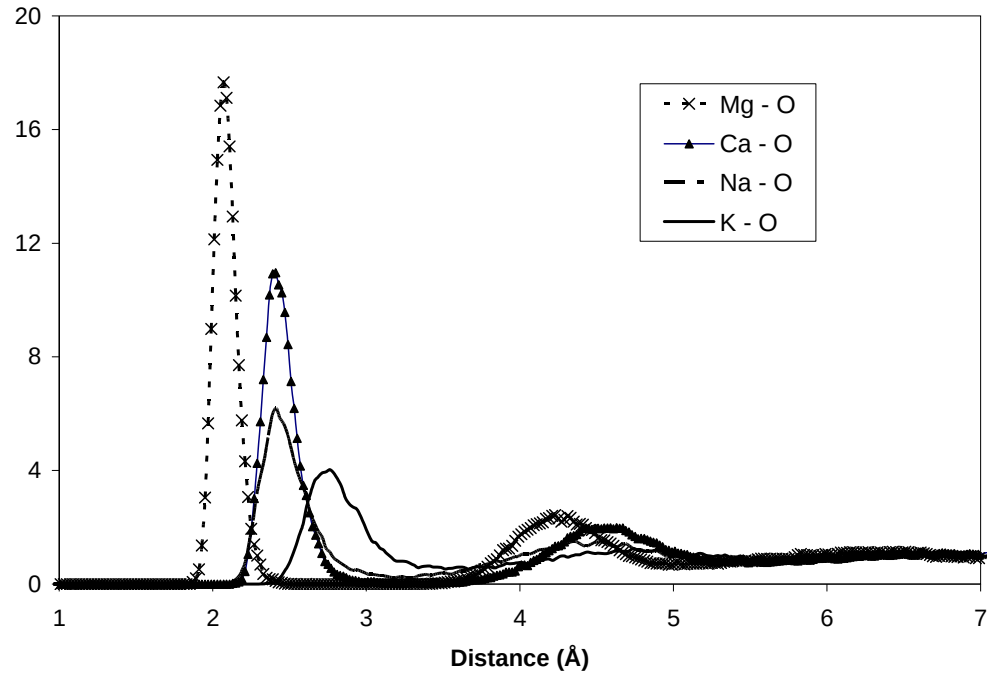


Figure 4. Comparison of RDFs of monovalent and divalent cations in water.

From the running integration of the RDF, the average coordination number for Mg^{2+} was found to be 6, in agreement with experimental(Caminiti et al., 1977) and *ab initio* MD results (Lightstone et al., 2001). For Ca^{2+}, a coordination number of 7.3 was obtained, consistent with an X-ray experimental value of 7.2±1.2. Recent *ab initio* MD simulations of Ca^{2+} in 60 water molecules reported a value of 6.2 or 7.0 depending on the flexibility of the water molecules used (Lightstone et al., 2005). Thus our model accurately describes the difference between Ca^{2+} and Mg^{2+} in water coordination.

To examine the effect of the ions on nearby solvent structure, the radial distribution functions of oxygen-oxygen pair in the first solvation shell have been

calculated. First, we have computed RDF for oxygen pairs of which at least one is in the first solvation shell. Comparison is made between the divalent ions and K^+ in Figure 5.

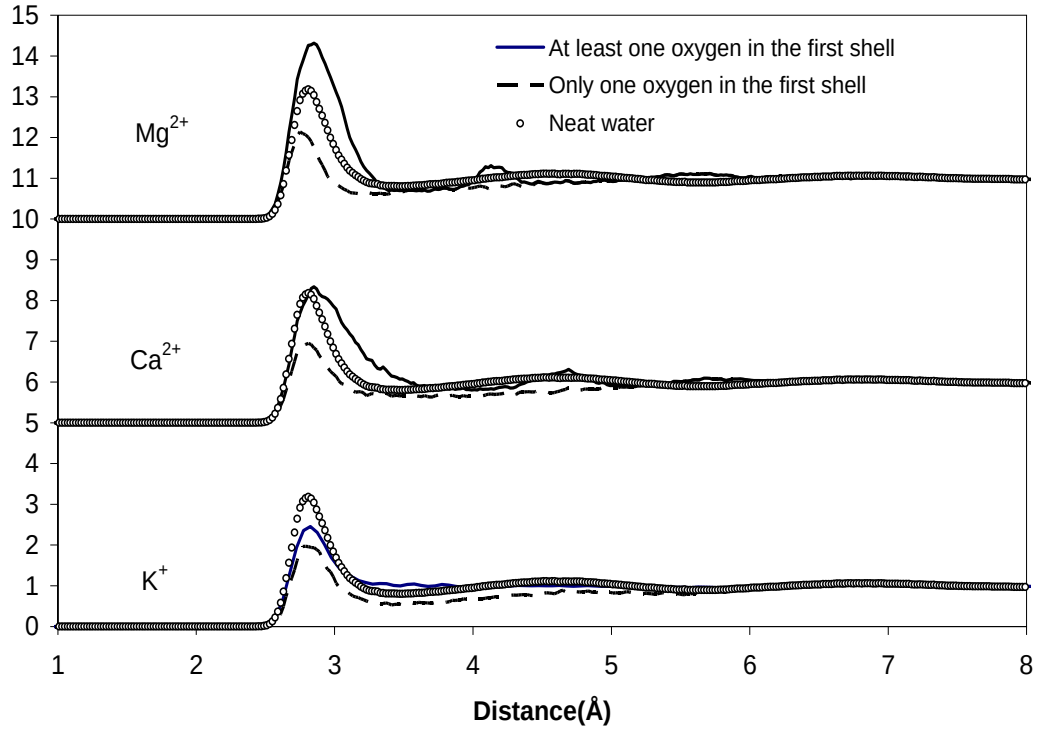


Figure 5. Oxygen-oxygen radial distribution functions of water molecules in the first solvation shell. The results for K^+ are taken from ref (Grossfield, 2005). RDFs for Ca^{2+} and Mg^{2+} are offset by 10 and 5, respectively.

Interestingly, the O-O RDFs around both Mg^{2+} and Ca^{2+} have more pronounced first peaks than K^+. In the case of Mg^{2+}, the first peak is even higher than the bulk water. The RDFs of Ca^{2+} and Mg^{2+} also display second peaks that do not exist for K^+. It is however possible that these peaks originate from oxygen pairs in the first solvation shell. We have therefore also computed RDFs between oxygen pairs with only one oxygen atom in the first shell. As shown in Figure 5, the resulting RDFs have significantly reduced first peaks and the second peaks completely disappear. This dramatic change

confirms that the water molecules in the first solvation shell are highly organized by the divalent ions. By contrast, the O-O RDFs of K^+ display less feature than bulk water no matter whether the pairs in the solvation shell are counted or not. The reduced correlation between the water in the first shell and surrounding water signifies the disruption of solvent structure by the cations.

To further describe the organization of water molecules immediately adjacent to the cations, the angle distributions of O-X-O, X=Ca^{2+} or Mg^{2+}, sampled from the MD simulations are plotted in Figure 6.

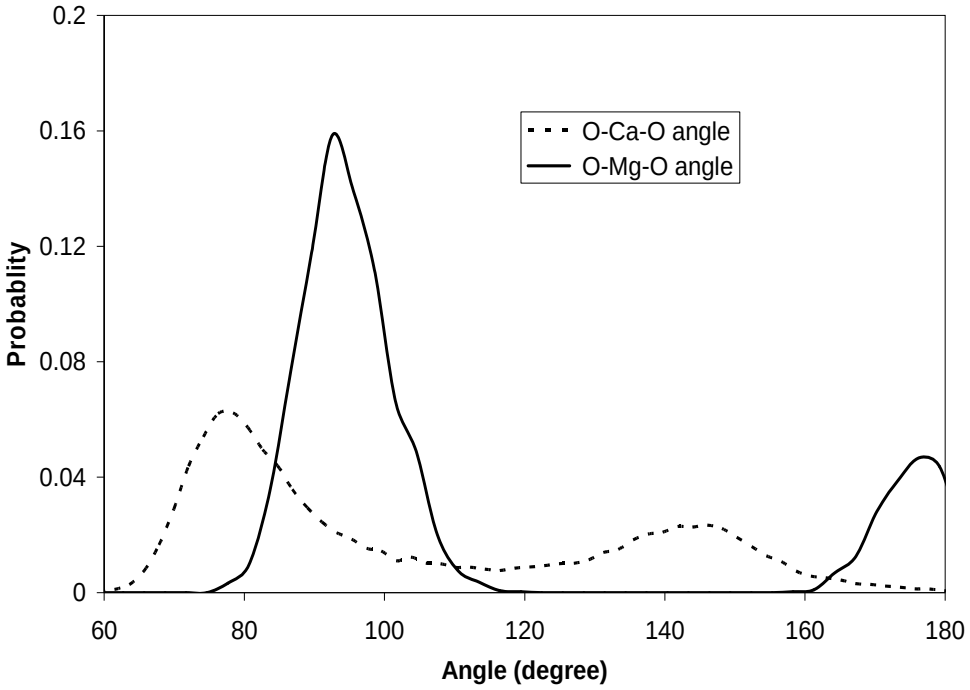


Figure 6. The O-ion-O angle distribution in the first solvation shell.

The O-Mg-O is predominantly distributed around 92˚ and 176°, indicating an octahedral coordination as also determined by X-ray experiment (Caminiti et al., 1977). In contrast, the O-Ca-O has a much broader distribution that peaks at 78˚ and 147°.

The ion solvation free energy is a thermodynamic indicator of how well an ion is solvated in the water whereas the relative solvation free energies among different solvents determine the partitioning of the ion between these solvents. However, biologically one must also consider ionic kinetics, which is of great importance whenever the ion changes environments, as when it enters a channel or binds to a protein. We have investigated the life time of ion-water coordination by examining the time correlation function of the instantaneous first shell coordination number. For Ca^{2+}, the relaxation time in the first solvation shell is 18 ps and coordination number fluctuate between 5 and 9 on a time scale of 1 ~ 2 ps. For Mg^{2+}, a relaxation time of 228 ps was obtained and the coordination number only deviate from 6 briefly during the whole 600 ps simulations. Relaxation times of 0.8 ps and 1.8 ps were reported previously for K^+ and Na^+ (Grossfield, 2005). Thus the water molecules in the first solvation shell of Mg^{2+} will remain bound for hundreds of picoseconds whilst the water molecules around Ca^{2+} and monovalent K^+ and Na^+ move in and out of the first shell much more frequently.

The self-diffusion coefficients were computed from the mean-squared displacement sampled during MD simulations. The Ca^{2+} exhibits a diffusion coefficient of 0.8×10^{-5} cm^2 s^{-1}, higher than that of Mg^{2+}, 0.3×10^{-5} cm^2 s^{-1}. The experimental diffusion coefficients for Ca^{2+} and Mg^{2+} are 0.79 and 0.71×10^{-5} cm^2 s^{-1}, respectively.

Spangberg and Hermansson also reported a somewhat low diffusion constant (0.4×10^{-5} cm^2 s^{-1}) for Mg^{2+} in water from MD simulations using polarizable potentials. The reduction of mobility from Ca^{2+} to Mg^{2+} in our simulation is most likely due to the strong interaction between the latter and water; effectively, we are measuring the diffusion coefficient as a much larger super-particle containing 6 waters and an ion.

The dipole moment distribution of the water molecule in the first solvation shell is shown in Figure 7.

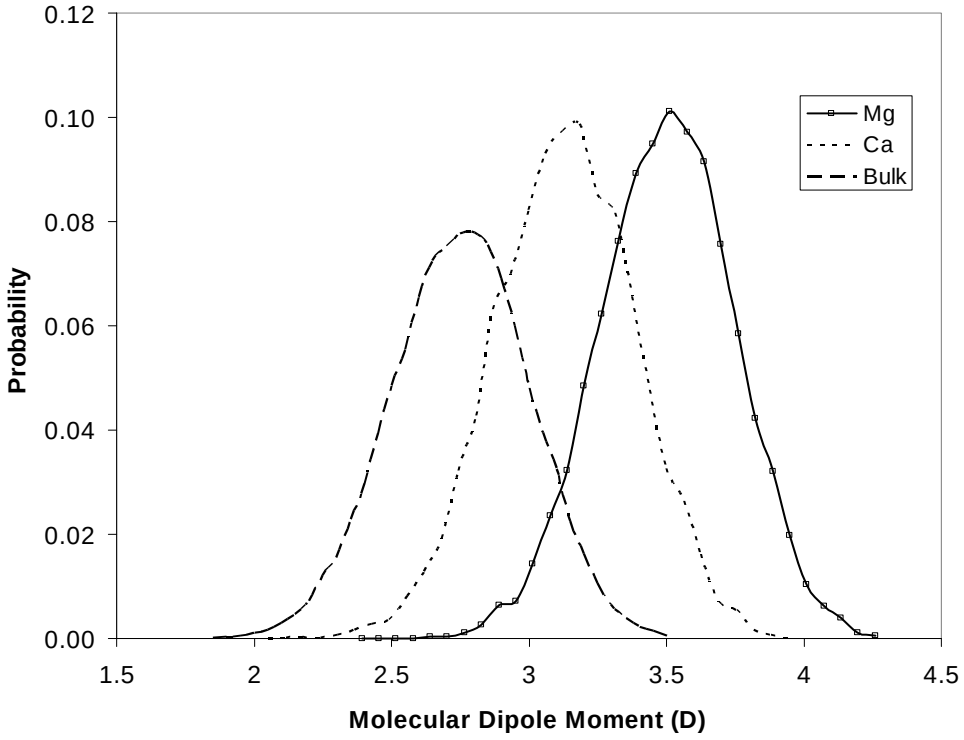


Figure 7. The dipole moment of water molecules in the first solvation shell.

The average dipole moments of water molecules around the ion are greater than that of bulk water (2.77 D) due to the polarization effect. Mg^{2+} displays stronger

induction on water than Ca^{2+} likely because of its smaller effective radii as discussed earlier. On the other hand, for Na^{+} and K^{+}, the average molecular dipole moments in the first shell were reported to be roughly the same as those in bulk (Grossfield, 2005). These results suggest that the dipole moment of solvent is affected by both the size and, more importantly, the valence of the ion species.

2.4 CONCLUSION

A polarizable model has been applied to the simulation of solvation of Mg^{2+} and Ca^{2+} ions in water. The parameters for the ions have been derived based on the *ab initio* ion-water dimer interaction energies in the gas-phase. The single ion solvation free energies predicted by molecular dynamics simulations agree well with other theoretical estimations. The resulting solvation free energies of Ca^{2+} and Mg^{2+} are four to five times greater than those of K^{+} and Na^{+}. The Bennett acceptance ratio method appears to be more accurate and computationally more efficient than the traditional free energy perturbation approach for free energy calculations. The use of a 216 water system is shown to be adequate for computing accurate solvation free energies.

The results of molecular dynamics simulations suggest that the divalent cations perturb the structure and dipole moments of the first solvation shell water considerably, in contrast to monovalent ions. The water structures in the first solvation shells of the divalent ions are more ordered that those around monovalent ions, demonstrated by the sharp first peak in the RDFs. Additionally, the separation between the first and second shell is more prominent. Based on the RDF and Born theory, the effective sizes of the

ions compared in this study are in the order of $K^+ > Na^+ \cong Ca^{2+} > Mg^{2+}$. The average water coordination numbers for Ca^{2+} and Mg^{2+} are 7.3 and 6, respectively. Furthermore, Mg^{2+} is found to bind tightly to six water molecules in an octahedral geometry in agreement with experiment. The dynamic fluctuations in the first shell coordination number indicate that the life time of Ca^{2+} - water coordination is about 18 ps, ten times longer than the relaxation time previously reported for K^+ or Na^+. Even though Mg^{2+} is only slightly smaller than Ca^{2+}, the life time of water molecules around Mg^{2+} is on the order of a few hundreds of picoseconds, such that the desolvation kinetics will have a strong influence on the ability of Mg^{2+} to bind other molecules.

3 Calculation of Absolute Protein-Ligand Binding Free Energy with a Polarizable Force Field

3.1 INTRODUCTION

Specific recognition of ligands by proteins is at the core of many crucial biological functions and systems such as enzyme catalysis and intracellular signaling. Binding affinity characterizes the strength of such recognition. With the recent advancements in computing, prediction of the binding affinity based on physical principles of molecular interaction has come to the forefront of active research and has been the subject of regular reviews (Brandsdal et al., 2003b; Gilson & Zhou, 2007; Jorgensen & Ei, 2004; Kollman et al., 2000; Lamb, Jorgensen, & Yw, 1997). All-atom molecular dynamics (MD) simulation with explicit solvent, coupled with efficient free energy sampling algorithms, can potentially offer accurate prediction of binding free energies of ligands to proteins (Gilson & Zhou, 2007). Common free energy simulation algorithms include the double decoupling method (DDM) and potential of mean force approach (PMF). Free energy perturbation (FEP), or thermodynamic integration (TI) can be employed to compute free energy differences in either DDM or PMF. It has been argued that DDM is problematic for charged systems, since the binding free energy is computed as a small difference between two large solvation energies in water and in protein (Woo & Roux, 2005). On the other hand, the PMF approach does not quantify absolute solvation energies of ligand, which makes it difficult to detect potential problems in treatment of long-range effect and boundary conditions (Burykin, Schutz,

Villa, & Warshel, 2002; Warshel, Sharma, Kato, & Parson, 2006). A comparison of PMF and FEP in ion channel study indicated the former suffered more seriously from hysteresis (Kato & Warshel, 2005). Alternatives to free energy pathway calculations include linear response analysis (LRA) (Lee, Chu, Bolger, & Warshel, 1992) and linear interaction energy (LIE) (Aqvist, Medina, & Samuelsson, 1994), where only the ligand-bound and unbound states are simulated. A semi-macroscopic approach based on Protein Dipoles Langevin Dipole (PDLD/S) was applied previously in LRA to further reduce the computational cost (Sham, Chu, Tao, & Warshel, 2000). Recent reviews have summarized some of the advantages and drawbacks of LRA and LIE (Gilson & Zhou, 2007; Warshel, Kato, & Pisliakov, 2007).

MD/FEP methods have been used to calculate the absolute binding free energies of different protein-ligand systems, such as L99A mutant of T4 lysozyme with benzene (Boresch, Tettinger, Leitgeb, & Karplus, 2003; Deng & Roux, 2006; Hermans & Wang, 1997), tyrosyl-tRNA-synthetase with tyrosine (Boresch et al., 2003), FKBP with several ligands (Fujitani, Tanida, Ito, Jayachandran et al., 2005; J. Y. Wang, Deng, & Roux, 2006), and human Lck SH2 domain with phosphotyrosine peptide (Woo & Roux, 2005), to name a few. Strong correlation between computed binding free energies with experimental values has been reported for a series of ligands binding to FKBP and lysozyme. Nonetheless, the calculated absolute binding free energies can still deviate from experimental measurement by several kilocalories. There have been a limited number of simulation studies of highly charged systems. Recently, the PMF approach

was used successfully in calculating the binding free energies of a charged peptide binding to the SH2 domain (Woo & Roux, 2005).

It has been recognized that the bottlenecks to achieving chemical accuracy in molecular simulation are the underlying physical models and the sampling convergence (Gilson & Zhou, 2007). The current-generation common force fields employ fixed atomic charges and therefore lack the ability to respond to the actual local electrostatic environment. Explicit treatment of polarization to provide realistic electrostatic representation dates back to Warshel and Levitt's use of atomic induced dipoles in the enzyme reaction study (Warshel & Levitt, 1976). Polarized force field (PFF) was later applied to estimating binding free energies in systems such as trypsin, antibody-antigen, and DNA polymerase (Florian, Goodman, & Warshel, 2002; Lee et al., 1992; Warshel, Sussman, & Hwang, 1988). History and development of PFF have been covered in recent reviews (Ponder & Case, 2003b; Warshel et al., 2007).

Aside from the physical potential, sampling convergence remains an enormous challenge in binding simulations with atomic force fields, especially when a large number of conformational and other degrees of freedom are involved (Kato, Braun-Sand, & Warshel, 2008). The sampling issue is usually due to slow barrier crossing on the rugged energy landscape of complex biomolecules and the relatively short simulation time. There are a variety of advanced techniques have been developed to improve sampling, some of which have been reviewed before (Berne & Straub, 1997; Lei & Duan, 2007; Liwo, Czaplewski, Oldziej, & Scheraga, 2008). First of all, the numerous local energy

barriers can be overcome by modifying the potential energy surface. Umbrella sampling, one of the widely used approaches, involves construction of a compensating function, aka. umbrella, which is added to the true potential energy function in order to bias the sampling to a particular set of conformations (Beutler & Vangunsteren, 1994; Torrie & Valleau, 1977). The weighted-histogram analysis method (WHAM) (Kumar, Bouzida, Swendsen, Kollman, & Rosenberg, 1992)is then applied to remove the contribution from the biasing potential. Hamelberg et al. later developed this noval approach based on umbrella sampling which adds a bias potential without prior knowledge of the conformations of interest (Hamelberg, Mongan, & McCammon, 2004). Second, replica-exchange method (REM) is one of the most effective sampling methods, in which *n* replica systems, each at a different temperature, are simulated (Hansmann, 1997). At given intervals, exchanges of temperatures are attempted between neighboring replicas. The broader application of this powerful sampling technique to larger systems has been hindered by the need for a homogeneous computer cluster with a large number of nodes. A few variants of REM have been developed recently to overcome this problem, for example, replica exchange with solute tempering (REST) (P. Liu, Kim, Friesner, & Berne, 2005) and partial replica exchange molecular dynamics (PREMD) and local replica exchange molecular dynamics (LREMD) (Cheng, Cui, Hornak, & Sinnnerling, 2005). Additionally, sampling can be improved by reducing the degrees of freedom which can be fulfilled by either restraints to internal coordinates (Deng & Roux, 2006; Karplus, 2003) or coarse-graining (Trylska, Tozzini, & McCammon, 2005).

In this study, we report the calculation of the absolute binding free energy of trypsin with the charged ligand benzamidine from molecular dynamics simulations with a polarizable force field. Trypsin is one of the typical serine proteases that are associated with digestion. Serine proteases are a class of enzymes that are characterized by the presence of a serine residue in the active site of the enzyme. They act as important targets for medicinal chemistry that are associated with a wide range of biologically critical processes (Talhout, Villa, Mark, & Engberts, 2003), including blood clotting, immunity, and inflammation. All serine proteases hydrolyze peptide bonds at the catalytic triad called S2 site (Peters & Merz, 2006). Trypsin is synthesized in pancreas and secreted into intestine. In certain circumstances, trypsins are activated in pancreas excessively which will destroy the healthy pancreas cells and leads to pancreatitis consequently. The common idea to suppress trypsin activation is to block the active site by inhibitors. The aspartic acid residue located in the S1 pocket next to the catalytic site S2 can be utilized to provide strong electrostatic interactions with counter-charged substrates. The most popular candidate benzamidine is such a positively charged peptide which forms a salt bridge with the aspartic acid in the S1 site of trypsin as shown in Figure 8 .

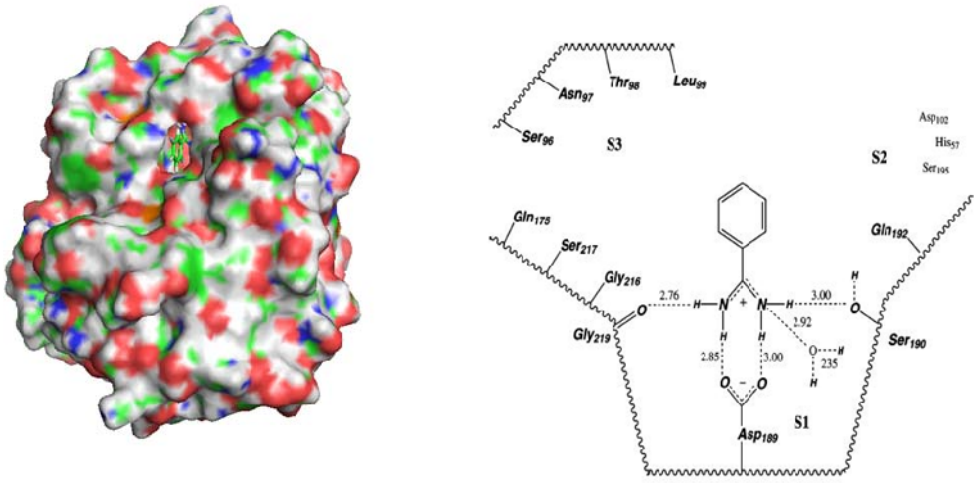


Figure 8. Structure of benzamidine-bound trypsin and the interactions between benzamidine and trypsin in the binding pocket.

Benzamidines carry net charges and are relatively small and rigid. This allows us to achieve adequate sampling and focus on the application of the polarizable potential in the calculation of binding free energies. Besides comparing calculated free energy with experimental literature, we also examined the role of electrostatics and polarization in ligand-protein binding.

3.2 METHOD

3.2.1 Ligand parameterization

The potential energy of the system, i.e. protein, water and ligand, is expressed as the sum of electrostatic, van der Waals and valence terms. The valence terms consist of typical harmonic function for bond stretching, angle bending, three-term Fourier torsional potential and out-of-plane term for trigonal centers, taken from the original MM3 potential developed by Allinger and co-workers (Allinger, Yuh, & Lii, 1989) .

Previously, we have developed potentials based on the above model for water (P. Y. Ren & Ponder, 2003, 2004), ions (Grossfield et al., 2003; Jiao, King, Grossfield, Darden, & Ren, 2006), organic molecules and peptides (Ponder & Case, 2003b; P. Y. Ren & Ponder, 2002), namely the Atomic Multipole Optimized Energetics for Biomolecular Applications model (AMOEBA). Independent studies using these models have also been reported (Jiang, Jordan, Taylor, 2007; Liang, Walsh, 2007; Liang, Walsh, 2006; Tuma, Jenicek, Jungwirth, 2005). Parameters for proteins are freely available with the TINKER modeling package (Ponder, 2004) at ftp://dasher.wustl.edu/pub/tinker/params. In the current work, the AMOEBA potential was used for trypsin without any modification. For the ligand, the parameterization is described as follows. The van der Waals (vdW), bond, angle, torsion, out-of-plane and atomic polarizability parameters of benzamidine were transferred from those of benzene and the guanidinium group of arginine of AMOEBA potential. The equilibrium bond and angle values were adjusted to match geometry given by *ab initio* optimization of the ligands at the level HF/6-31G*. The QM calculations were performed using the Gaussian 03 package (Miller, Hernandez, Handy, Jayatilaka, & Willetts, 1990). The electrostatic parameters including charge, dipole and quadrupole moments at each atom were derived from the density matrix output from the Gaussian 03, using the GDMA program (Stone, 2005). Computed from the GMDA 2.2, with H radius set to 0.31 Å, the multipole values of the benzamidine were found to be insensitive to the choice of the basis set. In this study, values from MP2/6-311++G(2d,2p) were used. The torsional parameters were all

transferred from benzene/histidine or guanidinium groups except for the middle bond that connects the ring and the amidine group, for which the value from C_ε-C_ζ-O-H torsion in AMOEBA tyrosine was found to be adequate for the ligand, as discussed in the Results and Discussion section. The detailed procedure of parameterization is provided in Appendix A.

3.2.2 Absolute binding free energy from double-decoupling simulations

The absolute free energy of benzamidine binding to trypsin was calculated using double-decoupling method by "disappearing" the ligand in both bulk water and solvated protein complex in two separate simulations.

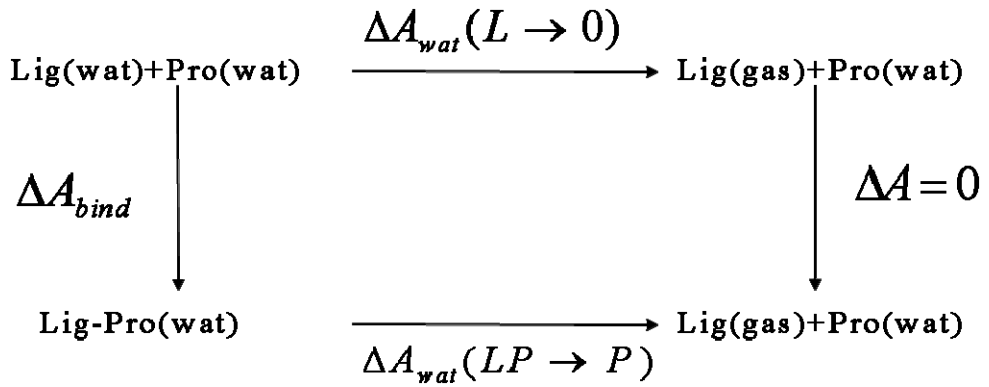


Figure 9. Thermodynamic cycle of absolute binding free energy calculation.

This scheme was originated from the thermodynamic cycle introduced by Lee et al in 1992 (Lee et al., 1992). Theoretically, the binding free energy is defined as the left leg of Figure 9. Due to the fact that free energy is a state function, it is path independent. As long as the endpoints are the same, the free energy can be calculated following another path. To complete the thermodynamic cycle, we computed the binding free

energy by "disappearing" the ligand in both water and protein. The two states on the right hand side of Figure 9 are virtually the same, so the free energy change of the right leg is zero. Therefore the binding free energy is computed as:

$$\Delta A_{bind} = \Delta A_{wat}(L \to 0) - \Delta A_{wat}(LP \to P) \tag{12}$$

The decoupling simulations involved gradually turning off the electrostatic and van der Waals interactions between the ligand and the rest of the system. We decided to turn off electrostatics before vdW in order to prevent atoms on top of each other with repulsion. The electrostatic interactions were decoupled in 10 steps by scaling down the electrostatic parameters of the benzamidine linearly, i.e., by applying the scaling factor λ = {1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0}, according to

$$\begin{aligned} q(\lambda) &= \lambda q_0 \\ \alpha(\lambda) &= \lambda \alpha_0 \end{aligned} \tag{13}$$

where q_o and α_o are original parameters of atomic multipole and polarizability, respectively.

Note that by diminishing the ligand's electrostatic parameters, not only were the electrostatic interactions between the ligand and the environment turned off, but so were the intramolecular electrostatic interactions within the ligand. To complete the thermodynamic cycle and to restore the intramolecular interactions, we recharged the ligand in vacuum after the ligand was decoupled from either water or protein environment. This step, however, is not necessary for computing the binding free energy as this recharging contribution is identical in both ligand-water and ligand-protein

decoupling and thus cancels exactly. This step was carried out merely to obtain the complete solvation free energy of the ligand in water. As described in the Results and Discussion section, this approach of scaling the electrostatic parameters to zero offers a numerical advantage because the resulting decoupling energies in both water and protein are rather small (< 5 kcal/mol). As a result, the calculation of the binding free energy no longer relies on the cancellation of two large numbers. To recharge the ligand in vacuum, we gradually scaled the multipoles and atomic polarizabilities back to their full values in 6 steps. A time step of 0.1 fs was used, with polarization convergence set to 1×10^{-5} D per atom.

The decoupling of the vdW interactions between the ligand and its environment was then carried out. Instead of scaling down the vdW parameters, we modified the pairwise interactions between benzamidine and its surroundings. To avoid singularity at small vdW interaction distances, a potential situation when ligand atoms are in very close contact with other atoms and vdW energy approaches infinity numerically, we replaced the buffered-14-7 vdW function with a soft-core modification (Beutler, Mark, Vanschaik, Gerber, & Vangunsteren, 1994):

$$U_{ij} = \lambda^n \varepsilon_{ij} \frac{1.07^7}{[\alpha(1-\lambda)^2 + (\rho + 0.07)^7]} \left(\frac{1.12}{\alpha(1-\lambda)^2 + \rho^7 + 0.12} - 2 \right) \qquad (14)$$

At λ=1, the above equation reduces to the original buffered-14-7 function. By scaling λ from 1.0 to 0.0, vdW interactions between ligand and its environment were turned off linearly in 10 uniform steps. Soft-core modification was implemented in AMBER/PMEMD (see Appendix C).

MD simulations have been performed in parallel for all steps along the above-mentioned decoupling pathway. The SANDER executable from the AMBER package (version pre9) was used (Case et al., 2005). The benzamidine-trypsin complex (1BTY) was placed in a periodic octahedral water box of 2222 water molecules. The initial dimension of the cube that enclosed the octahedron was 50Å on each side. A single 100 ps NPT dynamics run was performed and the system density was equilibrated to 1 g cm^{-3}. The resulting configuration was then used in all subsequent NVT simulations for decoupling electrostatic and vdW interactions. The same procedure was applied to prepare the ligand-water system. No counter ions were added to neutralize the system. In our previous study of ion hydration free energy, we discussed that it is best not to make correction due to the finite system size and periodicity (Grossfield et al., 2003). At each decoupling step, 1 to 3 ns simulations were carried out as specified in the Results and Discussion. A 1 fs time step was used for all condensed-phase simulations. Atomic coordinates of the simulation system were saved every 500 fs. The temperature was maintained at 298K using Berendsen thermostat (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984a). The vdW cutoff was set to 9 Å, with the long tail correction included. Particle Mesh Ewald (PME) (Sagui, Pomorski, Darden, & Roland, 2004; Toukmaji, Sagui, Board, & Darden, 2000) was used to compute electrostatic interactions, with a real-space cutoff of 7.0 Å. To speed up the simulations, induced dipoles were iterated until the RMS change between steps was less than 0.05 D per atom. In post-MD free energy analysis, a tighter convergence criteria, 10^{-5} D per atom, was used to re-

evaluate the potential energies of saved snapshots. We compared the electrostatic decoupling free energy of the ligand-water system obtained by using this approach to the result calculated using a convergence of 10^{-6} D per atom in both the MD simulation and the free energy analysis, and found the difference to be 0.12 kcal/mol, well within the statistical error. The speedup achieved by convergence of the induced dipole to only 0.05 D per atom in MD simulations was roughly a factor of two. For the fully solvated complex, we were able to attain a 50-ps molecular dynamics trajectory a day on a single processor.

The free energy differences between adjacent steps were computed using Bennett Acceptance Ratio method which has been discussed in the previous chapter. We used snapshots every 1 ps to compute the above uncertainty. The total statistical uncertainty of each vdW or electrostatic decoupling free energy was computed as the sum of the errors from individual steps.

While the interactions between the ligand and trypsin are being switched off, the ligand tends to sample more space. Chances are the ligand will drift out of the binding pocket of trypsin eventually which will pose a big trouble in sampling convergence. That's why a harmonic virtual bond was used to restrain the ligand to the protein pocket during the decoupling process (Boresch et al., 2003; Hamelberg & McCammon, 2004) as Figure 10.

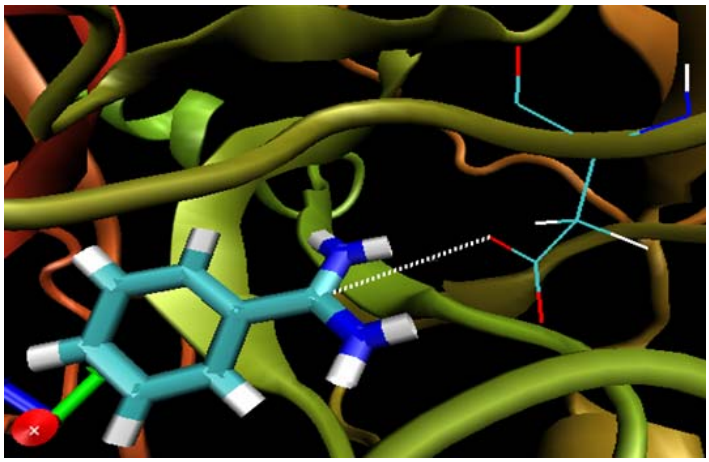


Figure 10. The restraint between the benzamidine (carbon atom of amidinium group) and the trypsin (oxygen atom of Asp 189).

The potential energy of the harmonic bond is expressed as

$$U(r) = \frac{k}{2}(r - r_0)^2 \tag{15}$$

where k is the force constant. The positional fluctuation of benzamidine in the binding pocket of trypsin was measured from 100 ps MD simulations without restraint. A force constant of 20 kcal/mol Å^{-2} was subsequently obtained from $k = 3RT / \langle \delta r^2 \rangle$, where δr is the atomic position fluctuation.

Since the restraint between the ligand and the protein was artificially introduced, we had to remove the bias in the final binding free energy as Figure 11.

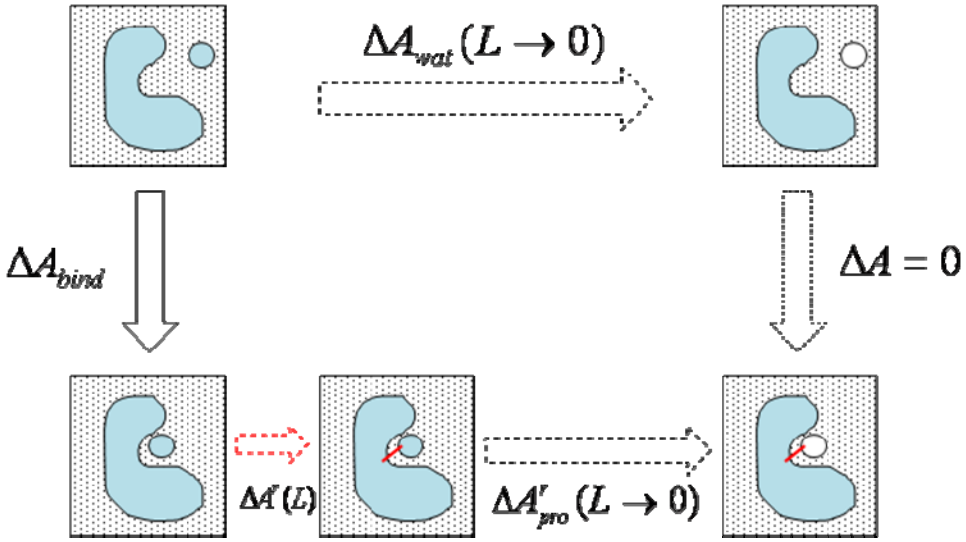


Figure 11. Thermodynamic cycle of double decoupling with restraint. Red line between ligand (circle) and protein represents the artificial restraint.

$$\Delta A_{bind} = \Delta A_{wat}(L \to 0) - \Delta A^{r}_{pro}(L \to 0) - \Delta A^{r}(L) - RT \ln\left[C^{o} \left(\frac{2\pi RT}{k} \right)^{3/2} \right] \qquad (16)$$

A correction was added in the Equation (17). The first correction *ΔA'(L)* is the free energy change via exerting the restraint when the interactions between the protein and the ligand are intact. This is expected to be a small contribution, because the full interations are strong enough to hold the ligand in the pocket even without a restraint. The second correction was calculated as $-RT\ln(C^{o}V)$, where C^{o} is the standard concentration and V is the sampling volume of the ligand under the restraint (Gilson, Given, Bush, & McCammon, 1997; Hamelberg & McCammon, 2004). This term amounts to 5.87 kcal/mol for k = 20 kcal/mol Å^{-2}, and 6.48 kcal/mol for k = 40 kcal/mol Å^{-2}.

3.2.3 Polarization

To evaluate the contribution of polarization to the binding free energy, we computed the free energy due to the induction between the ligand and its environment. In the current polarization model, short range atomic dipole induction is damped using a smeared charge distribution proposed by Thole (Thole, 1981)

$$\rho = \frac{3a}{4\pi}\exp(-au^3) \qquad (17)$$

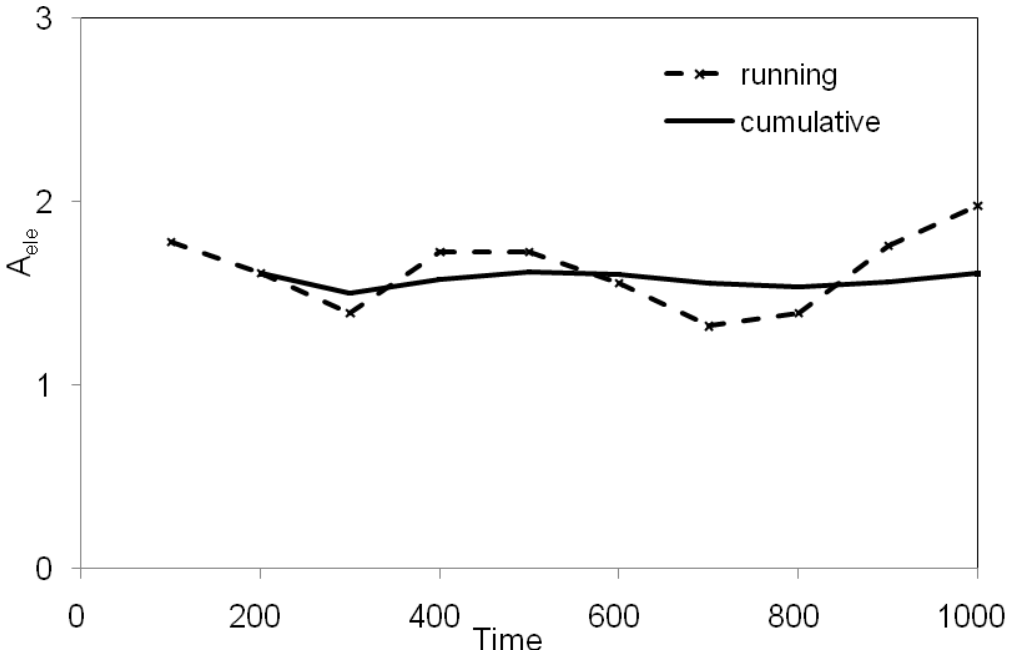
where u is the effective distance between the two atoms that polarize each other (P. Y. Ren & Ponder, 2003). The damping is critical to achieve anisotropic molecular response with isotropic atomic polarizability. Factor a controls the strength of damping. The smaller the damping factor is, the stronger the damping, and hence the weaker the polarization energy. By setting a to zero, it is then possible to turn off the polarization (dipole induction, interaction energy and force) between the specific pairs of atoms. We calculated the free energy changes arising from the polarization between ligand and water, and between ligand and protein in solution, by scaling a from the original 0.39 to zero in 5 steps (a = {0.39, 0.039, 0.0039, 0.00039, 0}). The polarization within water or protein-water was not modified. A 500 ps MD simulation run was carried out at each step. As dipole induction is short-ranged, a cut-off of 14 Å and 8 Å was used for damping in protein and in bulk water, respectively. We have verified that longer cut-off values do not affect the polarization free energy reported here.

3.3 RESULTS AND DISCUSSION

3.3.1Absolute binding free energy

To evaluate the absolute binding free energy of benzamidine to trypsin, the free energies of decoupling benzamidine from water and trypsin-water were computed from MD simulations respectively. The decoupling free energies were evaluated from a path in which the electrostatic and then the vdW interactions between benzamidine and its environment were turned off in steps. A soft-core version of buffered-14-7 potential with $n = 5/\alpha = 0.7$ was used in the vdW decoupling. A harmonic potential (k= 20 kcal/mol Å^{-2}) was used to restrain the benzamidine to the trypsin.

Figure 12 and Figure 13 show the decoupling free energies of the ligand-water system and the ligand-protein system, respectively.


3
2
1
0
A_{ele}
running
cumulative
0
200
400
600
800
1000
Time

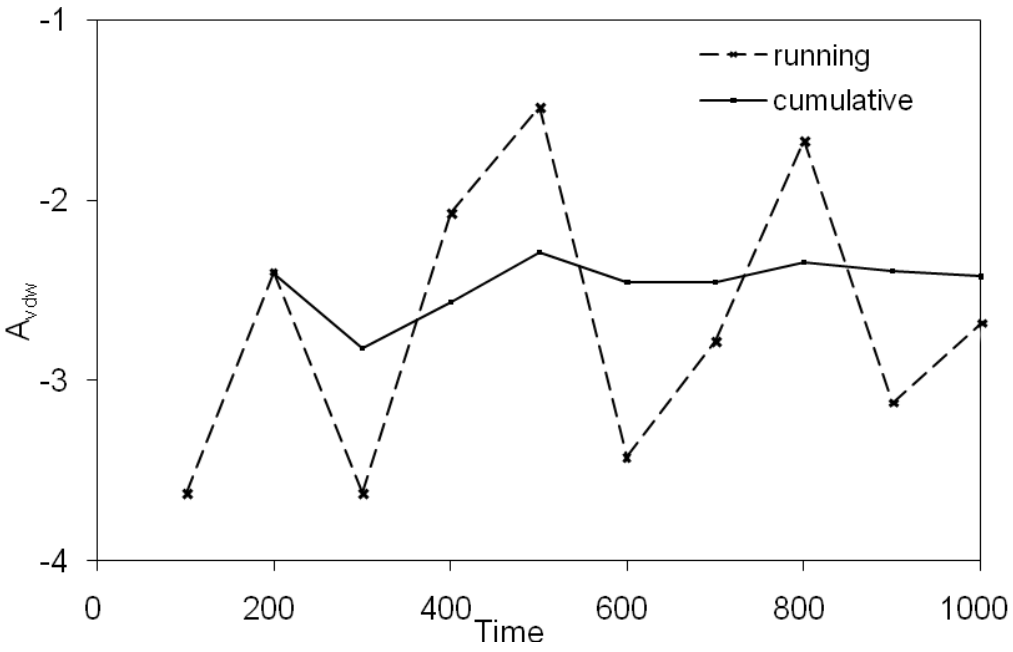
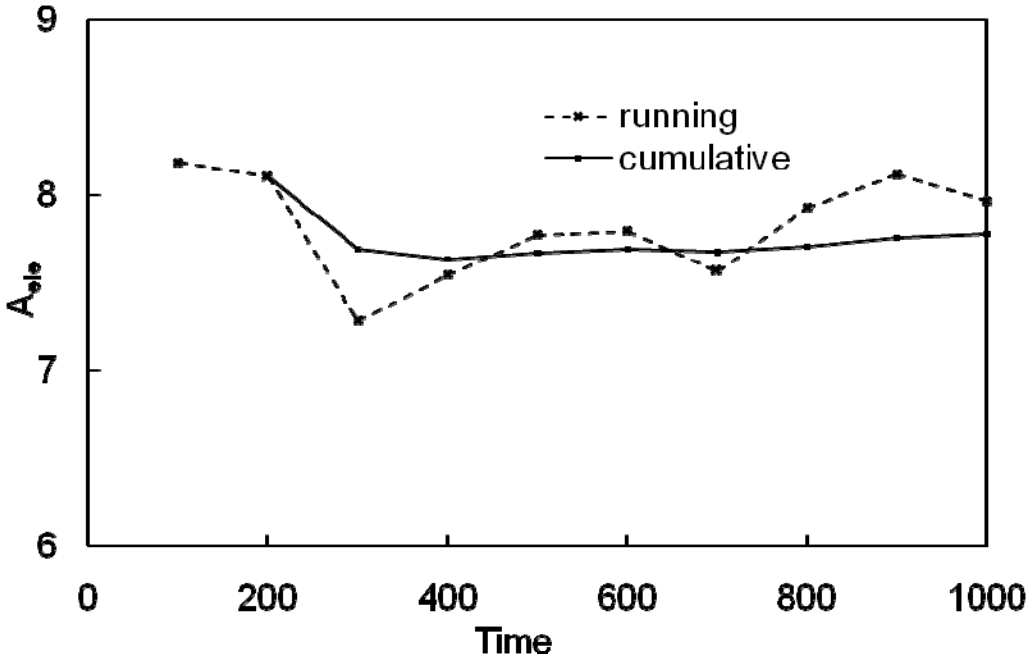


Figure 12. Electrostatic (a) and van der Waals (b) decoupling free energies (kcal mol) of ligand-water system. The dashed line with cross markers is the running average of every 100 ps block. The solid line with square markers is the cumulative average.


9
8
7
6
A_{ele}
running
cumulative
0
200
400
600
800
1000
Time

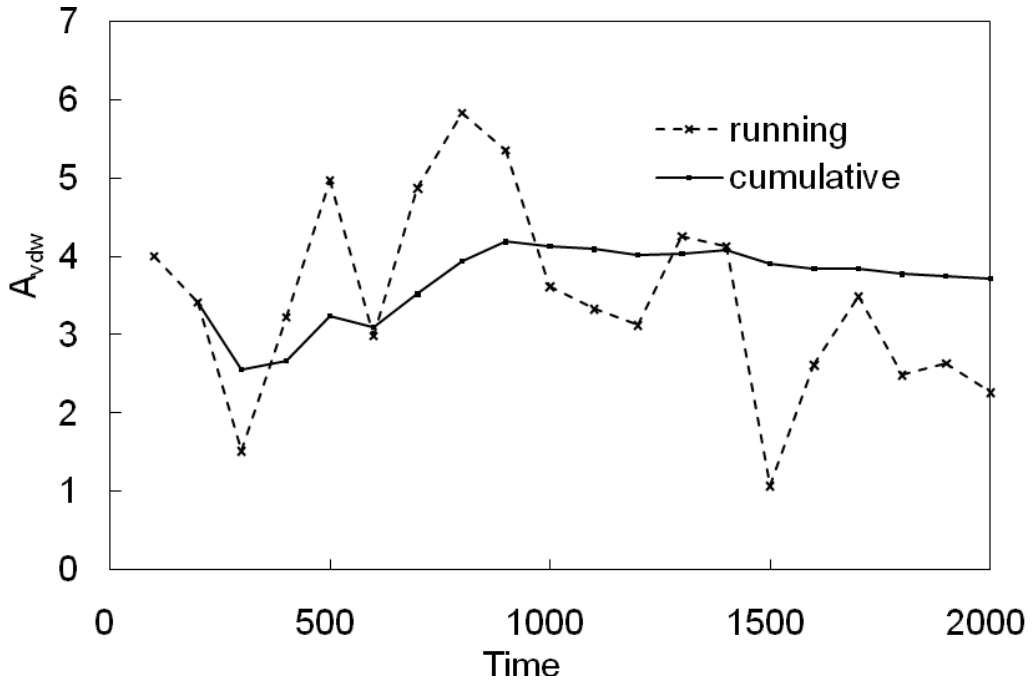

7
6
5
4
3
2
1
0
A_{vdw}
running
cumulative
0
500
1000
1500
2000
Time

Figure 13. Electrostatic (a) and van der Waals (b) decoupling free energies of ligand-protein system (kcal mol). The dashed lines with cross markers are the running average of every 100 ps block. The solid lines with square markers are the cumulative average.

The first 100 ps of MD trajectories in all simulations were considered as system equilibration and subsequently ignored in the free energy analysis. In addition to calculating cumulative averages, running averages were computed from 100 ps blocks of trajectories to illustrate the fluctuation. Note that the running averages do not reflect the statistical error in the final free energy. The free energies of the ligand-water system were reasonably converged in 1 ns (Figure 12), while the ligand-protein free energy took longer to settle down, even with a restraining bond between the ligand and protein (Figure 13). This is expected as trypsin-water is much more complex than bulk water and thus requires longer simulations. For both ligand-water and ligand-protein, the electrostatic free energy fluctuates much less than the vdW component. This is due to the full van der Waals interactions being present during electrostatic decoupling, which confines benzamidine to the pocket with low mobility. In contrast, as the vdW interactions between benzamidine and its environment were gradually decoupled, the benzamidine molecule occupied greater and greater regions of space, coming in close contact with and eventually penetrating the surrounding water and protein atoms. Additionally, water molecules were found to occupy the pocket after about 500 ps as benzamidine was annihilated from the S1 pocket of trypsin. All these factors contribute to the wild fluctuations in the vdW free energy.

The electrostatic and the van der Waals decoupling free energies were determined to be 1.27±0.2 kcal/mol and -2.42±0.4 kcal/mol, respectively, for benzamidine-water, and 7.78±0.2 kcal/mol and 3.72±0.3 kcal/mol for benzamidine-trypsin (Table 4).

Table 4. Absolute free energy of benzamidine binding to trypsin computed using different force constants and soft-core coefficients. The numbers in parenthesis are the standard errors.

Restraint constant (kcal/mol Å^{-2})	*Soft-core*	$\Delta A_{wat}(L\rightarrow 0)$		$\Delta A_{pro}(L\rightarrow 0)$		*Restraint correction*	ΔA_{calc}	ΔA_{exp}
		ΔA_{ele}	ΔA_{vdw}	ΔA_{ele}	ΔA_{vdw}			
20	0.5/4	1.27 (0.2)	-2.27 (0.4)	7.78 (0.2)	3.42 (0.3)	6.26	-6.72	-6.3[b] -7.3[c] -6.4[d] -6.7[e]
20	0.7/5	1.27[f]	-2.42 (0.4)	7.78[f]	3.72 (0.3)	6.26	-7.27	
40	0.7/5	1.27[f]	-2.35[a]	7.57 (0.2)	4.56 (0.2)	7.03	-7.28	

[a] Averaged from the two ligand-water vdW decoupling free energies in the rows above.
[b] Ref. (Schwarzl, Tschopp, Smith, Fischer, & Er, 2002)
[c] Ref. (Katz et al., 2001)
[d] Ref. (Maresguia, Nelson, & Rogana, 1977)
[e] Ref. (Mares-Guia, 1965)
[f] The value is taken from the row above

Thus, the free energy of benzamidine binding to trypsin is -7.27 kcal/mol, which includes a correction of 6.26 kcal/mol for the bias due to the restraint. Several experimental binding free energies were reported, ranging from 6.3 to 7.3 kcal/mol (Katz et al., 2001; Katz et al., 2000; Mares-Guia, 1965; Schwarzl, Tschopp, Smith, Fischer et al., 2002). It is not uncommon for the experimental binding affinity to vary up to a factor of 10 (1.3 kcal/mol), depending on the assay conditions (Schwarzl, Tschopp, Smith,

Fischer et al., 2002), or the specific experimental method such as spectrophotometry (Schwarzl, Tschopp, Smith, Fischer et al., 2002) and crystallography (Katz et al., 2000).

The free energy of decoupling benzamidine from water appears to be much lower than the expected solvation free energy of a charged molecule, which is typically several tens of kcal/mol. As noted in the Methods section, we decoupled the electrostatic interaction by zeroing out the atomic multipoles and polarizabilities of the benzamidine; and the end states of the ligand-water and ligand-protein simulations feature a "ghost" benzamidine molecule without intramolecular electrostatic interactions. We have determined that the recharging free energy of benzamidine in vacuum is 46.92 kcal/mol. Combining this value with the decoupling energy above, the electrostatic solvation free energy of benzamidine in water becomes -48.19 kcal/mol (Table 5).

Table 5. Absolute solvation free energies of benzamidine in water and trypsin.

	Protein		*Water*		*Binding*		
	ele[a]	*vdw*[b]	*ele*[a]	*vdw*	*ele*	*vdw*	*Total*
Benzamidine	-54.60[c]	2.17[c]	-48.19	2.35[c]	-6.41	-0.18	-6.59

[a] Intramolecular contribution of -46.92 kcal mol is included.
[b] Restraint correction is included in the vdW component.
[c] Averaged from values in Table 4.

Thus, a large portion of solvation free energy of charged benzamidine (both in water and trypsin environments) is actually not responsible for driving its binding to trypsin. Our electrostatic parameter scaling approach thus avoids the numerical problem associated with the double decoupling method when applied to charged systems (Woo &

Roux, 2005), as the binding free energy no longer relies on cancellation of two large solvation free energy values.

3.3.2 Effect of soft-core vdW potential

Free energy is a state function, and therefore independent of the sampling path. We investigated two soft-core modifications of the buffered-14-7 function, $n = 5/\alpha = 0.7$ and $n = 4/\alpha = 0.5$, in the calculation of vdW decoupling free energy in ligand-water and ligand-protein. The free energies computed using the two potentials converge toward each other after about 1 ns of simulation. Using $n = 4/\alpha = 0.5$, the van der Waals decoupling free energies were found to be -2.27±0.4 and 3.42±0.3 kcal/mol for ligand-protein and ligand-water, respectively, in comparison with -2.42±0.4 and 3.72±0.3 kcal/mol from simulations with $n = 5/\alpha = 0.7$ (Table 4). The differences between the two sets of values are comparable to the statistical error.

3.3.3 Free energy as driving force for binding.

It is possible to decode the physical driving force behind benzamidine binding to trypsin since our calculations offer detailed information on atomic interactions that are not easily measurable by experimental procedures. However, this is complicated due to the presence of the restraint between benzamidine and trypsin. We overcame this complication by comparing simulations results from different restraints.

Besides using a force constant of 20 kcal/mol $Å^{-2}$ for the restraint, we performed another set of simulations of trypsin-benzamidine with doubled restraint strength (k = 40 kcal/mol $Å^{-2}$). The electrostatic and the vdW decoupling free energies from the two sets

of simulations were compared. The electrostatic decoupling free energy is 7.57±0.2 kcal/mol from the simulation with k = 40, comparing to 7.78±0.2 kcal/mol when using the weaker restraint (Table 4). This similarity indicates that the restraint strength has little effect on the electrostatic decoupling. This is not surprising because benzamidine is likely to be confined within the trypsin binding pocket by vdW interactions at this stage. The restraint does not truly come into effect until the vdW interactions are turned off. Indeed, the van der Waals free energies differ by as much as 1.3 kcal depending on the restraint strength. On the other hand, the correction to the binding free energy due to the restraint also varies as the restraint strength increases from 20 to 40 by a similar amount (1.6 kcal/mol). Based on the above observation, we argue that the restraint affects mostly the vdW decoupling such that the correction should be applied mainly to the vdW decoupling free energy of benzamidine-trypsin. After taking these corrections into account, the van der Waals free energies of decoupling benzamidine from trypsin become -2.3 kcal/mol (averaged over values from two soft-core vdW potential at k=20) and -1.9 kcal/mol (k = 40). These quantities are fairly close to the vdW decoupling free energy of ligand-water (-2.27 and -2.42 kcal/mol, depending on the soft-core potential). In contrast, the electrostatic decoupling free energy of ligand-water and ligand-protein differ by -6.4 kcal/mol on average, which amounts to almost all of the binding free energy. Thus we conclude that the electrostatic interaction is responsible for the binding of benzamidine to trypsin.

The vdW decoupling free energy with stronger restraint (k = 40) seems to stabilize much quicker than that with k = 20. Nonetheless, it still drifts slightly over the 2 ns simulation timeframe. We have extended the simulation to 3 ns, during which the free energy value changed by -0.31 kcal/mol. Therefore, the stronger force constant may offer a quicker estimation of the binding free energy, although long simulations (~3 ns) are still necessary to obtain accurate results.

In Table 4, we summarized the three sets of the binding free energies computed using different soft-core vdW and restraint strengths. Consistency among the simulation results supports the premise that the sampling is adequate and the results are well-converged, owing to the presence of the restraint and to the fact that benzamidine is small and rigid. Our best estimate of the absolute binding free energy of benzamidine-trypsin is therefore 6.6 kcal/mol, averaged over all simulation results. The agreement between the calculated and experimental values is well within chemical accuracy.

3.3.4 Polarization effect

A unique feature of the present model is the explicit treatment of dipole polarization, which allows the electrostatics to respond to the environment, be it water or protein. It was suggested that accounting for polarization improves the transferability of a force field (Geerke, van Gunsteren, 2007), which would be critical for transferring ligand from bulk water into the protein. It is therefore of interest to examine the effect of polarization on the thermodynamics of benzamidine binding to trypsin. We turned off the dipole induction between benzamidine and its environment using free energy perturbation

to compute the "polarization free energy" in both bulk water and trypsin. Note that there is still polarization present between water-water and trypsin-water, although both water and trypsin are unable to feel the electric field due to benzamidine and vice versa. 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 (Figure 14).

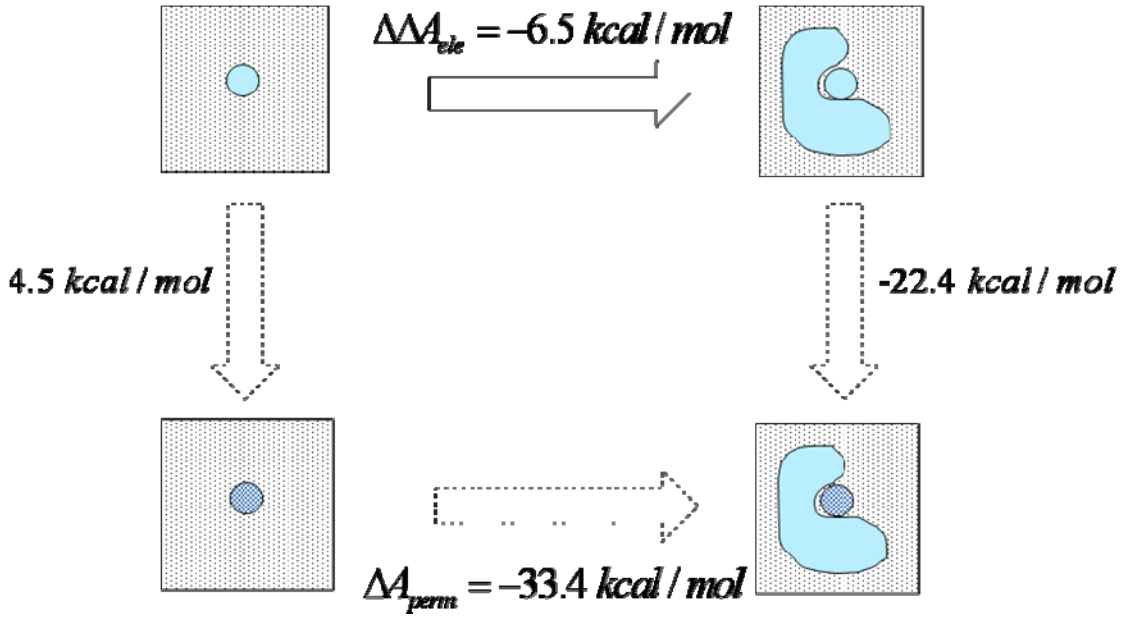


Figure 14. Polarization effect upon binding.

Not only are the magnitudes dramatically different, the sign is also opposite. While polarization seems to enhance the solvation of benzamidine in water, it weakens the association between benzamidine and trypsin. Overall, the polarization works to diminish the effect of permanent electrostatics in driving the binding of benzamidine to trypsin.

Turning on polarization leading to an increase in the energy of the protein-ligand complex may appear counterintuitive. Indeed, polarization energy always lowers the "total" system energy by making a negative contribution. However, the polarization energy of the complex became less negative when the polarization between benzamidine and trypsin-water was present. In other words, the gain in electrostatic energy due to permanent electrostatic interactions, e.g., salt-bridges, is counterbalanced by the loss in the polarization energy. The local polarization response to the association of two charged entities is to screen the electrostatic interactions, similar to the dielectric effect of water screening charge interactions. To verify this phenomenon, we computed the total dipole moment of the carboxyl group (CO_2^-) of trypsin's aspartic acid D189, which forms a salt bridge with benzamidine, before and after polarization was turned on. Consistent with our observation of energy, the dipole moment did decrease by 0.1 D when polarization was present.

In our model, polarization energies between benzamidine-water and benzamidine-trypsin differ by 27 kcal/mol. Our results agree with earlier findings that electrostatics is sensitive to local environment. Calculations by Lee et al. showed that the polarization energy in water and antibody-antigen complex varied by as much as 8 kcal/mol (Lee et al., 1992). A quantum mechanics/molecular mechanics (QM/MM) study of HIV-1 protease-inhibitor binding (Hensen et al., 2004) suggested that polarization contributed to about one-third of the total electrostatic interaction energy.

On the other hand, the artificial model we created by turning off polarization is not equivalent to fixed-charge based force fields. Previous LIE studies of trypsin-benzamidine using fixed-charge potentials reported electrostatic contribution to the binding free energy to be between -4.9 and -6.4 kcal/mol (Aqvist, 1996; Leiros et al., 2004; W. Wang, Wang, & Kollman, 1999). The values are close to -6.5 kcal/mol obtained in this study, although the electrostatic solvation energies of benzamidine in water differ from our values by -4 to -13 kcal/mol (Table 5). This suggests that it is possible for fixed-charge models to implicitly include the overall polarization effect in the binding equilibrium. Direct comparisons of polarizable and non-polarizable force fields in the free energy pathway calculations will perhaps offer further insight. It has also been discussed previously that unless divalent ions are involved (Warshel et al., 2007) or binding occurs at the protein interior (Warshel & Levitt, 1976), polarization plays a secondary role and may be absorbed by effective parameterization of fixed-charges.

3.3.5 Importance of including multipoles

One of the unique characteristics of the AMOEBA force field is that it includes higher order of electronic moments (quadrupole). One may ask how important a role multipole moment plays in the calculation of binding energy.

Based on the trajectories from simulations with full interactions, we reevaluated the electrostatic free energy change in the complex by turning off the multipoles of all the atoms within 4 Å around the benzamidine. These 110 atoms are the closest neighbors of benzamidine in the binding pocket. The lack of multipoles in the core region results in

overestimation of the electrostatic decoupling free energy by 2.19 kcal/mol, while the total is 10.06 kcal/mol compared to 7.78 kcal/mol with full electrostatic components (shown in Figure 15).

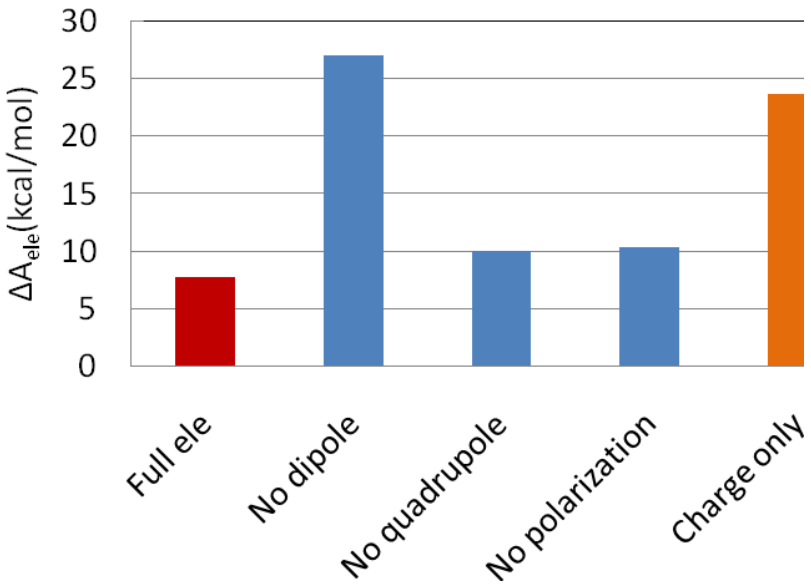


Figure 15. Electrostatic free energies in ligand-bound trypsin with different components turned off.

For the sake of comparison, we recomputed the electrostatic free energy by setting different electrostatic components to zero within the 4-Å region. It turned out that, zeroing out polarizabilities essentially has a similar result to turning off the quadrupoles with a comparable overestimation of 2.58 kcal/mol. This indicates that the higher order moments are as crucial as polarization for binding free energy calculation. Additionally, the free energy has been dramatically changed by turning off dipole parameters, regardless of whether the quadrupoles and polarizabilities are included (27.0 kcal/mol) or not (23.7 kcal/mol). Undoubtedly, dipole parameter is necessary in ligand-protein binding

free energy calculation. The contribution of dipole interactions weighs more than quadrupole and polarizability components.

We also did a series of free energy calculation by switching off electrostatic components of the outer region of the system. The scheme is described as follows. The inner region centered at the benzamidine had complete set of polarizabilities, dipoles and quadrupoles, whereas the polarizabilities/quadrupoles of all the atoms that belong to the outer region were zeroed out. The electrostatic decoupling free energy was reevaluated with the boundary at different positions, with a variable distance from the benzamidine.

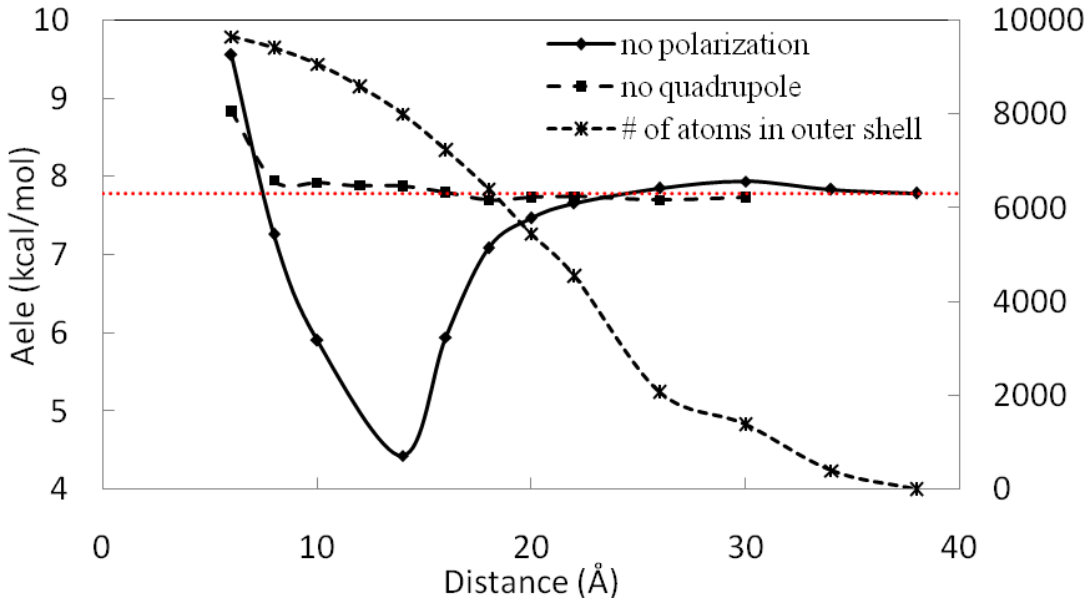


Figure 16. Electrostatic free energies of ligand-protein with zeroing out polarizabilities or quadrupoles of the outer region atoms. Solid line with diamond markers is the free energy of turning off polarization. Long dashed line with square markers represents the free energy of turning off quadrupoles. Short dashed line with star markers reprensents the number of atoms at the outer region of which polarization/quadrupole was turned off. Red dotted line is the free energy calculated with full interactions.

Figure 16 shows the electrostatic free energy change with respect to the position of the boundary. As a result, the free energy without quadrupoles of outer region converges fairly fast. It reaches the reference free energy (7.78 kcal/mol) with the boundary at 10 Å where roughly 90% of the atoms in the whole system have zero quadrupoles. On the contrary, the free energy computed with no polarizabilities in the outer region starts off with 9.56 kcal/mol at 6 Å, drops to the minimum 4.43 kcal/mol at 14 Å and finally comes back and converges gradually beyond 22 Å.

Not only does the free energy experience much more fluctuation by turning off polarizabilities of outer region atoms than turning off quadrupoles, but it also needs at least 50% atoms with full polarizabilities in the calculation to obtain a reasonable answer. In other words, quadrupole energy terms affect electrostatic calculation more locally, which is only crucial within a short range of the interaction site. However, polarization energy terms can affect atoms in a further distance than quadrupole and due to the many body effects, polarization effects can be passed on for a longer distance in the system. Therefore, ignoring multipoles of most atoms at the outer shell will barely influence the free energy calculation, while electrostatic free energy is much more sensitive to polarization which requires a minimum of half of the atoms in the system with full polarizabilities.

3.4 CONCLUSION

A polarizable force field was applied to compute the binding affinity of a positively charged ligand to protein. Parameters were either directly derived from QM or

transferred from the protein force field without modification or recalibration. Molecular dynamics simulations were performed with double decoupling of benzamidine from both water and trypsin binding site with free energy perturbation. Different thermodynamic sampling paths were employed, by varying the softness of vdW potential and the restraint strength, and the resulting free energies were consistent. The computed absolute binding free energy is well within experimental accuracy.

Our results indicate that the electrostatics is the driving force for benzamidine binding to trypsin. We have further evaluated the role of polarization in binding by "turning off" the dipole induction between the ligand and its environment. It was found that polarization response varies drastically depending on the nature of the environment, and its contribution to the decoupling free energy does not simply cancel between water and protein. As a result of this finding, we believe that it is critical to treat polarization explicitly in order to achieve chemical accuracy in predicting binding affinity of charged systems. Higher order of moments are also important in the calculation, without which the electrostatic decoupling free energy will be overestimated.

We have compared the results by turning off polarizabilities or quadrupoles in the outer region. By moving the position of the boundary between inner and outer region, we found that the electrostatic free energy is more sensitive to polarizabilities than quadrupoles. In order to acquire an acceptable free energy, at least 50% of the atoms around the ligand must have full polarizabilities, whereas 10% of the atoms in the vicinity of the ligand with complete quadrupoles are sufficient.

In summary, electrostatics and polarization play important roles in molecular recognition and need to be accounted for in quantitative modeling. Our study demonstrates that chemical accuracy in predicting protein-ligand binding free energy can be achieved with a polarizable potential energy function when adequate sampling is possible.

4 Calculation of relative ligand-protein binding free energies

4.1 INTRODUCTION

Molecular recognition plays a key role in many biomolecular processes such as enzyme catalysis, intracellular signaling and protein conformational switching. The modern drug discovery process begins with an identification of small molecules that interact with specific targets such as receptors, enzymes, hormones, ion channels and other macromolecules with high affinities. Physical-based molecular modeling has been sought after as the potential technique to accelerate and facilitate the drug discovery process. From rapid empirical docking to sophisticated quantum mechanical (QM) *ab initio* theory, from explicit-water molecular dynamics simulations to implicit solvent continuum approaches, a range of computational methods have been utilized to determine the binding affinity of small molecules to macromolecular targets (Brandsdal et al., 2003a; Gilson & Zhou, 2007; Gohlke & Klebe, 2002; Jorgensen, 2004; Kollman et al., 2000). Although great progress has been made in various fronts, including the rigorous treatment of long-range electrostatic interactions (Darden, 2008; Sagui & Darden, 1999) and the sophisticated sampling algorithms for free energy calculations (Chipot & Pohorille, 2007), there remain challenges in using molecular modeling to make reliable predictions of ligand binding affinities. Two immediate obstacles are limited sampling of protein-ligand-water interaction and accuracy of the potential energy function describing the atomic interactions.

In the previous chapter we have utilized a polarizable atomic multipole-based potential to calculate the absolute binding free energy of benzamidine to trypsin (Jiao, Golubkov, Darden, & Ren, 2008). We concluded that the electronic polarization renders very different effect in protein and water environments and should be taken into account explicitly for accurate free energy evaluation. Unlike the absolute binding free energy, relative binding free energy is more likely to be predicted accurately due to systematic error cancellation. There have been extensive studies of relative binding affinity using an array of explicit and continuum based methods (Gilson & Zhou, 2007; Gohlke & Klebe, 2002). Several have reported good agreements with experiment (Deng & Roux, 2006; Fujitani, Tanida, Ito, Shirts et al., 2005; J. Y. Wang & Roux, 2005). However, consistent prediction of relative binding affinity from molecular simulations is not yet robust or fully validated due to the relative computational expense compared to docking like approaches (Gilson & Zhou, 2007). Further work on a wide range of molecular systems is needed to gain a firm understanding of the capability of molecular modeling to rank ligand affinity *in silico*.

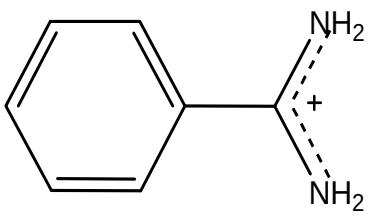
Trypsin-benzamidine has been a prototypical system for evaluating modeling techniques. A good number of ligands that inhibit trypsin and other serine proteases have been investigated via computer simulations. Free energy perturbation (FEP) (Essex, Severance, TiradoRives, & Jorgensen, 1997) and thermodynamic integration (TI) (Ota et al., 1999; Talhout & Engberts, 2004) with explicit solvent simulations, QM (Grater, Schwarzl, Dejaegere, Fischer, & Smith, 2005) and MM-based implicit solvent (Resat,

Marrone, & McCammon, 1997) and other approximated approaches (Radmer & Kollman, 1998) have all been attempted. Several of the studies have focused on a series of ligands of similar physicochemical properties, e.g. benzamidine derivatives with various *p*-substituted alkane groups. Essex and Jorgensen have suggested that, although more hydrogen bonds are found in more tightly bound ligand, the bulk solvation effect dominates the binding affinity, i.e. more polar ligands would be weaker inhibitors due to better solvation in bulk water (Essex et al., 1997). The underestimation of binding affinity for benzamidine has been attributed to a deficiency in the partial charges used. The opinion of bulk-solvation domination has been echoed by others (Talhout & Engberts, 2001). According to Grater *et al.,*(Grater et al., 2005) the van der Waals energy is the major energy term that favors binding to trypsin. A recent study shows that the relative binding affinity results obtained with a polarizable force field are much more correlated with experimental data than a non-polarizable force field, suggesting the inadequacy of the latter for charged systems (Khoruzhii et al., 2008). In this work, we present a study of a series of ligands with different aromatic and charged groups using a polarizable potential for the entire system of protein, ligand and water. Aside from free energy calculations, we have also examined the charge distribution in the ligands and the protein-ligand-water interactions in atomic detail.

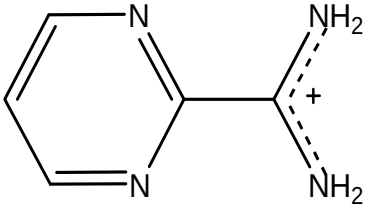
4.2 METHOD

4.2.1 Ligands of trypsin

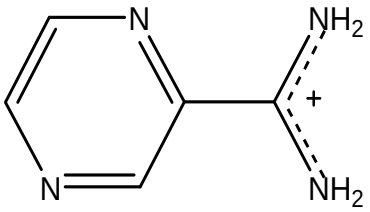
The signature aspartic acid residue located at the binding site of trypsin provides strong electrostatic interactions with counter-charged ligands. All five ligands we have investigated in this work contain a positively charged functional group (Figure 17).



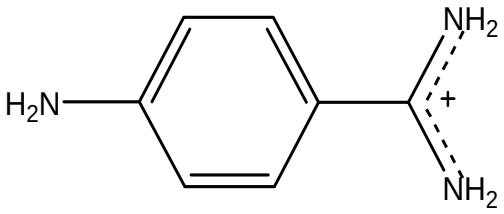
A



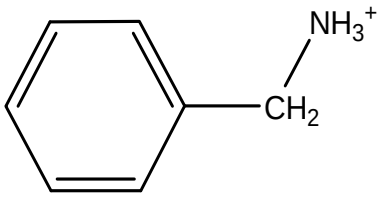
B



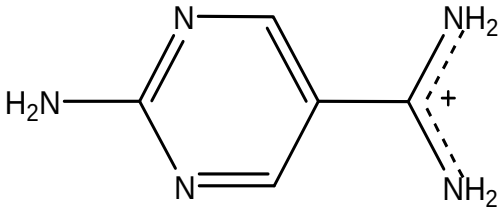
C



D



E



F

Figure 17. Chemical structures of trypsin ligands studied: A. benzamidine; B. 1,3-diazamidine; C. 1,4-diazamidine; D. 4-amino-benzamidine; E. Benzylamine; F. 4-amino-diazamidine.

Benzamidine (ligand A), consists of a hydrophobic phenyl ring and a positively charged amidinium group that forms a salt bridge with the aspartic acid. Ligand B and C are similar to benzamidine except that the phenyl ring is replaced by a 1,3-diazine (or pyrimidine) and 1,4-diazine (or pyrazine), respectively. Ligand D, 4-amino-benzamidine, contains a NH_2 substitution group at the 4 position of the phenyl ring. In ligand E, a protonated amine replaces the amidinium group. Ligand F is a hybrid ligand of B and E, 1,3-diazamidine plus a amino at 4' position of the ring. We picked this group of ligands because they all are analogs of benzamidine which we investigated already. It is interesting to see if our potential model can capture the binding free energy changes due to the (1) mutation, (2) elongation, or (3) more flexible charged group. The experimental or other computational binding free energies are available for benchmark.

4.2.2 Force field parameterization

Molecular mechanics simulations were performed using a polarizable force field for the entire system, including ligand, water and trypsin. The electrostatic interaction is represented by permanent atomic charges, dipoles and quadrupoles, plus a polarization effect via atomic induced dipoles. The model has been introduced previously for water (P. Y. Ren & Ponder, 2003), ions (Grossfield et al., 2003; Jiao et al., 2006) and dipeptides (P. Y. Ren & Ponder, 2002). The force field was recently applied to compute the absolute binding free energy of benzamidine to trypsin (Jiao et al., 2008). The parameters for

water and protein are available with the TINKER molecular modeling package (Ponder, 2006). Parameterization for the new ligands in this work is described as follows.

The structure of each ligand was optimized quantum mechanically at the level of HF/6-31G* using Gaussian 03 (Frisch, 2003). A single point energy calculation was performed subsequently at the MP2/6-311++G(2d,2p) level to compute the molecular dipole moment and the density matrix. The electrostatic parameters, including monopole, dipole and quadrupole moments, were derived from the density matrix using GDMA v2.2 (Stone, 2005). The hydrogen atomic radius parameter was set to 0.31.

The van der Waals (vdW), bond, angle, and atomic polarizability parameters of the ligands were transferred from the AMOEBA potential (amoebapro.prm) available in TINKER. The relevant parameters of the amidinium group were taken from the guanidinium group of arginine. The equilibrium bond and angle values were adjusted to match the geometry obtained from force field and QM optimizations.

Torsional parameters were obtained by fitting to the QM calculation. This is done at the last step after all the other parameters are defined. The structure of the ligand with the certain torsion at different angle values (ranging from 0 degree to 360 degrees) was optimized by Gaussian. Single point energies were calculated at MP2/6-311++G** level and hence the torsional profile in terms of torsion angles. The same calculation was then carried out by TINKER with the particular torsion parameters set to zero. The three fold Fourier series were then fit to the torsional energy difference between QM and MM calculation. For ligand A, B, and C, the only rotatable dihedral angle is the one that links

the aromatic ring and the amidinium group which is a partial double bond. We adopted a generic torsional energy term, E_{tor} = 2.70*(1-cos2φ) kcal/mol, for all three ligands. The same torsional parameters were applied to the bond between the 4-amino group and the phenyl ring in ligand D. The bonds between the phenyl ring and the amine group of ligand E are single bonds in nature, and the torsional contribution is insignificant to the overall rotational energy barrier (E_{tor} = 0.064*(1-cos2φ)+0.605*(1-cos3φ) kcal/mol).

The molecular dipole moment vector was computed for each ligand in gas-phase using the standard orientation from QM optimization. To calculate the ligand dipole moments in bulk water and solvated complex, the averaged atomic induced-dipole moments were collected from the molecular dynamics simulations. The permanent and induced multipoles were then applied to the same QM geometry to compute the ligand dipole moments. All the structure files and ligand parameters used in this chapter are included in the Appendix B.

4.2.3 Free energy perturbation

Free energy perturbation (FEP)(Jorgensen & Ravimohan, 1985) was utilized to compute the free energy change between two states. Relative binding free energy was calculated for ligands B through E by perturbing each into benzamidine (ligand A) in both neat water and the protein complex while only ligand F is perturbed from ligand D (Figure 18).

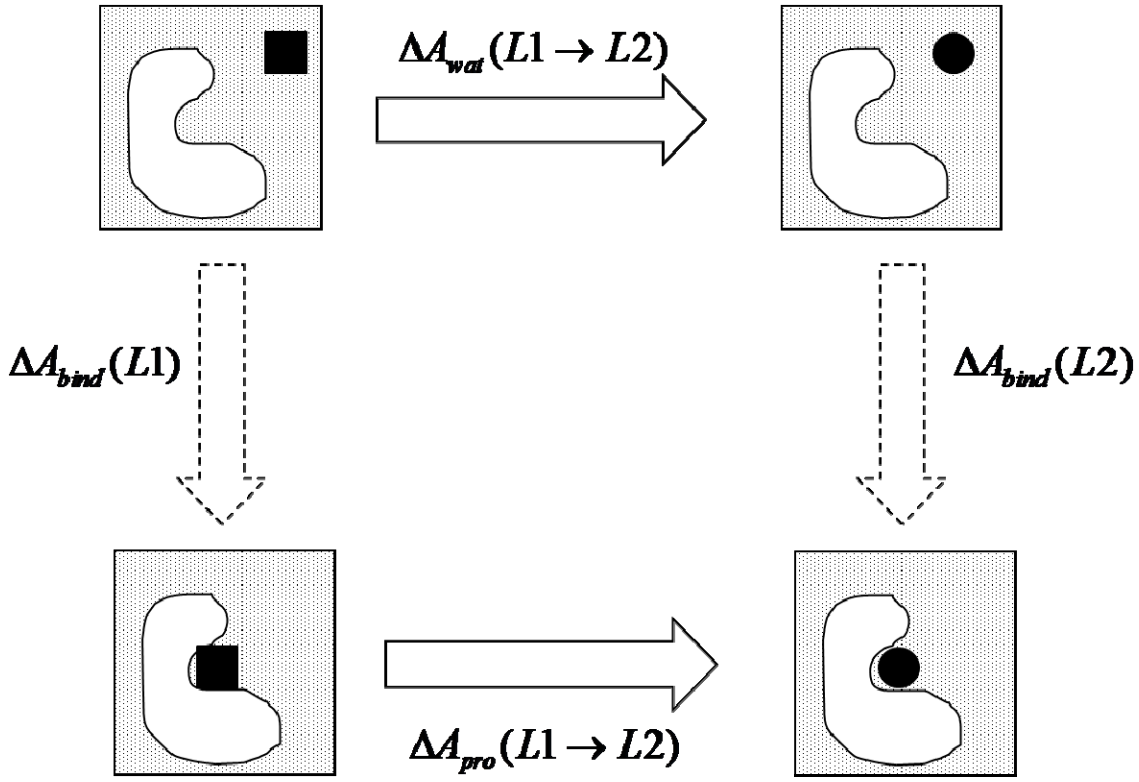


Figure 18. Thermodynamic scheme to calculate relative binding free energy. Squares and rounds represent ligand 1 and ligand 2.

The relative binding free energy between two ligands was computed as:

$$\Delta\Delta A_{bind}(L1 \rightarrow L2) = \Delta A_{pro}(L1 \rightarrow L2) - \Delta A_{wat}(L1 \rightarrow L2) \qquad (18)$$

The number of steps and perturbation path are determined by the structural variation of the two ligands. The perturbation involves parameters of one ligand gradually being changed to these of the other ligand following certain path by linearly interpolating the ligand electrostatic and vdW parameters between the two end states. Mutation between two types of atoms is the most straightforward scenario, where the full set of parameters of one type of particular atoms (including electrostatic and van der Waals parameters) were changed linearly to parameters of another type of atoms, e.g. the mutation of the

nitrogen atoms of 1,3-diazamidine to the carbon atoms of benzamidine. For these mutations that atoms are to be grown out, van der Waals parameters should be changed ahead of electrostatic parameters. In this way, the van der Waals interaction can prevent the new-born atoms sitting on top of other atoms (ligand A → ligand D). For the same reason, van der Waals parameters of atoms that are being annihilated should be turned off after electrostatic parameters are turned off. In case of ligand A → ligand E, both growth and annihilation occur. Not only nonbonded terms, such as electrostatics and van der Waals, but also internal valence terms including bond, angle torsion and out-of-band, need to be changed. Meantime, the vdW parameters of the growing atoms should be the first to change while those of the disappearing atoms should be the last to be switched off due to the same concern above (ligand A → ligand D). In the annihilation of vdW interactions, the soft-core approach was used to turn off the interactions between the dummy atoms and all other atoms in the system (Beutler et al., 1994). Fewer steps were required for relatively minor structural variantion, for instance, it took 12 steps for ligand A → ligand B/C, whereas more steps were needed for more complicated perturbation (16 steps for ligand A → ligand E).

The free energies between two neighboring states were calculated using the Bennett Acceptance Ratio estimator.(Bennett, 1976) The statistical errors in the free energy change between two steps were computed as well. The total statistical error in the solvation free energy in bulk water or complex was computed as the sum of the errors from individual perturbation steps.

4.2.4 Explicit solvent molecular dynamics simulations

At each perturbation step, molecular dynamics simulations of ligand in bulk water and protein were performed, respectively. The initial systems were prepared using TINKER. The benzamidine-trypsin crystal structure (1BTY) (Katz, Finermoore, Mortezaei, Rich, & Stroud, 1995) was used as a starting structure to generate new structures for the other ligands. The rule of thumb is to pick the ligand with more atoms as the starting point, so coordinates for all the atoms are defined. Structure of trypsin complexed with ligand D was created by placing ligand D in trypsin binding pocket by superimposing the phenyl ring onto that of benzamidine. The case of ligand E is a bit more complex. The amine group NH_3 was being mutated to one of the NH_2 groups of the amidinium, while the other NH_2 group was growing out from one of the hydrogen atoms attached to the SP3 carbon atom. To be more specific, one of the hydrogen atoms of the amine was turned into a dummy atom. Moreover, we attached two dummy atoms to the SP3 carbon hydrogen. When the hydrogen atom later became a nitrogen atom, the two dummy atoms bonded to the hydrogen were growing into hydrogen atoms eventually. We wound up constructing a hybrid ligand A/E and placing it in the trypsin at the active site with the same orientation as the benzamidine in the crystal complex structure (Figure 19).

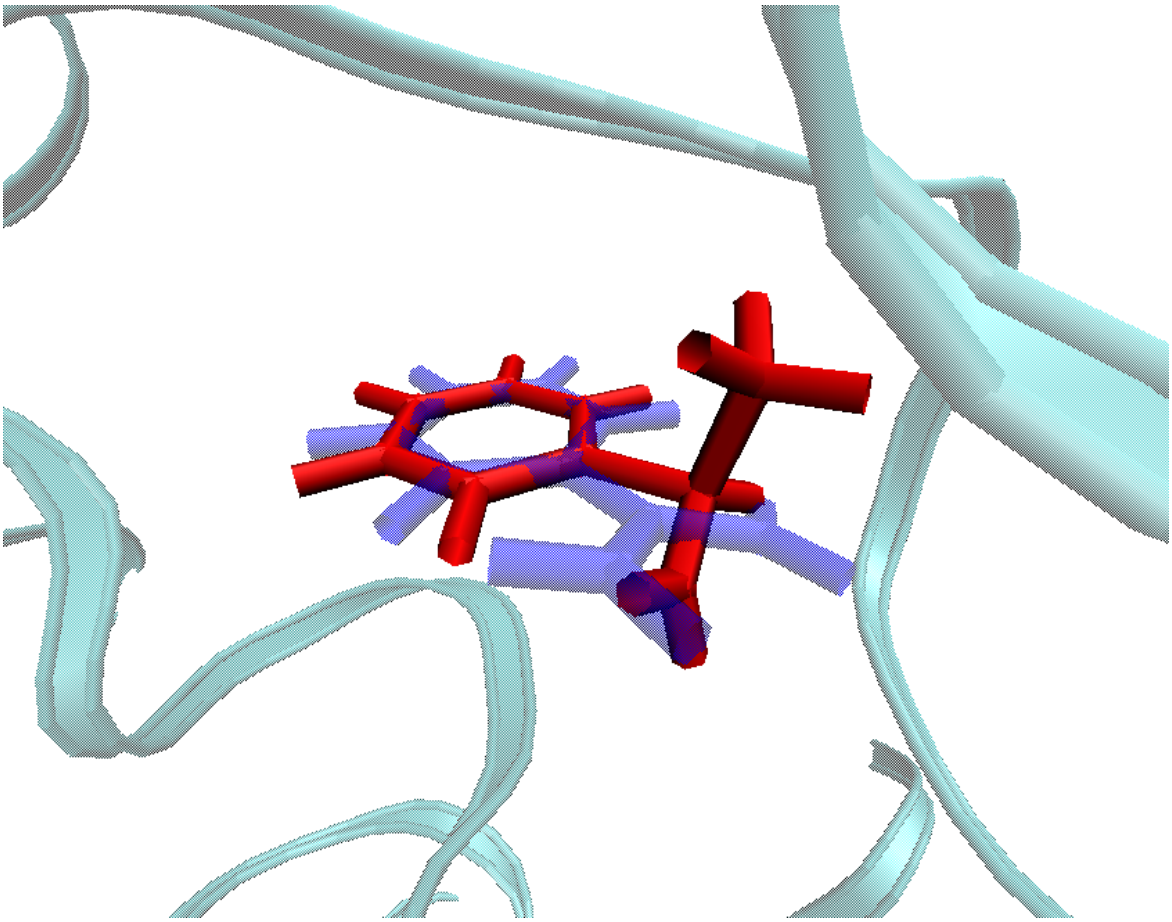


Figure 19. Superposition of hybrid ligand A/E and benzamidine in trypsin crystal structure. The protein structure is shown in format of ribbon (partial). The original benzamidine is shown in transparent blue. The hybrid ligand of A and E is shown in red.

Based on the crystal structure, HIS40 and HIS91 are deprotonated at εN while HIS57 is deprotonated at εN. The protein-ligand complex was placed in a periodic octahedral water box. For ligands A, B and C, we continued to use our previous system of 2222 water molecules (the containing cubic box is 51Å on each side) (Jiao et al., 2008). For ligands D, E and F, we adopted a bigger octahedron box with 4515 water molecules and 58 Å on each side. An internal water molecule is present in the crystal

structure, hydrogen-bonded to the amidinium group of benzamidine (Figure 20a). In our system construction, the trypsin-ligand complexes were soaked in a water box and internal water molecules were added into the binding site where space allowed, without utilizing the information on crystal water. TINKER placed one or two water molecules near the Asp189-amidinium/amine site as shown in Figure 20b through Figure 20e.

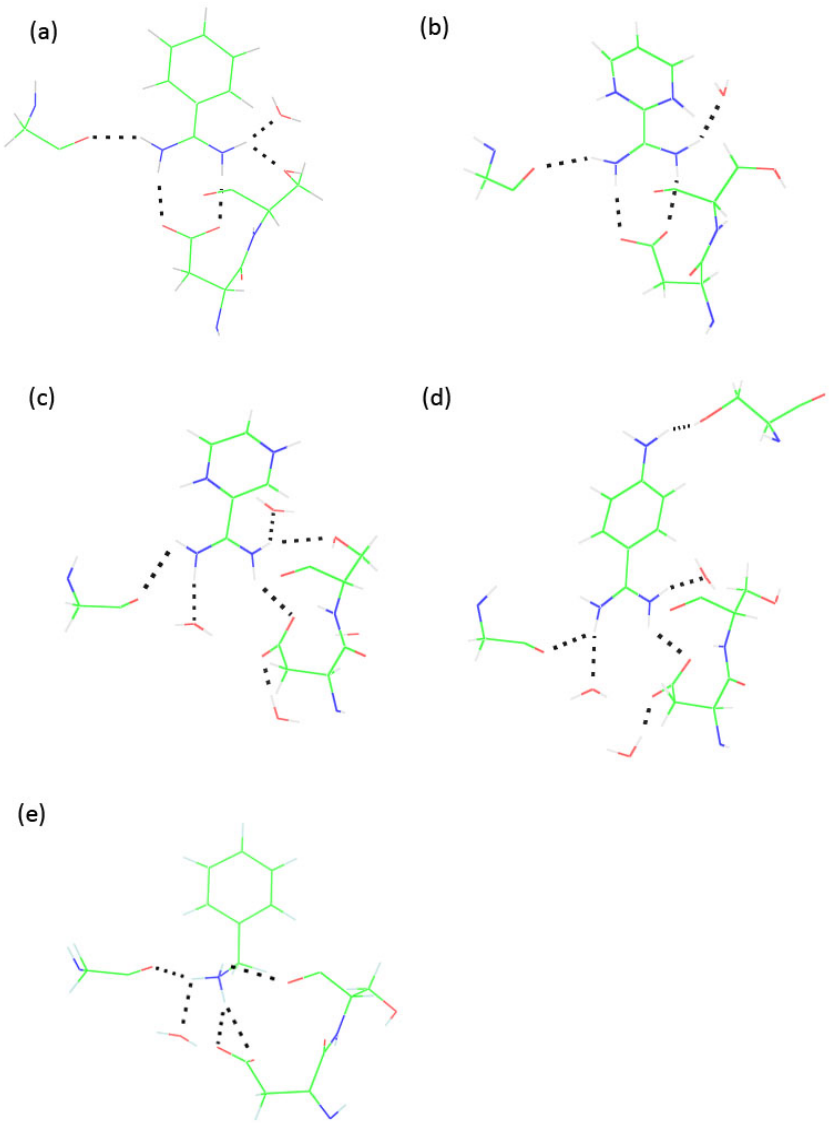


Figure 20. Intermolecular hydrogen bonding structures between ligands and trypsin at the binding site. (a) Crystal structure of trypsin in complex with ligand A (PDBID 1BTY). (b) to (e) are representative snapshots from explicit-water molecular dynamics simulations of trypsin with ligand B to E, respectively.

All production MD simulations were performed along the perturbation pathways described above using PMEMD in AMBER v9 (Case et al., 2006). We were able to

achieve more than 200 ps per day with an 8-core 2.8 GHz Xeon computer for the 58 Å system, which is a speedup of ~4x over a single core. A 100 ps NPT dynamics simulation was first performed to equilibrate the system. The resulting configuration was then subject to the NVT simulations with the density fixed at the NPT-average. The same procedure was applied to prepare the ligand-water systems which were also octahedron boxes containing about 400 water molecules. NVT dynamics simulations of 1 to 2 ns were performed on the trypsin-ligand systems at each perturbation step as required for statistical convergence, whereas 0.5 ns simulations were conducted for all ligand-water systems. A 1 fs time step was used. Atomic coordinates of the simulation system were saved every 500 fs. The temperature was maintained at 298 K using the Berendsen thermostat (Berendsen et al., 1984a). The vdW cutoff was set to 9 Å, with a long-tail correction included. Particle Mesh Ewald (PME) was used to treat the electrostatic interactions, with a real-space cutoff of 7.0 Å. To speed up the simulation, the induced dipoles were iterated until the root-mean-square change was below 0.05 D per atom. In the post-MD free energy analysis with the Bennett acceptance ratio (Bennett, 1976), we used a tighter convergence criterion, 10^{-5} D per atom.

4.3 RESULTS AND DISCUSSION

4.3.1 Relative binding free energy

We have performed a series of molecular dynamics simulations to perturb the ligand from benzamidine to another in both bulk water and in the solvated complex. The polarizable potential is applied to the entire system in all simulations. From these

simulations, we were able to compute the relative binding free energy for each ligand), as well as the absolute binding free energy based on the previously calculated value for benzamidine (Jiao et al., 2008). The results are summarized in Table 6.

Table 6. Relative and absolute binding free energies from the explicit solvent FEP simulations. All relative values were computed with respect to benzamidine (ligand A). Statistical errors are given in the parenthesis.

	A	B	C	D	E	F
ΔA_{wat}	0	-25.51(0.5)	-10.55(0.5)	-1.46(0.5)	-20.93(0.7)	-19.57
ΔA_{pro}	0	-23.76(0.4)	-8.74(0.3)	-1.74(0.4)	-19.60(0.6)	-17.40
$\Delta\Delta A_{bind}$	0	1.75	1.81	-0.28	1.33	2.17[7]
ΔA_{bind}	-6.7[8]	-5.0	-4.9	-7.0	-5.4	-4.8
Experiment	-6.3,[1] -6.4,[2] -7.3[3]	-4.7[1]	-4.8[1]	-7.0,[2] -7.2[4]	-3.8,[5] -4.7[6]	-5.0[4]
Other computation	-6.4[1]	-7.0[1]	-6.5[1]	-6.1[4]	-4.2[5] , -2.4[6]	-4.7[4]

1. Ref (Grater et al., 2005). Calculation using PB/SA combined with QM/MM.
2. Ref (Talhout & Engberts, 2001).
3. Ref (Katz et al., 2001).
4. Ref (Schwarzl, Tschopp, Smith, & Fischer, 2002). Calculation using PB/SA.
5. Ref (Ota et al., 1999). Non-Boltzmann thermodynamic integration (NBTI) MD simulations.
6. Ref (Leiros et al., 2004). Linear interaction energy (LIE).
7. Relative binding free energy to ligand D.
8. The benzamidine binding free energy was calculated previously using a harmonic restrain (k=20 kcal/mol) and soft-core vdW function (0.5/4). An additional correction of -0.37 kcal/mol is added to account for the removal of the restraint from the fully interacting protein-ligand.

The experimental binding free energies are based on inhibition constants determined by spectrophotometry or isothermal titration calorimetry under various assay conditions. The existence of multiple experimental values for the same ligand indicates that the experimental uncertainty is almost 1 kcal/mol in energy or one order of

magnitude in binding affinity. The relative affinity from the same source should be more reliable although for ligand D we find two sets of values that differ by 0.2 kcal/mol.

The calculated relative binding free energies are in excellent agreement with experimental measurements. We are satisfied that in all cases the sign of the binding affinity change has been predicted correctly. All relative solvation free energies of ligand B through E are negative, indicating these ligands are all better solvated than benzamidine (ligand A) in bulk as well as in trypsin binding site. Similarly, ligand F is solvated more favorably than ligand D in both water and protein. The free energy changes in both environments are fairly significant for B, C and E, on the order of -10 to -20 kcal/mol, when the phenyl ring is replaced by a diazine or the charged amidinium by an amine. This is again confirmed by the free energy change of ligand D → ligand E. However, the change in bulk water is mostly compensated by that in the complex, leading to a net decrease in the binding free energy of 1~2 kcal/mol. Thus, ligands B, C, and E are predicted to be somewhat less potent inhibitors than benzamidine, and ligand F less potent inhibitor than ligand D, in agreement with experiment. As for ligand D with an extra NH_2 substituent on the phenyl ring, the free energy changes in water and in the protein complex are relatively small: -1.46 and -1.74 kcal/mol, respectively. As a result, the binding affinity of ligand D increased slightly over that of benzamidine. However, the calculated magnitude of change is less significant than the experimental value.

Deng et al. (Deng & Roux, 2006) reported that the repulsive and dispersive interaction contribute significantly to the binding free energy from WCA decomposition,

while the electrostatic interaction is slightly unfavorable. However, these computations were limited to nonpolar ligands such as benzene, toluene and phenol. In contrast, these 6 ligands binding to trypsin is mainly determined by the electrostatic contributions ranging from -4.95 to -7.97 kcal/mol, while the contributions from other interactions are only from -0.50 to 2.60 kcal/mol (Figure 21). Thus the electrostatic interaction is indicated as the driving force of the binding of these highly charged ligands to trypsin.

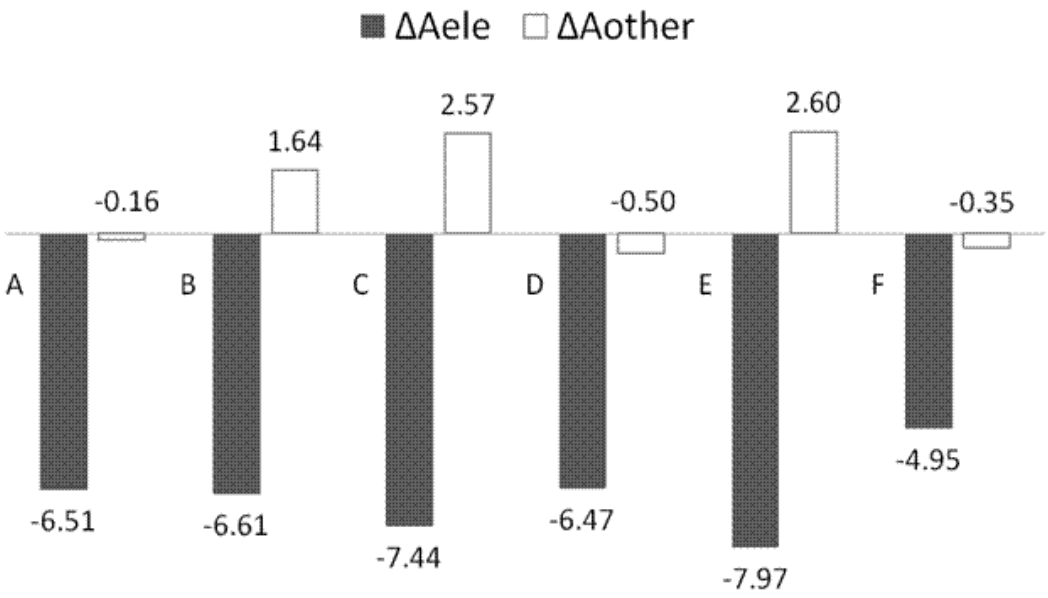


Figure 21. Decomposition of binding free energies (kcal/mol). Grey column is the electrostatic free energy and white column is the contribution of other free energy components including vdW and geometry.

Nontheless, the deciding factor for the binding selectivity of the ligands, i.e. the relative binding affinity, is not necessarily electrostatics. We found van der Waals interaction is the main cause for the decrease in the binding affinity from ligand A (benzamidine) to ligand B (1,3-diazamidine). Similarly, we found that for ligand C (1,4-

diazamidine), the vdW interaction remains the dominant factor in the relative binding affinity. By contrast, the electrostatics amounts to more than 60% of the change in the binding free energy for both ligands D (4-amino-benzamidine) and E (benzylamine). The separation of electrostatics and vdW contribution is somewhat artificial depending on the choice of the specific perturbation path. Nonetheless, the results suggest that the net change from benzamidine to ligand B or C, where two aromatic CH groups are replaced by two N atoms, is mostly a size effect as the electrostatic contribution to ΔA_{sol} compensate between the water and the protein environments.

4.3.2 Molecular dipole moments of the ligands

Electrostatic interactions are important factors to the trypsin-ligand recognition as the presence of the charged group is crucial (Talhout & Engberts, 2001). While the aromatic benzene is commonly considered as a "hydrophobic" group, the accurate account for hydrophobicity also depends on the details of electrostatic interaction with water and other surrounding atoms. In a previous study, we evaluated the effect of polarization in binding by switching off dipole induction between the benzamidine and its environment. We concluded that polarization actually worked to offset the permanent electrostatic attraction between benzamidine and trypsin. Here we have calculated the dipole moment of each ligand in gas phase, bulk water, and protein complex, to characterize the ligand charge distributions Table 7.

Table 7. Molecular dipole moments (Debye) in gas, in bulk water and protein-ligand complex from quantum mechanics *ab initio* calculations at MP2/6-311++G(2d,2p) and molecular dynamics simulations using the polarizable force field. For each ligand, the

dipole moments were calculated using the QM geometry in the inertia frames with the origin at the center of mass.

		Total	D_X	D_Y	D_Z
A	Gas phase (QM)	6.00	-6.00	0.00	0.00
	Gas phase	6.21	-6.21	0.00	0.00
	Water	6.67	-6.67	0.00	0.00
	Protein	6.88	-6.88	0.00	0.00
B	Gas phase (QM)	3.75	-3.75	0.00	0.00
	Gas phase	3.73	-3.73	0.00	0.00
	Water	3.83	-3.83	0.00	0.00
	Protein	3.98	-3.98	0.03	0.00
C	Gas phase (QM)	6.24	-6.23	0.00	0.00
	Gas phase	6.61	-6.60	0.39	0.00
	Water	7.11	-7.10	0.34	0.00
	Protein	7.17	-7.16	0.43	0.00
D	Gas phase (QM)	4.21	4.21	0.00	0.00
	Gas phase	4.37	4.37	0.00	0.00
	Water	4.79	4.79	0.00	0.00
	Protein	5.08	5.08	0.00	0.00
E	Gas phase (QM)	8.93	8.66	0.00	2.19
	Gas phase	9.50	9.15	0.00	2.57
	Water	10.28	9.93	0.02	2.66
	Protein	10.80	10.49	-0.15	2.59
F	Gas phase (QM)	6.32	-6.32	-0.00	0.00
	Gas phase	6.56	-6.56	0.00	0.00
	Water	6.80	6.80	0.00	0.00
	Protein	7.27	7.27	0.01	0.00

The molecular dipole moments computed from polarizable atomic multipoles, which have been derived from QM *ab initio* calculation, in principle reproduce the *ab initio* dipole moments exactly. The discrepancy between the gas phase dipole moments from QM and from force field calculations is due to the averaging of atomic multipoles over symmetric atoms, such as the hydrogen atoms in amine and amidinium groups. The

averaging is for the sake of simplicity, but is also necessary as we are unable to distinguish these atoms individually in our simulations. Electronic polarization in bulk water and in protein complex leads to an increase in the molecular dipole of up to 10%, with the protein environment consistently showing a greater effect than the bulk water environment. Ligand B (1,3-dizamidine) is the least affected by induction. Note that the polarization effect on the ligand is only a small fraction of that in the whole system; the protein is significantly polarized by the ligand as we have discussed previously (Jiao et al., 2008).

When a series of similar ligands are considered, it has been suggested by Essex (Essex et al., 1997) and Talhout (Talhout & Engberts, 2001)that there is a correlation between the molecular polarity and the binding affinity. They argued that the more polar ligand is better solvated in water and therefore has lower affinity binding to trypsin. In our calculation, there is, however, no evident correlation between the molecular dipole and binding affinity. The scattering plot of binding affinities and ligand dipole moments in Figure 22 does not imply any of such correlation, with a poor R square value of 0.026.

In changing ligand A to B, the phenyl ring is mutated into a pyrimidine. The two N atoms introduced in the ring in place of the two CH groups perturbed the charge distribution significantly (as evident by the 50% decrease in the dipole moment). Nonetheless, as seen from the atomic multipole parameters for both ligands, the perturbation is fairly “local”, restricted to the nitrogen atoms themselves and the carbon atoms immediately bonded to the nitrogen atoms. The atomic charges, dipoles and

quadrupoles of the amidinium group are essentially invariant between the two ligands. In changing ligand A to C, on the other hand, due to the broken symmetry in the pyrazine (or 1,4-diazine), the effect of the two nitrogen atoms cancels out and leaves the molecular dipole moment similar to that of benzamidine. In ligand D, the 4-amino substitution group donates π-electrons to the aromatic ring which reduces the molecular dipole moment relative to benzamidine. The amine group of ligand E (benzyl amine) causes a significant dipole moment (D_Z) out of the plane of the phenyl ring.

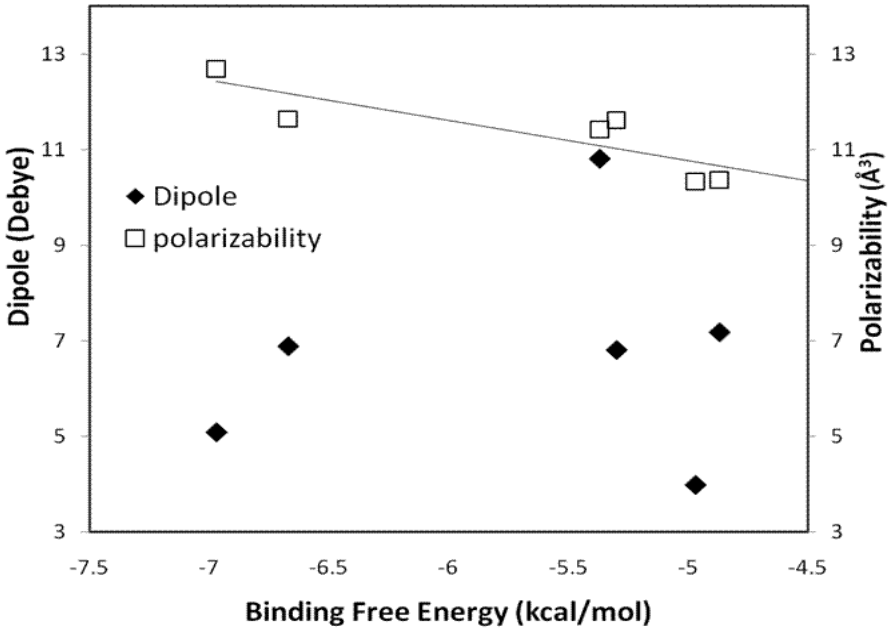


Figure 22. Correlation between dipole/polarizability of the ligands and binding free energy. Molecular dipole moments are in black diamond while polarizabilities are in open squares.

4.3.3 Structural analysis from explicit solvent simulations

Among the five ligands investigated, there is only one X-ray crystal structure available for benzamidine-trypsin (1BTY) (Katz et al., 1995). In the crystal structure, the surrounding residues and water molecules form specific interactions with the amidinium group of the benzamidine. The negatively charged Asp189 residue forms a salt bridge with the positively charged amidinium group by double hydrogen bonding with the two nitrogen atoms. At the same time, Gly219 carbonyl O is hydrogen-bonded to one nitrogen atom of the amidinium group while on the other side both Ser190 O and a water molecule form a bifurcated hydrogen bond with the other nitrogen atom. The thermodynamics of binding of *p*-substituted benzamidines to trypsin was investigated experimentally by Talhout *et al* (Talhout & Engberts, 2001; Talhout et al., 2003). It was suggested that both the hydrogen-bonding amidinium group and the hydrophobic phenyl ring of the benzamidine contributed to trypsin binding, with the former enthalpically favorable and the latter entropically favorable.

There is one internal water molecule in the proximity of the salt bridge between the benzamidine amidinium group and the Asp189 according to the crystal structure. The internal water is likely to be critical in stabilizing the binding complex. However, detailed information on internal water molecules is not always available for the inhibitors of interest, as with the ligands B through F in this study. Therefore in our models, the internal water molecules have been added into the binding site, after inhibitor is in place, based on the space availability. On average, there are more water molecules added into the complex than observed in the trypsin-benzamidine crystal structure. During the

simulation, one water molecule quickly moved to the location where the crystal water interacts with the amidinium together with Ser190, except for benzyl amine as discussed below. This water molecule formed a stable hydrogen bond with one NH_2 group of the amidinium in ligands A through D (Figure 23a).

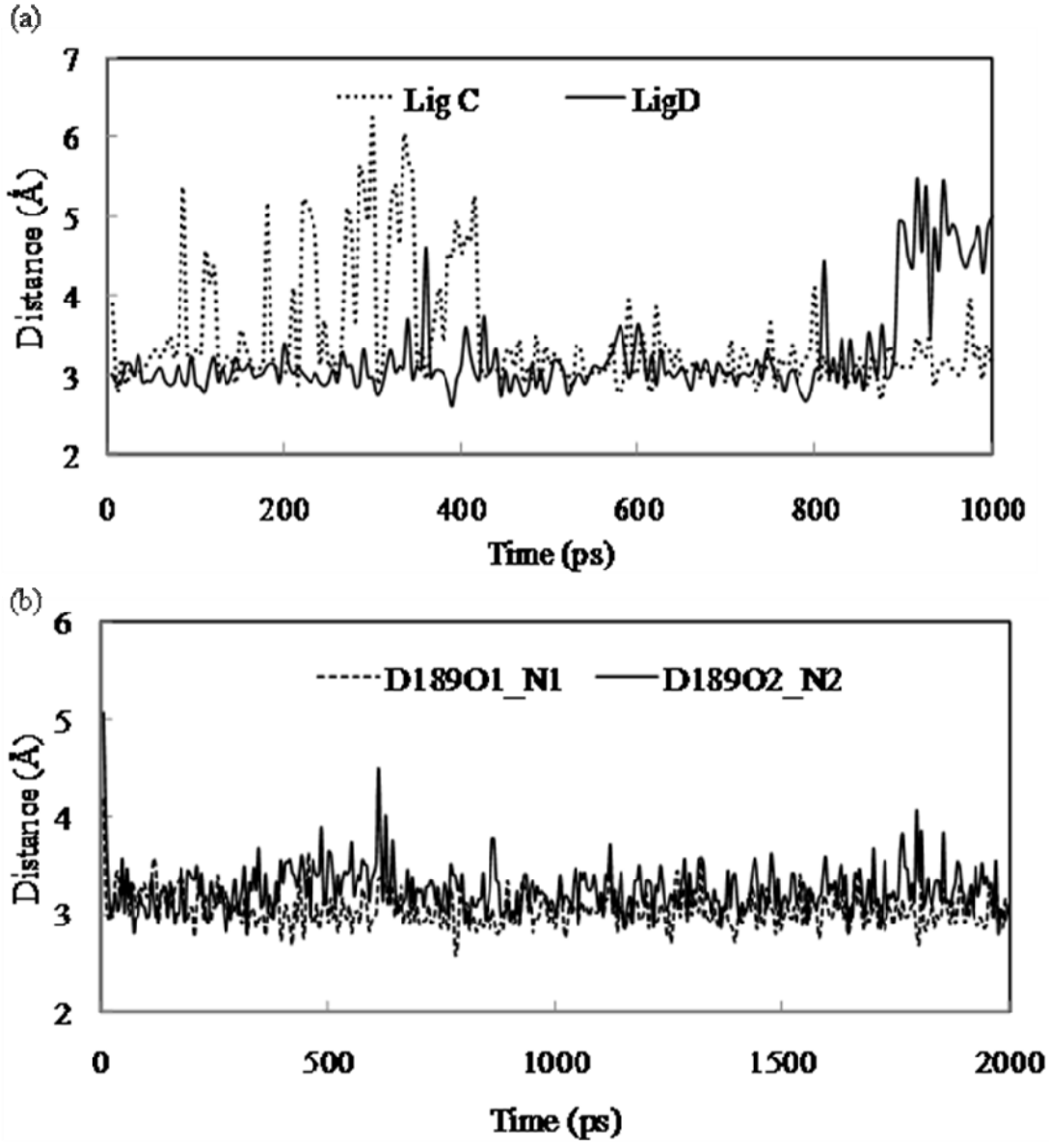


Figure 23. Evolution of the hydrogen bond distances between the ligand and the surroundings: (a) water hydrogen bonding to amidinium N2 from simulations of ligand C and D. A similar water molecule is present in the crystal structure of trypsin-benzamidine; (b) ligand B hydrogen bonding to the Asp189 residue in the binding pocket.

The double hydrogen bonding between Asp189 and amidinium was present in trypsin-ligand B complex throughout the entire simulation (Figure 20b and Figure 23b). In the case of ligand C or D, only one hydrogen bond between Asp189 (Oδ1) and the ligand (N2) was observed. The other initial hydrogen bond between Asp (Oδ2) and ligand (N1) was eventually replaced by a water molecule which was introduced into the pocket during system construction (Figure 20 c and d). The Asp Oδ2 that became free, however, bonded to another internal water molecule introduced during the soaking. Neither of the two water molecules is present in the crystal structure of trypsin-benzamidine (1BTY).

The hydrogen bond between Gly219 and the ligands is well conserved in all the simulations and the average bond distance (2.94 Å) is in good agreement with that in crystal structure (2.89 Å). By contrast, great variability was observed in the hydrogen bonding between Ser190 and the ligands. The hydrogen bond between the ligand and the Ser190 Oγ, which is present in crystal structure, was only observed in the simulations of trypsin-ligand A and trypsin-ligand C. For ligands B and D, the interaction seems rather weak and the Ser190 side chain essentially drifted away from the ligand.

Among all the ligands investigated, ligand D (4-amino-benzamdine) has the strongest binding affinity according to both experiment and our calculation. To evaluate the role of the amino group in binding, we have examined the possible interactions

between the amino group and trypsin and indentified a stable hydrogen bond between the Ser195 hydroxyl and the amino group (Figure 20 d). Ser 195 together with Asp 102 and His 57 constitute the catalytic triad that attacks the peptide bond. The interaction between Ser195 and ligand D likely enhances the binding of 4-amino-benzamidine to trypsin. However the gain in binding affinity associated with the extra hydrogen bond in trypsin seems to be mostly offset by the increase in solvation free energy in bulk water.

In ligand E, as the charged amine replaces the amidinium group, the capacity for hydrogen bonding decreases, which may have been the cause for the weaker binding affinity when comparing to the other ligands. There were a handful of hydrogen bond acceptors competing for the limited hydrogen bond donors on the amine group, including Gly219 O, Ser190 O, Asp189 Oδ1 and Oδ2, and a water molecule (Figure 20e). It is worth noting that the amine nitrogen of ligand E deviated from the symmetry axis of benzamidine and leaned towards Gly219 for the entire simulation, consistent with previous observations from computer simulation (Leiros et al., 2004).

Additionally, making the ring less hydrophobic does not improve the binding affinity. On the contrary, the ligands with nitrogen atoms in place of oxygen atoms in the phenyl ring have relatively weaker binding to the trypsin. To be more specific, ligand B and ligand C have higher binding free energies than ligand A. Moreover, the amidinium group (ligand A) has been proved to provide more interactions in the binding pocket than amine group (ligand E) and hence stronger binding. For ligand D, the amino group at 4-position of the phenyl ring formed an additional hydrogen bond with Ser 177 at the

catalytic site which enhanced binding by 0.36 kcal/mol.

4.4 CONCLUSION

We have computed the binding free energies of five positively-charged benzamidine analogs to trypsin using an empirical force field based on polarizable atomic multipole electrostatics. Molecular dynamics simulations were performed to perturb the ligands into (or from) benzamidine in both bulk water and in protein-ligand complex. The calculated relative binding free energies, both in sign and magnitude, are in excellent agreement with experimental measurements, with accuracy comparable to that found in experiment. Replacing the phenyl ring with another aromatic structure or amidinium with amine causes significant changes in solvation free energy in bulk water and in the complex, which however leads to a small net change in the overall binding free energy due to cancellation. The 4-amino substitution at the phenyl affects the solvation and binding free energy insignificantly according to our simulations. The molecular dipole moments of the ligands have been characterized in gas phase, in bulk water and in protein-ligand complex. For the ligands studied, molecular dipole moments show no correlation with either the solvation free energy in bulk water or the trypsin binding free energy. The charge redistribution resulting from the chemical change from benzamidine to the other ligands is fairly local – replacing benzene with diazine has no effect on the atomic multipoles at the charged amidinium group. Detailed structure analysis revealed that a few trypsin residues such as Asp189, gly219 and Ser190, and internal water molecules participate in and compete for hydrogen-bonding with the ligands. The

dynamic fluctuation observed for these interactions during the simulations manifests the challenges for sampling in free energy simulations.

5 Hydration Free Energy of Small Organic Molecules

5.1 INTRODUCTION

Hydration plays a significant role in various chemical and biological processes. The prediction of aqueous solvation free energy of molecules is of tremendous interest in areas of medicinal chemistry. For instance, estimating the desolvation penalty for a small ligand and substrate upon binding as a complex is a key issue in drug discovery (Shivakumar, Deng, & Roux, 2009). An accurate determination of solvation free energy also helps investigate the structural stability or folding of a molecule (Eisenberg & McLachlan, 1986). In the development of molecular mechanics force fields, the calculation of hydration free energy has been commonly performed to assess the accuracy of the physical models, since many hydration free energies have been measured experimentally (Jiang, Jordan, & Taylor, 2007; Kaminski, Duffy, Matsui, & Jorgensen, 1994).

A number of theoretical methods have been developed to calculate the hydration free energy. These fall into two major categories, the simulation method and the statistical method. Monte Carlo or molecular dynamics simulations have been widely used to compute the hydration free energies coupled with free energy perturbation theory or thermodynamic integration (Jorgensen & Ravimohan, 1985; Kaminski et al., 1994; Mobley, Bayly, Cooper, Shirts, & Dill, 2009). A significant number of implicit solvent methods have been proved to provide reasonable precision and accuracy in hydration free energy while remain computationally inexpensive (Sandberg, Casemyr, & Edholm, 2002;

Sitkoff, Sharp, & Honig, 1994). These methods are constructed on the basis of continuum electrostatic solvation models. There are statistical approaches which demonstrate good predictive power for hydration free energy. Additive constitutive models based on quantitative structure-activity (QSAR)/-property relationships (QSPR) have been developed in this category (Kravtsov, Karpov, Baskin, Palyulin, & Zefirov, 2007; Viswanadhan, Ghose, & Wendoloski, 2000), such as the HLOGS model which utilizes molecular holograms (Hurst, Heritage, & Clark, 1998), and the ALOGS model, which is an atomic constant approach (Viswanadhan, Ghose, Singh, & Wendoloski, 1999), and more empirical methods, based on a condensed surface representation free energy density, which allow for faster prediction (Jager & Kast, 2001). While such models produce rapid estimation of hydration properties, they are limited because they require a great deal of experimental data pertaining to different classes of organic compounds.

Despite the computational effort, molecular dynamics free energy perturbation with explicit solvent molecules provides the most realistic and most accurate estimation of hydration free energy. With recent computational and methodological advancement, FEP/MD with explicit solvent has become the golden standard of solvation calculation. Here, we applied our polarizable force field to calculate hydration free energies of small molecules via molecular dynamics simulation.

5.2 METHOD

5.2.1 Test set

We selected NMe-Formamide (NMA) and N-butane, two small organic molecules which are representative of the chemical structures and functionalities found in molecular recognition. NMA is a model compound which mimics the backbone of a protein and n-butane is the chain alkane isomer of butane, a common molecule which has two isomers, the other being tetrahedral isobutene (Figure 24).

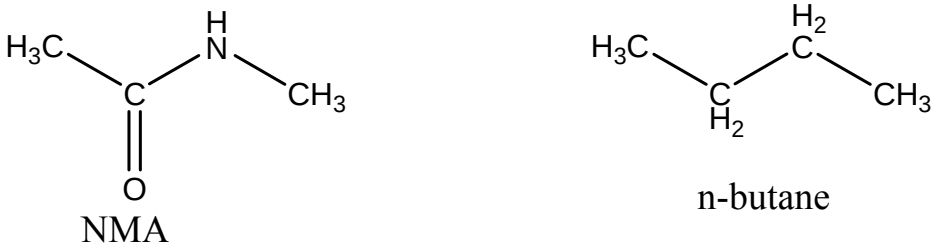


Figure 24 Structures of NMe-Formamide and n-butane

NMA is highly hydrophilic, and therefore it has favorable solvation in water. N-butane, however, is hydrophobic and therefore has poor solvation in a polar solvent, such as water. The atomic parameters of these two molecules are available in AMOEBA force field.

5.2.2 Free energy perturbation

The definition of solvation free energy is the free energy a molecule gains or loses when it is hydrated in a solvent. In other words, it is the difference in free energy before and after the molecule is solvated (19).

$$A_{hyd} = A_{aq} - A_{gas} - A_{wat} \tag{19}$$

The whole free energy perturbation can be described by the thermodynamic cycle (Figure 25).

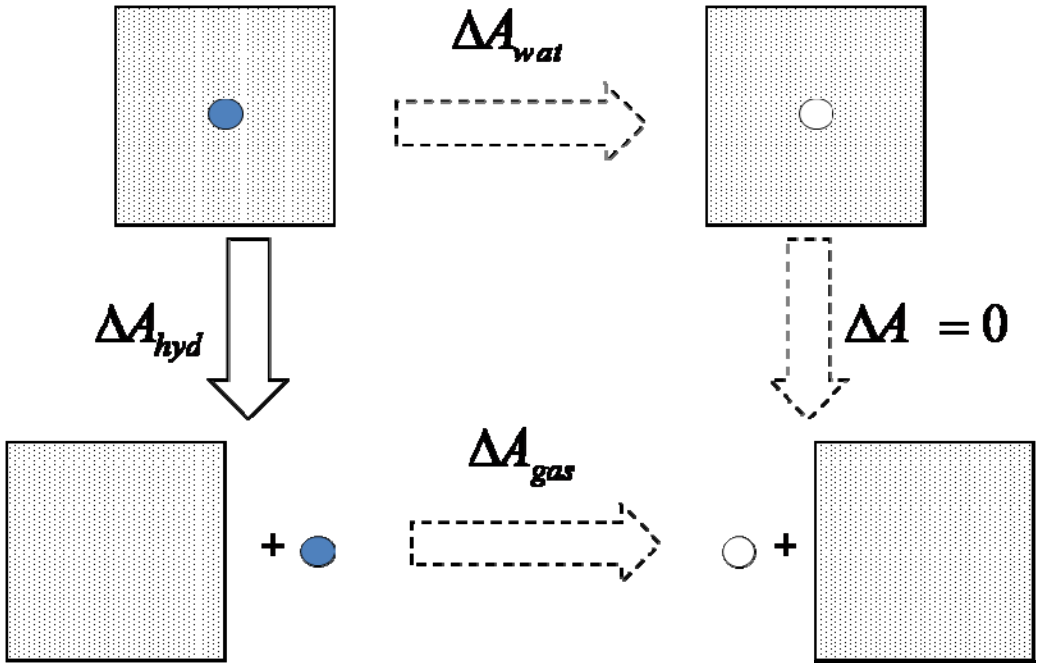


Figure 25. Thermodynamic cycle of hydration free energy calculation. Blue circles represent ligand with full interaction. Open circles represent ligand with no interaction.

In theory, the hydration free energy is calculated by the left leg of Figure 25. It can also be calculated through the other side of the cycle by taking the free energy difference between the top leg and bottom leg. The top leg is the process of turning off the interaction between ligand and water. This is essentially one of the double decoupling paths when computing the absolute ligand-protein binding free energy (see chapter 2). First electrostatics and then van der Waals interactions are turned off. While the inter-molecular forces are being zeroed off, the intra-molecular interactions of the small molecule are being turned off as well. In order to get the absolute solvation free energy of

the molecule, the intra-molecular interactions need to be grown back in the gas phase, which corresponds to the bottom leg. In the thermodynamic cycle, the end states of the two turn-off steps are basically identical, which makes the free energy change of the right leg zero. Therefore, the hydration free energy is calculated in two steps as follows:

$$\Delta A_{hyd} = \Delta A_{gas} - \Delta A_{wat} \qquad (20)$$

The first step is the decoupling of the molecule in water, the second is the growing of the intra-molecular interactions of molecule in the gas phase.

5.2.3 MD simulation

MD simulations have been performed in parallel for all FEP intermediate states along the above-mentioned decoupling pathway in water, using the SANDER executable from the AMBER package (version pre9) (Case et al., 2005). The molecule was soaked in a periodic cubic water box of 1115 water molecules. The cube that encloses the octahedron was 32Å on each side. A single 100 ps NPT dynamics run was performed and the system density was equilibrated to 1 g cm^{-3}. The resulting configuration was then used in all subsequent NVT simulations (~20) for decoupling electrostatic and vdW interactions. The long range electrostatics was treated using Particle Mesh Ewald summation. Induced dipoles were iterated until the changes in atomic induced dipoles were less than 0.01 Debye. Molecular dynamics simulations were performed with a 1 fs time step for 500 ps. Energetics and coordinates of all atoms were saved every 0.5 ps. The temperature was maintained at 298K using the Berendsen weak coupling method.

Intra-molecular simulations were run with TINKER package version 4.0. The electrostatic interactions were switched off in 10 steps. Van der Waals interactions remained unchanged because AMBER kept the intra-molecular vdW interactions intact while turning off vdW inter-molecular interactions. Each simulation was run for 500 ps with a 0.1 fs time step. Coordinates were saved every 0.1 ps. Energy information was printed out every 10 fs. Induced dipoles were iterated until the changes in atomic induced dipoles were less than 0.000001 Debye. The temperature was maintained at 298K using the stochastic thermostat. The first 100-ps was considered as equilibration and thus discarded. The last 400-ps simulation trajectory was analyzed by TINKER to obtain the free energies.

5.2.4 Soft-core modification of vdW long-range correction

In principle, the van der Waals potential has an infinite range, be it 12-6 Lenard-Jones or 14-7 buffered potential (used in this work). The dispersion term with r^{-6}/r^{-7} decays much faster than the electrostatic potential, not to mention the higher order repulsion component. It is customary to establish a cutoff radius r_c for computational expediency, beyond which, the pairwise potential is truncated. The effective vdW potential U(r) is

$$U(r) = \begin{cases} U_{vdW}(r) & r \le r_c \\ 0 & r > r_c \end{cases} \qquad (21)$$

With the assumption that the spatial correlations beyond the cutoff are unity, the contribution of the tail of the potential can be estimated by

$$U_{LRC} = 2\pi N\rho \int_{r_c}^{\infty} r^2 v(r) dr \qquad (22)$$

With a similar concept to the absolute binding free energy calculation, the van der Waals potential between the small molecule and the surroundings was modified with the soft-core method to prevent the singularity problem. Note that the soft-core modification should not only apply to the interactions within cutoff r_c but also to the long-range correction. VdW interactions beyond the cutoff involve any atom of the small molecules which was treated with the soft-core method. This has been implemented in AMBER (pre9) and PMEMD (see Appendix C).

5.2.5 Automation of hydration free energy calculation

The procedure of hydration free energy calculation of small organic molecules can be described by the following flow chart (Figure 26).

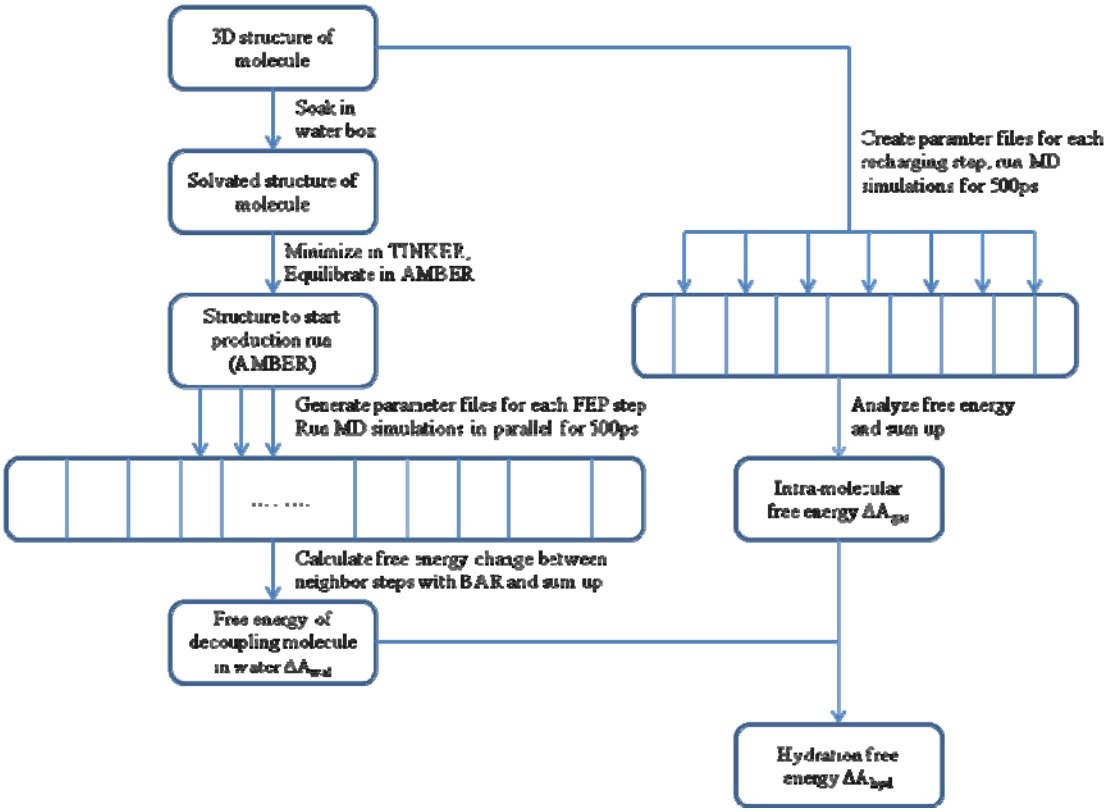


Figure 26. Flow chart of hydration free energy calculation.

The molecule was soaked in water box. AMBER files were generated for 100-ps equilibration followed by the minimization. Parameter files were then created for each FEP state. After the 500-ps simulations were finished, free energies between steps were computed with BAR. At the same time, the small molecule was recharged in multiple steps, and with different parameters. Also, the free energies of recharging the molecule were computed. At the end, the free energy difference between these two paths yielded the hydration free energy. The process was automated by a perl script.

5.3 RESULTS AND DISCUSSION

5.3.1 Hydration free energy

The hydration free energies of NMA and n-butane were calculated with the procedure discussed above. The results are shown in Table 8.

Table 8. Hydration free energies of NMA and butane. ΔA_{wat} is the decoupling free energy of the molecule in water. ΔA_{gas} is the intra-molecular free energy in gas phase. $\Delta A_{cal} = \Delta A_{gas} - \Delta A_{wat}$ is the calculated hydration free energy. ΔA_{exp} is the experimental free energy. All free energy units are kcal/mol.

	ΔA_{wat}		ΔA_{gas}	ΔA_{cal}	ΔA_{exp}
	ele	*vdW*			
NMA	24.21	-1.71	14.24	-8.26	-10.00
Butane	-2.62	-0.18	-1.01	1.79	2.15

We calculated hydration free energies of both NMA and butane within reasonable agreement with experimental data. Our results showed that the solvation of these two molecules behaves differently. NMA hydration is highly favorable in water, giving a negative solvation free energy, whereas the positive free energy loss given by hydration of butane in water indicates that the molecule has an unfavorable hydration in water. This result confirms that a polar molecule like NMA can be easily surrounded by a polar solvent, like water, and that a hydrophobic molecule like butane repels water and therefore has poor solvation.

We also decomposed the decoupling free energies of both molecules. We found that electrostatics is the major component in both cases (~20 times of vdW in magnitude), but in totally opposite direction. To be more specific, the electrostatic interactions play a

major role in favoring the solvation of NMA, while they help to screen the solvation of butane.

5.3.2 Decoupling paths

We tried turning off interactions between butane and water with different schemes for electrostatics and van der Waals by running several sets of simulations with different numbers of steps in Table 9.

Table 9. Comparison of hydration free energy with different perturbation paths

	Scheme	*ΔA*
ele	1.0,0.8,0.6,0.4,0.2,0.0	24.95
	1.0,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0.0	24.21
vdw	1.0,0.8,0.6,0.4,0.2,0.0	-1.87
	1.0,0.9,0.8,0.7,0.6,0.5,0.4,0.2,0.0	-1.60
	1.0,0.95,0.9,0.85,0.8,0.75,0.7,0.65,0.6,0.55,0.5,0.45,0.4,0.2,0.0	-1.91

For electrostatic decoupling, doubling the steps only resulted in a free energy change of 0.7 kcal/mol out of ~25 kcal/mol. Therefore, 11 steps are enough to obtain converged answer electrostatics perturbation. Van der Waals decoupling seemed to require more steps for a converged result. Theoretically, the more FEP steps the better, because there is more sufficient overlap between adjacent states. That is why we expanded the process to 15 steps, and this improved the result by 0.3 kcal/mol from 9-step perturbation. We also found that the free energy change at the beginning and at the

end of the perturbation was relatively small compared to the steps in the middle. Thus, we took out some of the unnecessary steps and made it routine for vdw perturbation with 11 steps (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0).

5.3.3 VdW long-range correction

The vdW free energies in Table 8 were simulated and calculated with 12 Å cutoff and long-range correction. We also ran simulations with small cutoff (9 Å).

Table 10. Van der Waals long-range correction with different cutoff. The free energies are in kcal/mol.

	R_c=9 Å	R_c=12 Å
NMA	0.397	0.123
Butane	0.003	0.125

As shown in Table 10, the contributions of long-range corrections for different cutoffs are insignificant in the total decoupling free energies (less than 5%). This indicates that the van der Waals interactions die off rapidly beyond a certain distance. For a molecule with similar size to butane or NMA, 9 Å is a reasonable cutoff.

5.3.4 Thermostat

We ran the FEP simulations for recharging NMA in 5 steps for 500 ps. Both Berendsen and Anderson (stochastic) thermostats were tested. We used the last 400 ps trajectory for analysis. The results indicate that although Berendsen is usually not considered a canonical thermostat, it yielded almost the same intramolecular free energy (14.29 kcal/mol) as a stochastic thermostat (14.24 kcal/mol).

5.4 CONCLUSION

Hydration is an essential element in drug design. A great deal of effort has been made to determine the hydration free energies of small molecules both by experiments and computations. We have applied our polarizable force field to compute the hydration free energy of the small organic molecules NMA and butane. The results of our simulations showed favorable agreement with experimental values. We also found that electrostatics is the driving force for the favorable solvation of NMA, whereas it plays a major part in damping the solvation of the hydrophobic butane molecule.

We tested the dependence of free energy on the number of FEP steps. Electrostatics decoupling needs fewer steps than van der Waals. We settled on 10 steps for electrostatics and 11 steps for van der Waals perturbation to provide reasonable results. Additionally, we ran simulations with different vdW cutoffs. Simulations with 9 Å and 12 Å cutoff gave similar answers. We explicitly computed long-range corrections, which turned out to be insignificant for both cases.

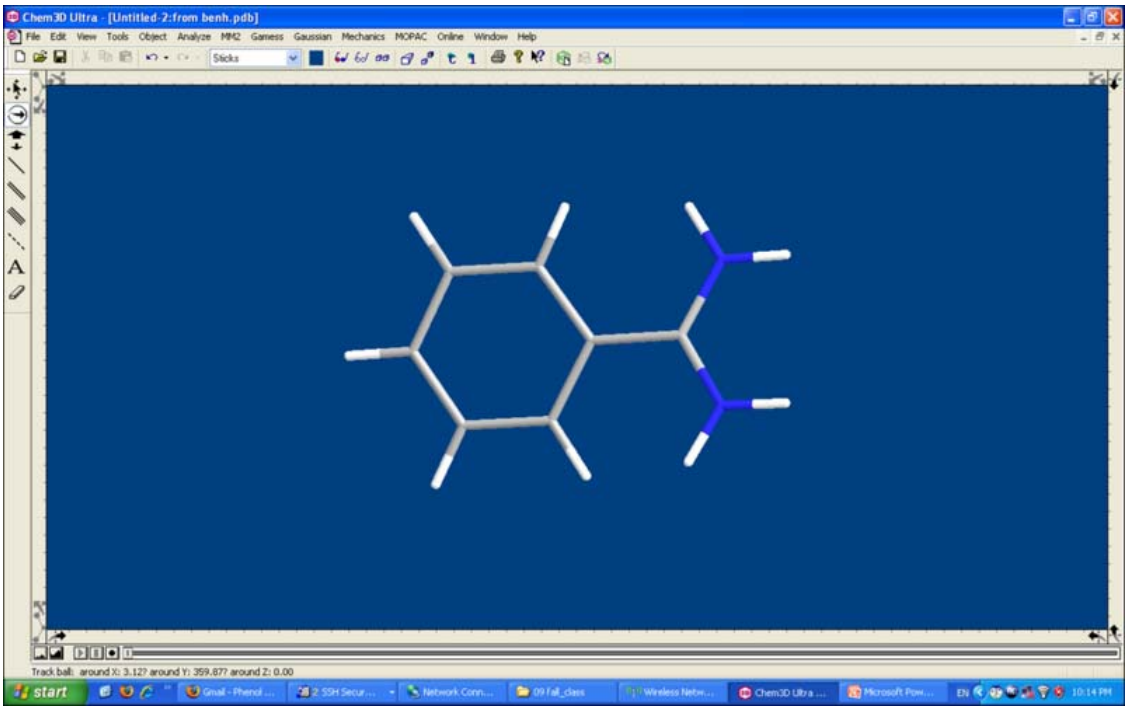
With respect to the intramolecular free energy calculation, the evidence showed no difference between Berendsen and stochastic thermostat.

Appendix A. Tutorial on ligand parameterization in AMOEBA force field

In this chapter, the procedure of parameterizing a small ligand in AMOEBA force field is described in details. It starts with building a molecule with 3D structure. After that, electrostatic parameters are derived from QM calculation. Then vdW and bonded parameters are even taken from pre-existed force field or fitted to QM.

A.1 BUILD THE MOLECULE

Before the parameterization, one must have a structure file of the certain ligand. Most of the time it is not available in PDB database, so the ligand has to be built it from scratch. There are many softwares that have the function to draw a chemical structure. In this work, I use Chem3D.



Once the molecule is built, minimize energy with any of the methods available in Chem3D, for example, MOPAC, MM2, Gaussian, etc. Although they are not very accurate, but this way unphysical geometry will be avoided. After that, save the structure as TINKER format (.xyz) file. We name the benzamidine we draw benz.xyz.

A.2 OPTIMIZE THE STRUCTURE

The structure of the ligand has to be further optimized with QM. Gaussian 03 is used. The input coordinates can be taken from the benz.xyz and pasted to the Gaussian input file format benz.com. Keyword "opt" indicates it is running optimization and 6-31g* is the basis set.

```
%chk=benz.chk
%mem=50MW
%nproc=2
 # opt hf/6-31g*

  opt energy

 1 1
 C                 -0.37610769    0.74520000    0.98303077
 H                 -0.43910769    1.74320000    0.99203077
 C                 -0.72410769    0.05120000    2.12303077
 H                 -1.03010769    0.54320000    2.93803077
 C                 -0.64810769   -1.30880000    2.13203077
 H                 -0.89810769   -1.82380000    2.95203077
 C                 -0.22510769   -1.96280000    1.00303077
 H                 -0.17010769   -2.96080000    1.00603077
 C                  0.12589231   -1.26980000   -0.13596923
 H                  0.43389231   -1.77480000   -0.94296923
 C                  0.05989231    0.11420000   -0.18296923
 C                  0.40489231    0.82420000   -1.33896923
 N                  0.00089231    2.06920000   -1.47396923
 H                  0.24089231    2.58620000   -2.29496923
 H                 -0.54610769    2.49820000   -0.75496923
 N                  1.12789231    0.22220000   -2.26396923
 H                  1.38389231    0.71320000   -3.09696923
 H                  1.41989231   -0.72480000   -2.13196923
<blank line>
```

Once you have the Gaussian input file, run "g03 benz.com" to optimize the structure. This will generate the output file benz.log after optimization.

A.3 SINGLE POINT CALCULATION

Extract the optimized coordinates from optimization output file (benz.log) and create a single point calculation input file benzsp.com. The keyword "sp" indicates this is a single point calculation. We use higher energy level and basis set MP2/6-311++G(2d,2p) in order to get accurate energy result. Lower energy level has been used for optimization since the energy is sensitive to energy level while structure is not.

```
%Mem=500MB
%Nosave
%Chk=benhsp.chk
%Nproc=2
 #MP2/6-311++G(2d,2p) Sp Density=MP2 MaxDisk=960MW

 sp energy

  1 1
C1  -0.219172  0.772015  1.045507
H2  -0.124354  1.841828  1.092625
C3  -0.556685  0.065459  2.183830
H4  -0.742000  0.589477  3.102333
C5  -0.648127 -1.318255  2.138044
H6  -0.916424 -1.863862  3.023554
C7  -0.395888 -2.003092  0.957999
H8  -0.479004 -3.072957  0.925387
C9  -0.043176 -1.306084 -0.181602
H10  0.125831 -1.839237 -1.099697
C11  0.041747  0.084714 -0.138928
C12  0.409652  0.833072 -1.353567
N13  -0.200543  1.960083 -1.635531
H14  0.060474  2.534099 -2.409732
H15  -0.999989  2.242440 -1.111118
N16  1.345776  0.368932 -2.147446
H17  1.581717  0.805673 -3.013818
H18  1.901226 -0.409704 -1.866286
<blank line>
```

A.4 MULTIPOLE CALCULATION

Use Generalized Distributed Multipole Analysis (GDMA) program to calculate multipole moments. Multipole information can be extracted from Gaussian check file (benzsp.chk). Check file is binary, run command “formchk benzsp.chk” to make a formatted check file benzsp.fchk.

GDMA input file should look like this:

```
Title

File benzsp.chk density MP2
Angstrom
AU
Multipoles
Limit 2
Radius H 0.31
Punch benzsp.punch

Start

Finish
```

With this input file gdmain, run gdma with command “gdma < gdmain > benzsp.gdmaout”. There are two output files benzsp.punch and benzsp.gdmaout. The latter has all the computed multipoles.

Edit the multipoles from the GDMA output to make it compatible with AMOEBA format with tinker command poledit.x.

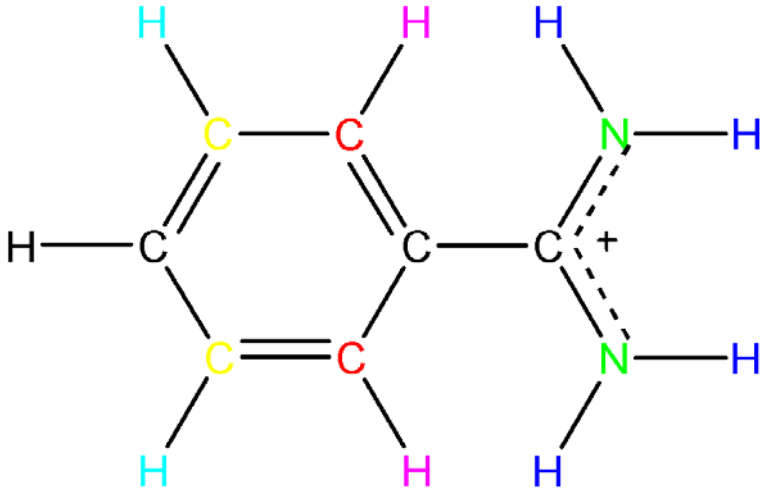
```
  The Multipole Editing Facility can Provide :

   (1) Multipole Parameters from GDMA Output
   (2) Alter Local Coordinate Frame Definitions
   (3) Removal of Intramolecular Polarization

  Enter the Number of the Desired Choice :  1

Site:       1         Name:   C         Atomic Number:       6

Coordinates:           -1.341132      -0.000208      -0.000073

Charge:                 -0.05287
Dipole:                  0.02333        0.00002        0.00006
Quadrupole:              0.34499
                        -0.00007        0.17026
                        -0.00020        0.06036       -0.51525

Site:       2         Name:   C         Atomic Number:       6

Coordinates:           -0.629089      -1.198661       0.039019

Charge:                 -0.38438
Dipole:                 -0.29800       -0.01789        0.04907
Quadrupole:             -0.61935
                        -0.50090        0.36155
                        -0.09178        0.03274        0.25780

  ...skipping....

  Local Frame Definition for Multipole Sites :

    Site     Name      Axis Type     Z Axis  X Axis  Y Axis

      1      C         Z-then-X         7       2       0
      2      C         Z-then-X         3       1       0
      3      C         Z-then-X        12       4       0
      4      C         Z-then-X        10       3       0
      5      C         Z-then-X        13       4       0
      6      C         Z-then-X         5       1       0
      7      C         Bisector         8       9       0
      8      N         Z-then-X         7      15       0
      9      N         Z-then-X         7      17       0
     10      N         Z-then-X         4      19       0
 ....skipping...

 Enter Altered Local Frame Definition [<CR>=Exit] :

 Atomic Polarizabilities for Multipole Sites :

    Site     Name       Polarize          Thole

      1      C            1.3340         0.3900
      2      C            1.3340         0.3900
      3      C            1.3340         0.3900
      4      C            1.3340         0.3900
      5      C            1.3340         0.3900
      6      C            1.3340         0.3900
  ..skipping....

  Enter Atom Number & Polarizability Values [<CR>=Exit] :

  ...skipping...
```

This will generate benzsp.xyz and benzsp.key. File benzsp.key has the electrostatic parameters as below:

```
atom           1    1    C      "ligsp               "     6     12.011    3
atom           2    2    C      "ligsp               "     6     12.011    3
atom           3    3    C      "ligsp               "     6     12.011    3
atom           4    4    C      "ligsp               "     6     12.011    3
atom           5    5    C      "ligsp               "     6     12.011    3
 ...skipping...
multipole      1    7    2             -0.05287
                                        -0.00001    0.00007   -0.02333
                                         0.16559
                                        -0.08263   -0.51058
                                        -0.00011    0.00007    0.34499
multipole      2    3    1             -0.38438
                                        -0.01893    0.04996   -0.29779
                                         0.36086
                                         0.03832    0.26085
                                        -0.49644   -0.10231   -0.62171
...skipping...
polarize       1              1.334      0.390       2    6     7
polarize       2              1.334      0.390       1    3    11
polarize       3              1.334      0.390       2    4    12
polarize       4              1.334      0.390       3    5    10
polarize       5              1.334      0.390       4    6    13
polarize       6              1.334      0.390       1    5    14
polarize       7              1.334      0.390       1    8     9
 ...skipping...
```

A.5 AVERAGE MULTIPOLES

Due to the symmetry of the ligand, average the multipoles of these symmetric atoms. Take benzamidine for example, these atoms in the same color need to be grouped with averaged multipoles.



There are a few things need to be done: (1) change the atom indices so that atoms in one group have the same index; (2) average multipoles; (3) shift the indices with a certain offset so that the parameters can be attached to the bottom of the parameter file. This can be done by our script avgmpole.pl.

A.6 REFIT THE ELECTROSTATICS

Averaging the multipoles might make the total dipole deviate from the QM a little. We need to refit the electrostatics again with TINKER program potential.x. Here is how it works. First create grid points for molecule.

```
The TINKER Electrostatic Potential Facility Can :

   (1) Get QM Potential from a Gaussian CUBE File
   (2) Calculate the Model Potential for a System
   (3) Compare the Model Potentials of Two Systems
   (4) Compare a Model Potential to a Target Grid
   (5) Fit Electrostatic Parameters to Target Grid

 Enter the Number of the Desired Choice :  2

 Output Potential Value at Each Grid Point [N] :Y

  Average Electrostatic Potential over Atoms :
     (Kcal/mole per unit charge)

   Atom          Points        Potential

      1             135          64.2195
      2             371          59.3085
      3             228          54.8976
      4             302          52.9730
      5             235          54.9886
      6             366          59.4935
      7             286          66.5887
      8             523          65.5094
      9             526          65.6609
     10             654          49.3012
     11             635          58.6693
     12            3555          45.6758
     13            3555          45.6982
     14             624          58.6372
     15            1243          62.0622
     16            1637          64.5757
     17            1230          62.0704
     18            1645          64.6416
     19            1071          48.1094
     20            1074          48.0477

Electrostatic Potential over all Grid Points :

Average Magnitude for Potential :             54.3511
```

The output file benzsp.grid has the coordinates of grid and benzsp.pot has the potentials for each grid points. Now use Gaussian command "cubegen 0 potential=MP2 benzsp.fchk benzsp.cube -5 h < benzsp.grid" to compute electrostatic potential of TINKER grid with QM. File benzsp.cube is the output which contains the QM potential for each TINKER grid points (both coordinates and energies) shown as below.

```
-1.341132000000 -0.000208000000 -3.200073000000 0.112800371358
-1.386006000000 0.192135000000 -3.193972000000 0.112181402714
-1.619669000000 0.018815000000 -3.187871000000 0.115556191658
-1.533760000000 -0.282519000000 -3.181770000000 0.116842473623
-1.223612000000 -0.376745000000 -3.175669000000 0.114038580816
-0.946495000000 -0.196588000000 -3.169568000000 0.110351219561
-0.876610000000 0.130788000000 -3.163467000000 0.108378055681
-1.041081000000 0.425788000000 -3.157366000000 0.108460613005
-1.354990000000 0.556388000000 -3.151265000000 0.110406541871
-1.687541000000 0.477711000000 -3.145164000000 0.113700818337
-1.920500000000 0.225795000000 -3.139063000000 0.117550862437
-1.259662000000 -0.730085000000 -3.114658000000 0.117304867056
-0.925333000000 -0.636162000000 -3.108557000000 0.112718952139
-0.667771000000 -0.402456000000 -3.102456000000 0.109223962529
-0.536946000000 -0.079718000000 -3.096355000000 0.107108578011

...skipping...
```

Now use potential.x to take QM potential from cube file (option 1) and then fit the parameters to the grid (option 5). New potential file converted from cube output is consistent with tinker pot file in format.

A.7 ADD OTHER PARAMETERS

Atom definition, including the molecular weight and mass, vdW, bond, angle, out-of-plane, most torsion and polarizability parameters can be obtained from the pre-existed AMOEBA force field (or other force field, e.g. MM3). You can also fine-tune the valence parameters such as bond length and angle values by adjust them to match the QM-optimized structure from early steps.

Not all the torsion parameters can be transferred. For special torsions which have not been defined in old parameter files, you can refine the trosion parameters by fitting to the QM energy profiles. This is done at the last step (i.e. after you get all the other parameters), then you can calculation the torsional profile using QM (restrained optimization at MP2/6-311++G** level, for example), and repeat the same calculation using TINKER (with the particular torsion parameter set to zero). Now fit the 3-term Fourier series to the energy difference between QM and TINKER with gnuplot script we use to do this. In case there are more than one torsion need to be fitted, fit one torsion at a time with the rest fixed at a certain angle.

A.8 FINAL CHECK

After all the parameters are done for the ligand, run TINKER command run "analyze ligand.xyz em" to print out the potential energy and electrostatic properties. If any parameter like bond, angle is missing, it will complain parameters are not defined. Compare the electrostatic properties with QM calculation especially the total charge and dipole moment and x y z components.

Appendix B. Structures and parameters of all the ligands

The TINKER structure (xyz) file is given for each ligand. The coordinates are in the standard orientation from QM optimization. The parameters that follow each ligand xyz file can be appended to amoebapro.prm file for use with TINKER 4.3.

B.1 LIGAND A: BENZAMIDINE

```
 18
  1 C     0.522647   1.178939  -0.271916   274    2    3   11
  2 H    -0.005589   2.083569  -0.513415   275    1
  3 C     1.904253   1.172404  -0.278031   276    1    4    5
  4 H     2.441574   2.074099  -0.503456   277    3
  5 C     2.592187  -0.000006   0.000044   278    3    6    7
  6 H     3.666337  -0.000005   0.000082   279    5
  7 C     1.904225  -1.172418   0.278057   276    5    8    9
  8 H     2.441523  -2.074112   0.503537   277    7
  9 C     0.522619  -1.178950   0.271800   274    7   10   11
 10 H    -0.005653  -2.083554   0.513326   275    9
 11 C    -0.169850   0.000009  -0.000038   284    1    9   12
 12 C    -1.643194   0.000005   0.000022   285   11   13   16
 13 N    -2.295704   1.012262   0.521138   286   12   14   15
 14 H    -3.290703   1.087271   0.484242   287   13
 15 H    -1.807144   1.712397   1.035889   287   13
 16 N    -2.295700  -1.012250  -0.521105   286   12   17   18
 17 H    -3.290695  -1.087298  -0.484163   287   16
 18 H    -1.807134  -1.712345  -1.035906   287   16

atom  274   60   C   "BenC"                6      12.000  3
atom  275   61   H   "BenH"                1       1.008  1
atom  276   60   C   "BenC"                6      12.000  3
atom  277   61   H   "BenH"                1       1.008  1
atom  278   60   C   "BenC"                6      12.000  3
atom  279   61   H   "BenH"                1       1.008  1
atom  280   60   C   "BenC"                6      12.000  3
atom  281   61   H   "BenH"                1       1.008  1
atom  282   60   C   "BenC"                6      12.000  3
atom  283   61   H   "BenH"                1       1.008  1
atom  284   60   C   "BenC"                6      12.000  3
atom  285   62   C   "CN2(C)"      6       12.000  3
atom  286   63   N   "CN(H2)"      7       14.003  3
atom  287   64   H   "HN"                  1       1.008  1
atom  288   64   H   "HN"                  1       1.008  1
atom  289   63   N   "CN(H2)"      7       14.003  3
atom  290   64   H   "HN"                  1       1.008  1
atom  291   64   H   "HN"                  1       1.008  1

   ##  Van der Waals Parameters  ##
```

```
vdw   60   3.800   0.0890
vdw   61   2.980   0.0260  0.92
vdw   62   3.650   0.1010
vdw   63   3.710   0.1100
vdw   64   2.590   0.0220  0.90

    ##  Bond Stretching Parameters

bond   60   60   472.0   1.3887
bond   60   61   370.0   1.0820
bond   60   62   323.0   1.5250
bond   62   63   491.4   1.3250
bond   63   64   487.0   1.0280

    ##  Angle Bending Parameters  ##

angle   60   60   60   54.67   121.700
angle   60   60   61   35.25   120.000
angle   60   60   62   54.67   121.700
angle   60   62   63   28.80   120.000
angle   63   62   63   28.80   120.000
angle   62   63   64   41.70   120.500
angle   64   63   64   29.50   123.000

    ##  Out-of-Plane Bend Parameters  ##

opbend   60   61   0.110
opbend   60   60   0.200
opbend   60   62   0.100
opbend   62   60   0.020
opbend   62   63   0.020
opbend   63   62   0.050
opbend   63   64   0.180

    ##  Torsional Parameters  ##

torsion  60  60  60  60  -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  61  60  60  61   0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  60  60  60  61   0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  60  60  60  62  -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  61  60  60  62   0.000 0.0 1   6.104 180.0 2   0.000 0.0 3
torsion  60  60  62  63  -0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  60  62  63  64   0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
torsion  63  62  63  64   0.000 0.0 1   4.000 180.0 2   0.000 0.0 3

    ##    Polarization Parameters    ##

polarize   274  1.750  275   276   283   284
polarize   275  0.686  274
polarize   276  1.750  274   277   278   281
polarize   277  0.686  276
polarize   278  1.750  276   279   280
polarize   279  0.686  278
polarize   284  1.750  274   282   285
polarize   285  0.496  284   286   289
```

```
polarize   286  1.073  285   287   288
polarize   287  0.496  286

    ##  Atomic Multipole Parameters  ##

multipole   274  276 -284          -0.09524
                                    0.03212    0.00000    0.20067
                                    0.67177
                                    0.00000   -1.18961
                                    0.03229    0.00000    0.51784
multipole   275  274  276                0.14638
                                   -0.01944    0.00000    0.13578
                                    0.02066
                                    0.00000   -0.13728
                                    0.01944    0.00000    0.11663
multipole   276  274 -278               -0.07039
                                    0.00000    0.00000    0.16226
                                    0.51509
                                      0.00000   -1.05821
                                    0.00000    0.00000    0.54312
multipole   277  276  274                0.14253
                                    0.00000    0.00000    0.13208
                                   -0.01770
                                    0.00000   -0.12726
                                    0.00000    0.00000    0.14496
multipole   278  276 -276               -0.06499
                                    0.00000    0.00000    0.15549
                                    0.50678
                                    0.00000   -1.03721
                                    0.00000    0.00000    0.53042
multipole   279  278  276                0.14138
                                   0.00000    0.00000    0.13189
                                  -0.01988
                                   0.00000   -0.12630
                                   0.00000    0.00000    0.14618
multipole   284  285  274               -0.08387
                                  -0.00001    0.00000   -0.01676
                                   0.47376
                                   0.00000   -1.29131
                                  -0.00002    0.00000    0.81754
multipole   285  286 -286                0.15997
                                   0.00000    0.00000    0.02424
                                   0.39018
                                   0.00000   -0.79213
                                  -0.00001    0.00000    0.40196
multipole   286  285  287               -0.20566
                                   0.00836    0.00000    0.11713
                                   0.44050
                                   0.00000   -1.36729
                                  -0.03427    0.00000    0.92679
multipole   287  286  285                0.25307
                                   0.02830    0.00000    0.06289
                                  -0.01360
                                   0.00000   -0.12859
                                   0.00625    0.00000    0.14219
```

B.2 LIGAND B: 1,3-DIAZAMIDINE

```
  16
   1  N     0.463183    1.177164   -0.000363    292    2    9
   2  C     1.787875    1.179836    0.000113    293    1    3    4
   3  H     2.270630    2.138708    0.000076    294    2
   4  C     2.510786   -0.000089    0.000135    295    2    5    6
   5  H     3.582989   -0.000149    0.000379    296    4
   6  C     1.787674   -1.179877   -0.000176    293    4    7    8
   7  H     2.270229   -2.138845    0.000334    294    6
   8  N     0.462970   -1.177093    0.000073    292    6    9
   9  C    -0.107263    0.000040   -0.000071    300    1    8   10
  10  C    -1.609090    0.000034    0.000034    301    9   11   14
  11  N    -2.219301    1.149974    0.000218    302   10   12   13
  12  H    -3.213569    1.237182   -0.000610    303   11
  13  H    -1.665164    1.983107    0.000338    303   11
  14  N    -2.219213   -1.149956   -0.000097    302   10   15   16
  15  H    -3.213473   -1.237252    0.000750    303   14
  16  H    -1.665011   -1.983047   -0.000283    303   14

atom   292   65   N   "BenN"                7      14.003  3
atom   293   66   C   "BenC"                6      12.000  3
atom   294   67   H   "BenH"                1       1.008  1
atom   295   66   C   "BenC"                6      12.000  3
atom   296   67   H   "BenH"                1       1.008  1
atom   297   66   C   "BenC"                6      12.000  3
atom   298   67   H   "BenH"                1       1.008  1
atom   299   65   N   "BenN"                7      14.003  3
atom   300   66   C   "BenC"                6      12.000  3
atom   301   68   C   "CN2(C)"       6      12.000  3
atom   302   69   N   "CN(H2)"       7      14.003  3
atom   303   70   H   "HN"                  1       1.008  1
atom   304   70   H   "HN"                  1       1.008  1
atom   305   69   N   "CN(H2)"       7      14.003  3
atom   306   70   H   "HN"                  1       1.008  1
atom   307   70   H   "HN"                  1       1.008  1

    ##  Van der Waals Parameters  ##

vdw   65   3.710   0.1100
vdw   66   3.800   0.0890
vdw   67   2.980   0.0260  0.92
vdw   68   3.650   0.1010
vdw   69   3.710   0.1100
vdw   70   2.590   0.0220  0.90

    ##  Bond Stretching Parameters  ##

bond   65   66   670.0   1.3370
bond   66   66   472.0   1.3887
bond   66   67   370.0   1.0820
bond   66   68   323.0   1.5250
bond   68   69   491.4   1.3250
bond   69   70   487.0   1.0280

    ##  Angle Bending Parameters  ##
```

```
angle   66   65   66   35.30   117.157
angle   65   66   65   47.50   126.755
angle   65   66   66   47.50   120.020
angle   65   66   67   35.25   119.880
angle   65   66   68   47.50   116.623
angle   66   66   66   54.67   121.700
angle   66   66   67   35.25   120.000
angle   66   68   69   28.80   120.000
angle   69   68   69   28.80   120.000
angle   68   69   70   41.70   120.500
angle   70   69   70   29.50   123.000

    ##  Out-of-Plane Bend Parameters  ##

opbend   65   66   0.210
opbend   66   65   0.200
opbend   66   66   0.200
opbend   66   67   0.110
opbend   66   68   0.100
opbend   68   66   0.020
opbend   68   69   0.020
opbend   69   68   0.050
opbend   69   70   0.180

    ##  Torsional Parameters  ##

torsion  65  66  65  66  -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  65  66  66  66  -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  66  65  66  66  -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  66  65  66  67   0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  65  66  66  67   0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  66  65  66  68  -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  66  66  66  67   0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  67  66  66  67   0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  65  66  68  69   0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  66  68  69  70   0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
torsion  69  68  69  70   0.000 0.0 1   4.000 180.0 2   0.000 0.0 3

    ##   Polarization Parameters   ##

polarize   292  1.073  300   293   308
polarize   293  1.750  292   294   295
polarize   294  0.686  293
polarize   295  1.750  293   296
polarize   296  0.686  295
polarize   300  1.750  292   301
polarize   301  0.496  300   302
polarize   302  1.073  301   303
polarize   303  0.496  302

    ##  Atomic Multipole Parameters  ##

multipole   292  293 -300                          -0.25285
                                  0.03363    0.00000    0.66730
                                  1.11342
```

0.00000 -0.59969
0.01927 0.00000 -0.51374
multipole 293 292 -295 0.04945
0.00001 0.00000 0.15954
0.45452
0.00000 -0.84484
-0.00003 0.00000 0.39032
multipole 294 293 292 0.15632
-0.00001 0.00000 0.12125
-0.01510
0.00000 -0.11896
0.00000 0.00000 0.13405
multipole 295 293 -293 -0.04632
0.00000 0.00000 0.16274
0.43060
0.00000 -0.96120
0.00003 0.00000 0.53061
multipole 296 295 293 0.15813
0.00001 0.00000 0.11979
-0.02823
0.00000 -0.12034
0.00000 0.00000 0.14857
multipole 300 301 292 0.10764
-0.00006 0.00000 -0.03403
0.49803
0.00000 -0.89888
-0.00003 0.00000 0.40085
multipole 301 302 -302 0.17445
0.00002 0.00000 -0.00670
0.44128
0.00000 -0.75184
-0.00005 0.00000 0.31056
multipole 302 301 303 -0.18046
-0.01961 0.00000 0.12808
0.46261
0.00000 -1.30082
-0.06304 0.00000 0.83821
multipole 303 302 301 0.26529
0.03449 0.00000 0.05559
-0.01078
0.00000 -0.12920
-0.00018 0.00000 0.13998

B.3 LIGAND C: 1,4-DIAZAMIDINE

16
1 C 0.604300 -1.213985 0.000130 310 2 3 9
2 H 0.160980 -2.192087 0.000275 311 1
3 N 1.928861 -1.185364 0.000122 312 1 4
4 C 2.508510 -0.013523 0.000001 313 3 5 6
5 H 3.582146 0.009248 0.000039 314 4
6 C 1.764479 1.177590 -0.000155 315 4 7 8
7 H 2.245777 2.136896 -0.000170 316 6
8 N 0.462390 1.145064 -0.000113 317 6 9
9 C -0.127962 -0.046682 -0.000025 318 1 8 10

```
   10  C    -1.613208    0.004411   -0.000013   319    9   11   14
   11  N    -2.335781   -1.091037   -0.000162   320   10   12   13
   12  H    -3.333753   -1.061728   -0.000293   321   11
   13  H    -1.923817   -1.996812   -0.000379   321   11
   14  N    -2.170805    1.182039    0.000192   323   10   15   16
   15  H    -3.160111    1.312789    0.000334   324   14
   16  H    -1.580596    1.989911    0.000293   324   14


atom   310   72   C   "BenC"                6      12.000  3
atom   311   73   H   "BenH"          1     1.008  1
atom   312   74   N   "BenN"                7      14.003  3
atom   313   72   C   "BenC"                6      12.000  3
atom   314   73   H   "BenH"                1       1.008  1
atom   315   72   C   "BenC"                6      12.000  3
atom   316   73   H   "BenH"                1       1.008  1
atom   317   74   N   "BenN"                7      14.003  3
atom   318   72   C   "BenC"                6      12.000  3
atom   319   75   C   "CN2(C)"       6      12.000  3
atom   320   76   N   "CN(H2)"       7      14.003  3
atom   321   77   H   "HN"                  1       1.008  1
atom   322   77   H   "HN"                  1       1.008  1
atom   323   76   N   "CN(H2)"       7      14.003  3
atom   324   77   H   "HN"                  1       1.008  1
atom   325   77   H   "HN"                  1       1.008  1

    ##  Van der Waals Parameters  ##

vdw   72   3.800   0.0890
vdw   73   2.980   0.0260  0.92
vdw   74   3.710   0.1100
vdw   75   3.650   0.1010
vdw   76   3.710   0.1100
vdw   77   2.590   0.0220  0.90
    ##  Bond Stretching Parameters  ##

bond   72   72   472.0   1.3887
bond   72   73   370.0   1.0820
bond   72   74   670.0   1.3370
bond   72   75   323.0   1.5250
bond   75   76   491.4   1.3250
bond   76   77   487.0   1.0280

    ##  Angle Bending Parameters  ##

angle   72   72   73   35.25   120.000
angle   72   72   74   47.50   120.020
angle   72   72   75   54.67   121.700
angle   73   72   74   35.25   119.880
angle   74   72   75   47.50   116.623
angle   72   74   72   35.30   117.157
angle   72   75   76   28.80   120.000
angle   76   75   76   28.80   120.000
angle   75   76   77   41.70   120.500
angle   77   76   77   29.50   123.000
```

```
    ##  Out-of-Plane Bend Parameters  ##

opbend   72   72   0.200
opbend   72   73   0.110
opbend   72   74   0.200
opbend   74   72   0.210
opbend   72   75   0.100
opbend   75   72   0.020
opbend   75   76   0.020
opbend   76   75   0.050
opbend   76   77   0.180

    ##  Torsional Parameters  ##

torsion  73  72  72  73  0.000 0.0 1    7.072 180.0 2   0.000 0.0 3
torsion  73  72  72  74  0.250 0.0 1    5.534 180.0 2  -0.550 0.0 3
torsion  73  72  72  75  0.000 0.0 1    6.104 180.0 2   0.000 0.0 3
torsion  74  72  72  74 -0.670 0.0 1    4.304 180.0 2   0.000 0.0 3
torsion  74  72  72  75 -0.610 0.0 1    4.212 180.0 2   0.000 0.0 3
torsion  72  72  74  72 -0.670 0.0 1    4.304 180.0 2   0.000 0.0 3
torsion  73  72  74  72  0.250 0.0 1    5.534 180.0 2  -0.550 0.0 3
torsion  75  72  74  72 -0.610 0.0 1    4.212 180.0 2   0.000 0.0 3
torsion  72  72  75  76  0.000 0.0 1    2.700 180.0 2   0.000 0.0 3
torsion  74  72  75  76  0.000 0.0 1    2.700 180.0 2   0.000 0.0 3
torsion  72  75  76  77  0.000 0.0 1    4.000 180.0 2   0.000 0.0 3
torsion  76  75  76  77  0.000 0.0 1    4.000 180.0 2   0.000 0.0 3

    ##    Polarization Parameters    ##

polarize  310  1.750  311   312   318
polarize  311  0.686  310
polarize  312  1.073  310   313
polarize  313  1.750  312   314   315
polarize  314  0.686  313
polarize  315  1.750  313   316   317
polarize  316  0.686  315
polarize  317  1.073  315   318
polarize  318  1.750  310   317   319
polarize  319  0.496  318   320   323
polarize  320  1.073  319   321
polarize  321  0.496  320
polarize  323  1.073  319   324
polarize  324  0.496  323

    ##  Atomic Multipole Parameters  ##

multipole   310  312  318                        0.02476
                                  0.14037   0.00000   0.15372
                                  0.50450
                                  0.00000  -0.90755
                                 -0.17577   0.00000   0.40306
multipole   311  310  312                        0.15130
                                  0.00414   0.00000   0.13024
                                  0.02084
                                  0.00000  -0.13307
                                  0.00641   0.00000   0.11224
```

```
multipole   312  310 -313          -0.19410
                                  0.00884    0.00000    0.58918
                                  0.96765
                                  0.00000   -0.51764
                                 -0.01616    0.00000   -0.45002
multipole   313  312  315           0.04844
                                  0.09082    0.00000    0.15185
                                  0.40930
                                  0.00000   -0.80780
                                 -0.04371    0.00000    0.39850
multipole   314  313  312                      0.15718
                                  0.01192    0.00000    0.12078
                                 -0.01837
                                  0.00000   -0.11813
                                 -0.00790    0.00000    0.13650
multipole   315  317  313           0.04789
                                  0.09376    0.00000    0.16313
                                  0.39676
                                  0.00000   -0.81948
                                 -0.02540    0.00000    0.42272
multipole   316  315  317           0.15752
                                  0.00998    0.00000    0.12151
                                 -0.01557
                                  0.00000   -0.11867
                                 -0.00507    0.00000    0.13424
multipole   317  315 -318          -0.25803
                                  0.01478    0.00000    0.66169
                                  1.15004
                                  0.00000   -0.61433
                                 -0.02304    0.00000   -0.53571
multipole   318  317  310           0.03112
                                 -0.01759    0.00000    0.10623
                                  0.62636
                                  0.00000   -1.03684
                                  0.02580    0.00000    0.41048
multipole   319  320 -323           0.17597
                                  0.02140    0.00000    0.00283
                                  0.40503
                                  0.00000   -0.75528
                                 -0.04821    0.00000    0.35025
multipole   320  319  321          -0.19811
                                  0.00587    0.00000    0.11354
                                  0.41901
                                  0.00000   -1.32735
                                 -0.01283    0.00000    0.90834
multipole   321  320  319           0.25054
                                  0.02356    0.00000    0.06421
                                 -0.01171
                                  0.00000   -0.13173
                                  0.01002    0.00000    0.14345
multipole   323  319  324          -0.17929
                                 -0.02427    0.00000    0.12421
                                  0.45863
                                  0.00000   -1.29230
                                 -0.06574    0.00000    0.83367
multipole   324  323  319           0.26715
```

```
 0.03463    0.00000    0.05523
-0.00973
 0.00000   -0.12834
-0.00054    0.00000    0.13807
```

B.4 LIGAND D: 4-AMINO-BENZAMIDINE

```
20
 1  C     0.640305   -0.000002   -0.000115   399     2     6     7
 2  C    -0.077336    1.181820    0.220245   400     1     3    11
 3  C    -1.442045    1.184957    0.229039   401     2     4    12
 4  C    -2.168671    0.000011   -0.000002   402     3     5    10
 5  C    -1.442076   -1.184886   -0.229205   401     4     6    13
 6  C    -0.077339   -1.181733   -0.220616   400     1     5    14
 7  C     2.086424   -0.000023    0.000053   405     1     8     9
 8  N     2.752711   -1.060052    0.421119   406     7    15    16
 9  N     2.752826    1.059994   -0.420893   406     7    17    18
10  N    -3.511974   -0.000022    0.000156   408     4    19    20
11  H     0.439795    2.097802    0.429444   409     2
12  H    -1.971455    2.095079    0.422585   410     3
13  H    -1.971483   -2.095031   -0.422614   410     5
14  H     0.439730   -2.097790   -0.429671   409     6
15  H     2.283663   -1.801606    0.898995   413     8
16  H     3.744657   -1.144049    0.319152   413     8
17  H     2.283924    1.801460   -0.899053   413     9
18  H     3.744751    1.143959   -0.318723   413     9
19  H    -4.032047   -0.837065   -0.154879   417    10
20  H    -4.032046    0.836924    0.155699   417    10


atom   399    111    C   "BenC"                    6       12.000  3
atom   400    111    C   "BenC"                    6       12.000  3
atom   401    111    C   "BenC"                    6       12.000  3
atom   402    111    C   "BenC"                    6       12.000  3
atom   403    111    C   "BenC"                    6       12.000  3
atom   404    111    C   "BenC"                    6       12.000  3
atom   405    112    C   "C2N"             6       12.000  3
atom   406    113    N   "CN(H2)"          7       14.003  3
atom   407    113    N   "CN(H2)"          7       14.003  3
atom   408    114    N   "NH2"             7       14.003  3
atom   409    115    H   "BenH"            1        1.008  1
atom   410    115    H   "BenH"            1        1.008  1
atom   411    115    H   "BenH"            1        1.008  1
atom   412    115    H   "BenH"            1        1.008  1
atom   413    116    H   "Amidine HN"      1        1.008  1
atom   414    116    H   "Amidine HN"      1        1.008  1
atom   415    116    H   "Amidine HN"      1        1.008  1
atom   416    116    H   "Amidine HN"      1        1.008  1
atom   417    117    H   "Amine HN"        1        1.008  1
atom   418    117    H   "Amine HN"        1        1.008  1


    ##  Van der Waals Parameters  ##

vdw   111   3.800   0.0890
```

```
vdw   112   3.650   0.1010
vdw   113   3.710   0.1100
vdw   114   3.710   0.1100
vdw   115   2.980   0.0260  0.92
vdw   116   2.590   0.0220  0.90
vdw   117   2.590   0.0220  0.90

    ##  Bond Stretching Parameters  ##

bond   111   111   472.0   1.3887
bond   111   112   323.0   1.5250
bond   111   114   454.7   1.3780
bond   111   115   370.0   1.0820
bond   112   113   491.4   1.3250
bond   113   116   487.0   1.0280
bond   114   117   487.0   1.0280

    ##  Angle Bending Parameters  ##

angle   111   111   111   54.67   121.700
angle   111   111   112   54.67   121.700
angle   111   111   114   54.67   121.700
angle   111   111   115   35.25   120.000
angle   111   112   113   28.80   120.000
angle   113   112   113   28.80   120.000
angle   112   113   116   41.70   120.500
angle   116   113   116   29.50   123.000
angle   111   114   117   41.70   120.500
angle   117   114   117   29.50   117.000

    ##  Out-of-Plane Bend Parameters  ##

opbend   111   111   0.200
opbend   111   112   0.100
opbend   111   114   0.100
opbend   111   115   0.110
opbend   112   111   0.020
opbend   112   113   0.020
opbend   113   112   0.050
opbend   113   116   0.180
opbend   114   111   0.050
opbend   114   117   0.180

    ##  Torsional Parameters  ##

torsion  111  111  111  111   -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  111  111  111  112   -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  111  111  111  114   -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  111  111  111  115    0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  115  111  111  115    0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  112  111  111  115    0.000 0.0 1   6.104 180.0 2   0.000 0.0 3
torsion  114  111  111  115    0.000 0.0 1   6.104 180.0 2   0.000 0.0 3
torsion  111  111  112  113    0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  111  111  114  117    0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  111  112  113  116    0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
torsion  113  112  113  116    0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
```

```
    ##    Polarization Parameters    ##

polarize   399  1.750   400   405
polarize   400  1.750   399   401   409
polarize   401  1.750   400   402   410
polarize   402  1.750   401   408
polarize   405  1.334   399   406
polarize   406  1.073   405   413
polarize   408  1.073   402   417
polarize   409  0.686   400
polarize   410  0.686   401
polarize   413  0.496   406
polarize   417  0.496   408

    ##  Atomic Multipole Parameters  ##

multipole   399  405  400          -0.10188
                        -0.00002    0.00000   -0.02490
                         0.46307
                         0.00000   -1.34444
                         0.00004    0.00000    0.88136
multipole   400  399 -401          -0.09514
                        -0.02834    0.00000    0.19892
                         0.70210
                         0.00000   -1.21645
                        -0.01604    0.00000    0.51436
multipole   401  402  400          -0.09102
                         0.13012    0.00000    0.11869
                         0.57374
                         0.00000   -1.17332
                        -0.08665    0.00000    0.59957
multipole   402  408  401           0.03361
                        -0.00001    0.00000   -0.00496
                         0.38207
                         0.00000   -1.01552
                         0.00001    0.00000    0.63344
multipole   405  406 -406           0.15359
                        -0.00001    0.00000    0.01917
                         0.38618
                         0.00000   -0.82522
                         0.00000    0.00000    0.43904
multipole   406  405  413          -0.22600
                        -0.00388    0.00000    0.12452
                         0.47166
                         0.00000   -1.42109
                         0.02898    0.00000    0.94943
multipole   408  402  417          -0.26238
                         0.00000    0.00000    0.12935
                         0.53916
                         0.00000   -1.49851
                         0.00000    0.00000    0.95935
multipole   409  400  399           0.14230
                         0.02080    0.00000    0.13977
                         0.02360
```

```
                          0.00000   -0.13649
                         -0.02284    0.00000    0.11289
multipole   410  401  400          0.14061
                         -0.00658    0.00000    0.14188
                          0.00639
                          0.00000   -0.13281
                          0.00091    0.00000    0.12641
multipole   413  406  405          0.24767
                          0.02930    0.00000    0.06637
                         -0.01163
                          0.00000   -0.13268
                          0.00559    0.00000    0.14431
multipole   417  408  402          0.22244
                          0.02820    0.00000    0.07940
                         -0.00998
                          0.00000   -0.15144
                          0.01090    0.00000    0.16142
```

B.5 LIGAND E: BENZYLAMINE

```
  18
   1  C    -2.323653   -0.000280    0.252084   328    2    6    9
   2  C    -1.650592    1.202357    0.099398   329    1    3   10
   3  C    -0.299716    1.203562   -0.201968   330    2    4   11
   4  C     0.382500    0.000262   -0.347520   331    3    5    7
   5  C    -0.299321   -1.203304   -0.202349   330    4    6   12
   6  C    -1.650173   -1.202657    0.099089   329    1    5   13
   7  C     1.855005    0.000529   -0.637404   334    4    8   14   15
   8  N     2.632718   -0.000347    0.680010   335    7   16   17   18
   9  H    -3.373444   -0.000500    0.479903   336    1
  10  H    -2.176796    2.132757    0.204086   337    2
  11  H     0.212312    2.140872   -0.341484   338    3
  12  H     0.213013   -2.140396   -0.342202   338    5
  13  H    -2.176075   -2.133251    0.203522   337    6
  14  H     2.182214    0.881721   -1.168466   341    7
  15  H     2.182277   -0.879907   -1.169686   342    7
  16  H     3.634512   -0.000231    0.539614   343    8
  17  H     2.394346   -0.812440    1.232787   343    8
  18  H     2.394307    0.810980    1.233886   345    8


atom   328    78    C   "BenC"              6       12.000  3
atom   329    78    C   "BenC"              6       12.000  3
atom   330    78    C   "BenC"              6       12.000  3
atom   331    78    C   "BenC"              6       12.000  3
atom   332    78    C   "BenC"              6       12.000  3
atom   333    78    C   "BenC"              6       12.000  3
atom   334    79    C   "CN(C)"             6       12.000  4
atom   335    80    N   "NH3+"              7       14.003  4
atom   336    81    H   "BenH"              1        1.008  1
atom   337    81    H   "BenH"              1        1.008  1
atom   338    81    H   "BenH"              1        1.008  1
atom   339    81    H   "BenH"              1        1.008  1
atom   340    81    H   "BenH"              1        1.008  1
atom   341    82    H   "HC"               1        1.008  1
```

```
atom  342   83   H   "HC"            1       1.008  3
atom  343   84   H   "HN"            1       1.008  1
atom  344   84   H   "HN"            1       1.008  1
atom  345   86   H   "HN"            1       1.008  1

   ##  Van der Waals Parameters  ##

vdw   78   3.800   0.0890
vdw   79   3.820   0.1010
vdw   80   3.810   0.1050
vdw   81   2.980   0.0260  0.92
vdw   82   2.980   0.0240  0.92
vdw   83   2.980   0.0240  0.92
vdw   84   2.700   0.0200  0.91
vdw   86   2.700   0.0200  0.91

   ##  Bond Stretching Parameters  ##

bond   78   78   472.0   1.3887
bond   78   79   453.2   1.4990
bond   78   81   370.0   1.0820
bond   79   80   381.3   1.4480
bond   79   82   341.0   1.1120
bond   79   83   341.0   1.1120
bond   80   84   461.9   1.0150
bond   80   86   461.9   1.0150

   ##  Angle Bending Parameters  ##

angle   78   78   78   54.67   121.700
angle   78   78   79   33.80   122.300
angle   78   78   81   35.25   120.000
angle   78   79   80   56.10   109.500
angle   78   79   82   42.00   110.700
angle   78   79   83   42.00   110.700
angle   80   79   82   59.00   109.300
angle   80   79   83   59.00   109.300
angle   82   79   83   40.00   107.800
angle   79   80   84   43.20   110.900
angle   84   80   84   43.50   107.000
angle   79   80   86   43.20   110.900
angle   84   80   86   43.50   107.000


   ##  Out-of-Plane Bend Parameters  ##

opbend   78   81   0.110
opbend   78   78   0.200
opbend   78   79   0.200

   ##  Torsional Parameters  ##

torsion  78  78  78  78  -0.670 0.0 1   4.304 180.0 2   0.000 0.0 3
torsion  78  78  78  81   0.250 0.0 1   5.534 180.0 2  -0.550 0.0 3
torsion  78  78  78  79  -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  79  78  78  81   0.000 0.0 1   4.104 180.0 2   0.000 0.0 3
```

```
torsion  81  78  78  81   0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  78  78  79  80   0.000 0.0 1   0.064 180.0 2   0.605 0.0 3
torsion  78  78  79  82   0.000 0.0 1   0.000 180.0 2  -0.090 0.0 3
torsion  78  78  79  83   0.000 0.0 1   0.000 180.0 2  -0.090 0.0 3
torsion  78  79  80  84   0.000 0.0 1   0.000 180.0 2  -0.110 0.0 3
torsion  82  79  80  84   0.000 0.0 1  -0.081 180.0 2   0.370 0.0 3
torsion  83  79  80  84   0.000 0.0 1  -0.081 180.0 2   0.370 0.0 3
torsion  78  79  80  86   0.000 0.0 1   0.000 180.0 2  -0.110 0.0 3
torsion  82  79  80  86   0.000 0.0 1  -0.081 180.0 2   0.370 0.0 3
torsion  83  79  80  86   0.000 0.0 1  -0.081 180.0 2   0.370 0.0 3

    ##    Polarization Parameters    ##

polarize   328  1.750   329   336
polarize   329  1.750   328   330   337
polarize   330  1.750   329   331   338
polarize   331  1.750   330   334
polarize   334  1.334   331   335   341  342
polarize   335  1.073   334   343   345
polarize   336  0.686   328
polarize   337  0.686   329
polarize   338  0.686   330
polarize   341  0.496   334
polarize   342  0.496   334   346
polarize   343  0.496   335
polarize   345  0.496   335

    ##  Atomic Multipole Parameters  ##

multipole   328  329 -329              -0.07408
                                  0.00000   0.00000   0.15870
                                  0.51699
                                  0.00000  -1.06836
                                 -0.00001   0.00000   0.55137
multipole   329  328 -330                          -0.08138
                                  0.00003   0.00000   0.16577
                                  0.52829
                                  0.00000  -1.09584
                                  0.01945   0.00000   0.56755
multipole   330  329 -331                          -0.11292
                                  0.00000   0.00000   0.20599
                                  0.67867
                                  0.00000  -1.23659
                                  0.00000   0.00000   0.55792
multipole   331  330 -330             -0.12801
                                  0.00002   0.00000   0.09059
                                  0.47941
                                  0.00000  -1.23942
                                 -0.00004   0.00000   0.76001
multipole   334  335  331             -0.06391
                                  0.10809   0.00000   0.02748
                                  0.13346
                                  0.00000  -0.23493
                                 -0.07570   0.00000   0.10148
multipole   335  334  343             -0.12534
                                 -0.03575   0.00000  -0.03660
```

```
                                0.01711
                                0.00000   -0.05790
                               -0.14169    0.00000    0.04079
multipole   336  328  329              0.13801
                                0.00000    0.00000    0.13472
                               -0.01886
                                0.00000   -0.12785
                                0.00000    0.00000    0.14671
multipole   337  329  328              0.13745
                                0.00153    0.00000    0.13586
                               -0.01691
                                0.00000   -0.12930
                               -0.00240    0.00000    0.14620
multipole   338  330  329              0.13151
                                0.00003    0.00000    0.15000
                                0.01108
                                0.00000   -0.13138
                               -0.00005    0.00000    0.12030
multipole   341  334  335              0.15411
                               -0.00436    0.00000    0.12855
                               -0.03204
                                0.00000   -0.07323
                                0.01366    0.00000    0.10528
multipole   342  334  335              0.15412
                               -0.00437    0.00000    0.12854
                               -0.03205
                                0.00000   -0.07323
                                0.01367    0.00000    0.10528
multipole   343  335  334              0.26526
                                0.03475    0.00000    0.05654
                               -0.04905
                                0.00000   -0.07559
                               -0.00129    0.00000    0.12464
multipole   345  335  334              0.26526
                                0.03475    0.00000    0.05654
                               -0.04905
                                0.00000   -0.07559
                               -0.00129    0.00000    0.12464
```

B.6 F. 4-AMINO-DIAZAMIDINE

```
 20
 1  C     0.647036   -0.000004   -0.000108   442    2     6     7
 2  C    -0.070605    1.181818    0.220252   439    1     3    11
 3  N    -1.435314    1.184955    0.229046   448    2     4
 4  C    -2.161940    0.000009    0.000005   441    3     5    10
 5  N    -1.435345   -1.184888   -0.229198   448    4     6
 6  C    -0.070608   -1.181735   -0.220609   439    1     5    12
 7  C     2.093155   -0.000025    0.000060   443    1     8     9
 8  N     2.759442   -1.060054    0.421126   444    7    13    14
 9  N     2.759557    1.059992   -0.420886   444    7    15    16
10  N    -3.505243   -0.000024    0.000163   446    4    17    18
11  H     0.446526    2.097800    0.429451   440    2
12  H     0.446461   -2.097792   -0.429664   440    6
13  H     2.290394   -1.801608    0.899002   445    8
```

```
   14  H     3.751388   -1.144051    0.319159   445     8
   15  H     2.290655    1.801458   -0.899046   445     9
   16  H     3.751482    1.143957   -0.318716   445     9
   17  H    -4.025316   -0.837067   -0.154872   447    10
   18  H    -4.025315    0.836922    0.155706   447    10

atom   439   125   C   "BenC"              6      12.000  3
atom   440   126   H   "BenH"              1       1.008  1
atom   441   125   C   "BenC"              6      12.000  3
atom   439   125   C   "BenC"              6      12.000  3
atom   440   126   H   "BenH"              1       1.008  1
atom   442   125   C   "BenC"              6      12.000  3
atom   443   127   C   "CN2(C)"            6      12.000  3
atom   444   128   N   "CN(H2)"            7      14.003  3
atom   445   129   H   "Amidine HN"        1       1.008  1
atom   445   129   H   "Amidine HN"        1       1.008  1
atom   444   128   N   "CN(H2)"            7      14.003  3
atom   445   129   H   "Amidine HN"        1       1.008  1
atom   445   129   H   "Amidine HN"        1       1.008  1
atom   446   130   N   "NH2"               7      14.003  3
atom   447   131   H   "Amine HN"          1       1.008  1
atom   447   131   H   "Amine HN"          1       1.008  1
atom   448   132   N   "BenN"              6      12.000  3
atom   448   132   N   "BenN"              6      12.000  3

##  Van der Waals Parameters  ##

vdw   125   3.800   0.0890
vdw   126   2.980   0.0260  0.92
vdw   127   3.650   0.1010
vdw   128   3.710   0.1100
vdw   129   2.590   0.0220  0.90
vdw   130   3.710   0.1100
vdw   131   2.590   0.0220  0.90
vdw   132   3.800   0.0890
vdw   133   2.980   0.0260  0.92

    ##  Bond Stretching Parameters  ##

bond   125   125   472.0   1.3887
bond   125   126   370.0   1.0820
bond   125   127   323.0   1.5250
bond   125   130   454.7   1.3780
bond   125   132   472.0   1.3887
bond   127   128   491.4   1.3250
bond   128   129   487.0   1.0280
bond   130   131   487.0   1.0280
bond   132   133   370.0   1.0820

    ##  Angle Bending Parameters  ##

angle   125   125   125   54.67   121.700
angle   125   125   126   35.25   120.000
angle   125   125   127   54.67   121.700
angle   125   125   132   47.50   120.020
angle   126   125   132   35.25   119.880
```

```
angle   125   127   128   28.80   120.000
angle   127   128   129   41.70   120.000
angle   125   132   125   54.67   121.700
angle   132   125   130   54.67   121.700
angle   125   130   131   41.70   120.500
angle   128   127   128   28.80   120.000
angle   129   128   129   29.50   123.000
angle   131   130   131   35.00   120.000
angle   132   125   132   54.67   121.700
angle   125   132   133   35.25   120.000

    ##  Out-of-Plane Bend Parameters  ##

opbend   125   125   0.200
opbend   125   126   0.110
opbend   125   132   0.200
opbend   132   125   0.200
opbend   125   127   0.100
opbend   127   125   0.100
opbend   127   128   0.020
opbend   128   127   0.050
opbend   128   129   0.180
opbend   130   131   0.180
opbend   130   125   0.100
opbend   125   130   0.200
opbend   132   133   0.110

    ##  Torsional Parameters  ##

torsion  131 130 125 132     0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  125 132 125 130    -0.610 0.0 1   4.212 180.0 2   0.000 0.0 3
torsion  132 125 132 125    -0.670 0.0 1   7.304 180.0 2   0.000 0.0 3
torsion  125 132 125 125    -0.670 0.0 1   7.304 180.0 2   0.000 0.0 3
torsion  125 132 125 126     0.250 0.0 1   7.534 180.0 2  -0.550 0.0 3
torsion  132 125 125 125    -0.670 0.0 1   7.304 180.0 2   0.000 0.0 3
torsion  132 125 125 127    -0.610 0.0 1   7.212 180.0 2   0.000 0.0 3
torsion  127 125 125 126     0.000 0.0 1   6.104 180.0 2   0.000 0.0 3
torsion  125 125 125 126     0.250 0.0 1   7.534 180.0 2  -0.550 0.0 3
torsion  125 127 128 129     0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
torsion  125 125 127 128     0.000 0.0 1   2.700 180.0 2   0.000 0.0 3
torsion  128 127 128 129     0.000 0.0 1   4.000 180.0 2   0.000 0.0 3
torsion  126 125 125 126     0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  126 125 125 130     0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  125 125 125 125    -0.670 0.0 1   7.304 180.0 2   0.000 0.0 3
torsion  133 132 125 130     0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  133 132 125 132     0.250 0.0 1   7.534 180.0 2  -0.550 0.0 3
torsion  133 132 125 126     0.000 0.0 1   7.072 180.0 2   0.000 0.0 3
torsion  133 132 125 125     0.250 0.0 1   7.534 180.0 2  -0.550 0.0 3

    ##    Polarization Parameters    ##

polarize   439    1.750    440    442    448
polarize   440    0.686    439
polarize   441    1.750    446    448
polarize   442    1.750    439    443
polarize   443    1.334    442    444
```

```
polarize   444    1.073    443    445
polarize   445    0.496    444
polarize   446    1.073    441    447
polarize   447    0.496    446
polarize   448    1.120    439    441

    ##  Atomic Multipole Parameters  ##

multipole   439  442 -448              -0.09514
                                  -0.02834    0.00000    0.19892
                                   0.70210
                                   0.00000   -1.21645
                                  -0.01604    0.00000    0.51436
multipole   440  439  442               0.14230
                                   0.02080    0.00000    0.13977
                                   0.02360
                                   0.00000   -0.13649
                                  -0.02284    0.00000    0.11289
multipole   441  446  448               0.03361
                                  -0.00001    0.00000   -0.00496
                                   0.38207
                                   0.00000   -1.01552
                                   0.00001    0.00000    0.63344
multipole   442  443  439                         -0.10188
                                  -0.00002    0.00000   -0.02490
                                   0.46307
                                   0.00000   -1.34444
                                   0.00004    0.00000    0.88136
multipole   443  444 -444               0.15359
                                  -0.00001    0.00000    0.01917
                                   0.38618
                                   0.00000   -0.82522
                                   0.00000    0.00000    0.43904
multipole   444  443  445              -0.22600
                                  -0.00388    0.00000    0.12452
                                   0.47166
                                   0.00000   -1.42109
                                   0.02898    0.00000    0.94943
multipole   445  444  443               0.24767
                                   0.02930    0.00000    0.06637
                                  -0.01163
                                   0.00000   -0.13268
                                   0.00559    0.00000    0.14431
multipole   446  441  447              -0.26238
                                   0.00000    0.00000    0.12935
                                   0.53916
                                   0.00000   -1.49851
                                   0.00000    0.00000    0.95935
multipole   447  446  441               0.22244
                                   0.02820    0.00000    0.07940
                                  -0.00998
                                   0.00000   -0.15144
                                   0.01090    0.00000    0.16142
multipole   448  441  439              -0.09102
                                   0.13012    0.00000    0.11869
                                   0.57374
```

```
                                  0.00000   -1.17332
                                 -0.08665    0.00000    0.59957
multipole   410  448  439                    0.14061
                                 -0.00658    0.00000    0.14188
                                  0.00639
                                  0.00000   -0.13281
                                  0.00091    0.00000    0.12641
```

Appendix C. Softcore modification of vdW in AMBER

C.1 INTRODUCTION

Softcore modification of vdW interaction was incorporated in AMBER and PMEMD (both pre-9 version). In order to run simulation with modified AMBER or PMEMD, you ought to have a file name "soft_atm.txt" in your work directory which includes the range(s) of atoms. The format of soft_atm.txt obeys following rules: (1) These atom numbers should be consistent with xyz file (or pdb file); (2) Each line contains only one segment; (3) If it is a range, use hyphen to connect the first atom and the last atom of the segment; and (4) maximum 20 lines are allowed.

Keywords related to softcore should be included in mdin file (input file of amber/pmemd). In AMBER, these keywords are

- softcore_alpha (soft_alpha in pmemd): coefficient of softcore. Default 0.5.
- softcore_lamda (soft_lamda in pmemd): scaling factor of decoupling, default 1.0.
- softcore_expo (soft_expo in pmemd): exponent of scaling factor, default 4.
- vdw_longrange_lambda: scaling factor of vdw longrange correction, default 1.0 (no LRC); 0.0 when there is full LRC

C.2 MODIFICATION DETAILS

C.2.1 Modification of AMBER pre-9

In AMBER, the files that have been modified are amoeba_mdin.f and amoeba_vdw.f. In amoeba_mdin.f, add the keywords to the variable definitions and assign initial values.

```
module amoeba_mdin
 implicit none
 private
 integer,save :: do_amoeba=0,do_amoeba_valence=1,do_amoeba_nonbond=1, &
                 do_vdw_taper=1,do_vdw_longrange=1
 _REAL_,save :: sor_coefficient = 0.75d0
 _REAL_,save :: dipole_scf_tol = 0.01d0
 integer,save :: dipole_scf_iter_max = 50
 _REAL_,save :: ee_dsum_cut=7.d0
 _REAL_,save :: ee_damped_cut=4.5d0
 _REAL_,save :: vdw_taper = 0.9d0
 _REAL_,save :: thole_expon_coeff=0.39d0
 _REAL_,save :: compress = 0.000046d0
 _REAL_,save :: softcore_lamda = 1.0d0
 _REAL_,save :: softcore_alpha = 0.5d0
 _REAL_,save :: softcore_expo = 4
 _REAL_,save :: vdw_longrange_lambda = 1.0d0
logical, save :: verbose=.false.
integer,save :: beeman_integrator = 0

!variables and arrays for softcore file
integer,save :: soft_atom_range1(20),soft_atom_range2(20),soft_line

public AMOEBA_read_mdin,do_amoeba,do_amoeba_valence, &
       do_amoeba_nonbond,verbose,beeman_integrator, &
       sor_coefficient,dipole_scf_tol,dipole_scf_iter_max, &
       ee_dsum_cut,ee_damped_cut,thole_expon_coeff,vdw_taper, &
       compress,do_vdw_taper,do_vdw_longrange, &
       softcore_lamda,softcore_alpha,softcore_expo, &
       AMOEBA_read_soft, soft_atom_range1,soft_atom_range2, soft_line,
&
        vdw_longrange_lambda
contains
```

In the file, soft_atom_range1 and soft_atom_range2 are two arrays store the ranges of softcore atoms taken from soft_atom.txt. soft_line is the number of lines of soft_atom.txt.

AMOEBA_read_soft is a subroutine which read the soft_atom.txt file which will be called in amoeba_vdw.f.

Now include the new keywords in amoeba_mdin namelist and add subroutine AMOEBA_read_soft as follows:

```
subroutine AMOEBA_read_mdin(nf)
   ...skipping...
 namelist/amoeba/do_amoeba,do_amoeba_valence,do_amoeba_nonbond, &
                 do_bond,do_ureyb,do_reg_angle,  &
                 do_trig_angle,do_opbend,do_torsion,do_str_torsion, &
                 do_pi_torsion,do_strbend,do_torsion_torsion, &
                 do_recip,do_adjust,do_direct,do_self, &
                 do_vdw,do_induced,verbose,beeman_integrator, &
                 sor_coefficient,dipole_scf_tol, &
                 dipole_scf_iter_max,ee_dsum_cut,ee_damped_cut, &

thole_expon_coeff,vdw_taper,do_vdw_taper,do_vdw_longrange, &

compress,softcore_lamda,softcore_alpha,softcore_expo,vdw_longrange_lamb
da
```

```
subroutine AMOEBA_read_soft()
  integer pos_dash,length,i_range,i
  character(len=20) atm_range
  character(len=6) temp

  !flag whether soft_atm.txt exists
  logical alive
  inquire(file='soft_atm.txt',exist=alive)
  if(alive) then
     do i=1,20
        soft_atom_range1(i)=0;
        soft_atom_range2(i)=0;
     end do

     soft_line=0
     open (11,FILE='soft_atm.txt')
     do while (.true.)

        read(11,*,end=99) atm_range
        !write(*,*) atm_range
        pos_dash=scan(atm_range,'-')
        !write(*,*) "position",pos_dash
        length=len_trim(atm_range)
        if (length > 0) then
```

```
           soft_line=soft_line+1
        endif
        !write(*,*) "length",length
        if(pos_dash.gt.0) then
           temp=atm_range(1:pos_dash-1)
           read(temp,*) soft_atom_range1(soft_line)
           temp=atm_range(pos_dash+1:length)
           read(temp,*) soft_atom_range2(soft_line)

        else
           read(atm_range,*) soft_atom_range1(soft_line)
           read(atm_range,*) soft_atom_range2(soft_line)
        endif
      end do
   else
     write(6,*) 'soft_atm.txt not found'
     call mexit(6,1)
   endif
   99 continue
   close(11)

end subroutine AMOEBA_read_soft
```

Now add softcore modification of vdw interaction in amoeba_vdw.f file.

```
subroutine AM_VDW_DIRECT_ene_frc_i(i,ipairs,numtot,xk,yk,zk, &
                                   crd,ene_vdw,frc,virial)
  use nblist, only: bckptr,imagcrds,tranvec
  use constants, only : zero,one,two,three,four,five,seven
  use amoeba_mdin, only : do_vdw_taper,&
                          softcore_lamda,softcore_alpha,softcore_expo,
&
                         soft_atom_range1,soft_atom_range2, soft_line
 integer,intent(in) :: i,ipairs(*),numtot
 _REAL_,intent(in) :: xk,yk,zk,crd(3,*)
 _REAL_,intent(inout) :: ene_vdw,frc(3,*),virial(3,3)
 integer :: itran,it,jt,ih,jh,j,m,np,mask27,i_range

     if ( delr2 < vdw_switch_off_2 )then
       jt = vdw_atom_type(j)
       eps = vdw_epsilon(jt,it)
       rad = vdw_radius(jt,it)
       delr = sqrt(delr2)
       rho = delr / rad
       rho6 = rho**6
       rho7 = rho6*rho
       t1 = ((one + vdw_buf_delta) / (rho + vdw_buf_delta))**7
       t2 = (one + vdw_buf_gamma) / (rho7 + vdw_buf_gamma)
       dt1drho = -seven*t1 / (rho + vdw_buf_delta)
```

```
       dt2drho = -seven*t2 * (rho6 / (rho7 + vdw_buf_gamma))
       drhodr = one / rad
       do i_range = 1,soft_line
        if ((i.ge.soft_atom_range1(i_range).and.
i.le.soft_atom_range2(i_range)).xor.j.ge.soft_atom_range1(i_range).and.
j.le.soft_atom_range2(i_range)) )then
         eps = eps * softcore_lamda ** softcore_expo
         t1 = (one + vdw_buf_delta)**7 / (softcore_alpha * (1 -
softcore_lamda)**2 + (rho + vdw_buf_delta)**7)
         t2 = (one + vdw_buf_gamma) / (softcore_alpha * (1 -
softcore_lamda)**2 + rho7 + vdw_buf_gamma)
         dt1drho = (-seven * (rho + vdw_buf_delta)**6 * t1) /
(softcore_alpha * (1 - softcore_lamda)**2 &
          + (rho + vdw_buf_delta)**7)
         dt2drho = (-seven * rho6 * t2) / (softcore_alpha * (1 -
softcore_lamda)**2 + rho7 + vdw_buf_gamma)
        endif
       end do
       f = eps*t1*(t2 - two)
       dfdr = eps*(dt1drho*(t2 - two) + t1*dt2drho)*drhodr
```

Include longrange correction softcore modification in subroutine

AM_VDW_longrange_factor.

```
subroutine AM_VDW_longrange_factor(num_atoms)
 use amoeba_mdin, only : do_vdw_taper,&

softcore_lamda,soft_atom_range1,soft_atom_range2,&

soft_line,vdw_longrange_lambda,AMOEBA_read_soft

use constants, only : zero,one,two,three,four,five,seven,pi
integer,intent(in) :: num_atoms

integer ier,n,nt,kdel,ndel,i,j,i_range,i_soft,i_soft_type
integer,save,allocatable :: lig_type_ct(:)
_REAL_ ::
r,r1,r2,r3,r4,r5,req,eps,f,sume,sumv,t1,t2,rho,switch,delr,rho6,rho7
_REAL_ :: dt1drho,dt2drho,drhodr,dfdr,dswitch_dr,g1,g2
allocate(vdw_type_count(num_vdw_atom_types),stat=ier); REQUIRE(ier==0)
allocate(lig_type_ct(num_vdw_atom_types),stat=ier); REQUIRE(ier==0)

 if (vdw_longrange_lambda .ne. 1.0 .or. softcore_lamda .ne. 1.0) then
   call AMOEBA_read_soft()
endif
do n = 1,num_vdw_atom_types
  vdw_type_count(n) = 0
  lig_type_ct(n) = 0
```

```
enddo
do n = 1,num_atoms
  nt = vdw_atom_type(n)
  vdw_type_count(nt) = vdw_type_count(nt) + 1
  do i_range = 1, soft_line
      if (n .ge. soft_atom_range1(i_range).and.
n.le.soft_atom_range2(i_range)) then
          lig_type_ct(nt) = lig_type_ct(nt) + 1
      endif
  enddo
enddo

 ...skipping...

 ! note the 2*pi below not 4*pi---since we do each i,j pair 2x
    ene_vdw_longrange_factor = ene_vdw_longrange_factor + two*pi*sume*
(vdw_type_count(i)*vdw_type_count(j) &
    -(1-
vdw_longrange_lambda)*(lig_type_ct(i)*vdw_type_count(j)+(vdw_type_count
(i)-lig_type_ct(i))*lig_type_ct(j)))
     vir_vdw_longrange_factor = vir_vdw_longrange_factor +
two*pi*sumv*(vdw_type_count(i)*vdw_type_count(j) &
   -(1-
vdw_longrange_lambda)*(lig_type_ct(i)*vdw_type_count(j)+(vdw_type_count
(i)-lig_type_ct(i))*lig_type_ct(j)))
   enddo
 enddo

end subroutine AM_VDW_longrange_factor
```

C.2.2 Modification of AMBER pre-9

In PMEMD, there are three files that need to be modified. They are mdin_amoeba_dat.fpp, amoeba_vdw.fpp and amoeba_direct.fpp. First of all, in the mdin_amoeba_dat.fpp file, add the parameters to the variable definitions in module mdin_amoeba_dat_mod and include the subroutine that reads the softcore atom ranges

```
module mdin_amoeba_dat_mod
#ifdef AMOEBA

 implicit none

! Data that should be broadcast to slave processes from the master:

 integer, parameter    :: mdin_amoeba_int_cnt = 24
```

```
 integer                      do_amoeba_valence, do_amoeba_nonbond,
do_bond, &
                              do_ureyb, do_reg_angle, do_trig_angle,
do_opbend, &
                              do_torsion, do_pi_torsion, do_strbend, &
                              do_torsion_torsion, do_str_torsion,
do_recip, &
                              do_adjust, do_direct, do_self, do_vdw, &
                              do_induced, do_vdw_taper, do_vdw_longrange,
&
                              beeman_integrator, dipole_scf_iter_max, &
                              amoeba_verbose, soft_expo

integer,save         ::
soft_atom_range1(20),soft_atom_range2(20),soft_line

common / mdin_amoeba_int / do_amoeba_valence, do_amoeba_nonbond,
do_bond, &
                           do_ureyb, do_reg_angle, do_trig_angle,
do_opbend, &
                           do_torsion, do_pi_torsion, do_strbend, &
                           do_torsion_torsion, do_str_torsion,
do_recip, &
                           do_adjust, do_direct, do_self, do_vdw, &
                           do_induced, do_vdw_taper, do_vdw_longrange,
&
                           beeman_integrator, dipole_scf_iter_max, &
                           amoeba_verbose,soft_expo

common                     soft_atom_range1,soft_atom_range2, soft_line

save  :: / mdin_amoeba_int /

integer, parameter    :: mdin_amoeba_dbl_cnt = 10

double precision             compress, dipole_scf_tol, ee_dsum_cut, &
                             ee_damped_cut, sor_coefficient, &
                             thole_expon_coeff, vdw_taper, &
                             soft_lamda,
soft_alpha,vdw_longrange_lambda

common / mdin_amoeba_dbl /   compress, dipole_scf_tol, ee_dsum_cut, &
                             ee_damped_cut, sor_coefficient, &
                             thole_expon_coeff, vdw_taper, &
                             soft_lamda,
soft_alpha,vdw_longrange_lambda

save  :: / mdin_amoeba_dbl /

contains
```

Assign the initial values for these softcore parameters in subroutine init_mdin_amoeba_dat

```
subroutine init_mdin_amoeba_dat

 use file_io_dat_mod
 use file_io_mod

 implicit none

!  Local variables:

 integer       :: ifind

 namelist /amoeba/ do_amoeba_valence, do_amoeba_nonbond, &
                   do_bond, do_ureyb, do_reg_angle,  &
                   do_trig_angle, do_opbend, do_torsion,
do_str_torsion, &
                   do_pi_torsion, do_strbend, do_torsion_torsion, &
                   do_recip, do_adjust, do_direct, do_self, &
                   do_vdw, do_induced, amoeba_verbose,
beeman_integrator, &
                   sor_coefficient, dipole_scf_tol, &
                   dipole_scf_iter_max, ee_dsum_cut, ee_damped_cut, &
                   thole_expon_coeff, vdw_taper, do_vdw_taper, &
                   do_vdw_longrange, compress, &
                   soft_lamda, soft_alpha, soft_expo,
vdw_longrange_lambda

   ...skipping...

 soft_lamda=1.d0
soft_alpha=0.5d0
soft_expo=4
vdw_longrange_lambda=1.d0
```

Add subroutine AMOEBA_read_soft

```
subroutine AMOEBA_read_soft

 use parallel_dat_mod

 integer pos_dash,length,i_range,i
 character(len=20) atm_range
 character(len=6) temp
 logical alive
 inquire(file='soft_atm.txt',exist=alive)
```

```
 if (alive) then
    do i=1,20
       soft_atom_range1(i)=0;
       soft_atom_range2(i)=0;
    end do
    soft_line=0

    open (11,FILE='soft_atm.txt')
    do while (.true.)
       soft_line=soft_line+1
       read(11,*,end=99) atm_range
       pos_dash=scan(atm_range,'-')
       length=len_trim(atm_range)
       if(pos_dash.gt.0) then
          temp=atm_range(1:pos_dash-1)
          read(temp,*) soft_atom_range1(soft_line)
          temp=atm_range(pos_dash+1:length)
          read(temp,*) soft_atom_range2(soft_line)
       else
          read(atm_range,*) soft_atom_range1(soft_line)
          read(atm_range,*) soft_atom_range2(soft_line)
       endif
    end do
  else
     write(6,*)'soft_atm.txt not found'
     call mexit(6, 1)
  endif
 99 continue
 close(11)

end subroutine AMOEBA_read_soft
```

PMEMD is different in handling vdW calculation in that the vdW interaction is calculated within amoeba_direct.fpp, while the long-range correction is done by amoeba_vdw.fpp. Both files need to be modified with softcore method. In the amoeba_direct.fpp, softcore modification is included in subroutine am_vdw_direct_ene_frc_i.

```
subroutine am_vdw_direct_ene_frc_i(atm_i, img, ipairs_sublst, x_tran, &
                                   pair_cnt, crd, ene_vdw, frc, img_frc,
&
                                   virial, img_atm_map)
```

```
 use mdin_amoeba_dat_mod, only : do_vdw_taper, soft_lamda, soft_expo,
soft_alpha, dipole_scf_iter_max, &
                                 soft_atom_range1,soft_atom_range2,
soft_line
 use amoeba_flags_mod
 use img_mod

  implicit none


! Formal arguments:

 integer, intent(in)                  :: atm_i
 type(img_rec), intent(in)            :: img(*)
 integer, intent(in)                  :: ipairs_sublst(*)
 double precision                     :: x_tran(1:3, 0:17)
 integer, intent(in)                  :: pair_cnt
 double precision, intent(in)         :: crd(3, *)
 double precision, intent(in out)     :: ene_vdw
 double precision, intent(in out)     :: frc(3, *)
 double precision, intent(in out)     :: img_frc(3, *)
 double precision, intent(in out)     :: virial(3, 3)
 integer, intent(in)                  :: img_atm_map(*)


! Local variables:

 integer               :: itran, it, jt, ih, jh, idx
 integer               :: sublst_idx
 integer               :: atm_j, img_j
 integer               :: i_range
double precision      :: wi, wj
double precision      :: delx, dely, delz
double precision      :: delr, delr2
double precision      :: eps
double precision      :: rad
double precision      :: rho
double precision      :: t1, t2
double precision      :: dt1drho, dt2drho, drhodr
double precision      :: term
double precision      :: dfx, dfy, dfz
double precision      :: rho6, rho7
double precision      :: vxx, vxy, vxz, vyy, vyz, vzz
double precision      :: switch, dswitch_dr
double precision      :: f, dfdr
double precision      :: delr3, delr4, delr5
integer, parameter    :: mask27 = Z"07FFFFFF"

...skipping...

  if (delr2 .lt. vdw_switch_off_2) then
```

```
    jt = vdw_atom_type(atm_j)
    eps = vdw_epsilon(jt, it)
    rad = vdw_radius(jt, it)
    delr = sqrt(delr2)
    rho = delr / rad
    rho6 = rho**6
    rho7 = rho6 * rho
    t1 = ((1.d0 + vdw_buf_delta) / (rho + vdw_buf_delta))**7
    t2 = (1.d0 + vdw_buf_gamma) / (rho7 + vdw_buf_gamma)
    dt1drho = -7.d0 * t1 / (rho + vdw_buf_delta)
    dt2drho = -7.d0 * t2 * (rho6 / (rho7 + vdw_buf_gamma))
    drhodr = 1.d0 / rad
    do i_range = 1,soft_line
       if(((atm_i.ge.soft_atom_range1(i_range)).and.
(atm_i.le.soft_atom_range2(i_range))).or.
((atm_j.ge.soft_atom_range1(i_range)) .and.
(atm_j.le.soft_atom_range2(i_range)))) then
          eps = eps * soft_lamda ** soft_expo
          t1 = (1.d0 + vdw_buf_delta)**7 / (soft_alpha * (1.d0 -
soft_lamda)**2 + (rho + vdw_buf_delta)**7)
          t2 = (1.d0 + vdw_buf_gamma) / (soft_alpha * (1.d0 -
soft_lamda)**2 + rho**7 + vdw_buf_gamma)
          dt1drho = (-7.d0 * (rho + vdw_buf_delta)**6 * t1) /
(soft_alpha * (1.d0 - soft_lamda)**2 + (rho + vdw_buf_delta)**7)
          dt2drho = (-7.d0 * rho**6 * t2) / (soft_alpha * (1.d0 -
soft_lamda)**2 + rho**7 + vdw_buf_gamma)
       endif
    end do
    f = eps * t1 * (t2 - 2.d0)
    dfdr = eps * (dt1drho * (t2 - 2.d0) + t1 * dt2drho) * drhodr
```

Long-range correction subroutine am_vdw_longrange_factor in amoeba_vdw.fpp file now has the soft-core modification as below:

```
subroutine am_vdw_longrange_factor(atm_cnt)

 use mdin_amoeba_dat_mod, only : do_vdw_taper,
soft_lamda,AMOEBA_read_soft, &
soft_atom_range1,soft_atom_range2,soft_line,&
                                 vdw_longrange_lambda
use gbl_constants_mod, only : PI

implicit none

! Formal arguments:

integer, intent(in)   :: atm_cnt
```

```
! Local variables:

integer               :: ier, n, nt, kdel, ndel, i, j
integer               :: i_range,i_soft,i_soft_type
double precision      :: r, r1, r2, r3, r4, r5
double precision      :: req, eps, f, sume, sumv, t1, t2, rho, switch,
delr
double precision      :: dt1drho, dt2drho, drhodr, dfdr, dswitch_dr,
g1, g2
integer,save,allocatable :: lig_type_ct(:)

if (vdw_longrange_lambda .ne. 1.0 .or. soft_lamda .ne. 1.0) then
   call AMOEBA_read_soft()
endif
allocate(lig_type_ct(vdw_param_cnt),stat=ier)
do n = 1, vdw_param_cnt
  vdw_type_count(n) = 0
  lig_type_ct(n) = 0
end do
do n = 1, atm_cnt
  nt = vdw_atom_type(n)
  vdw_type_count(nt) = vdw_type_count(nt) + 1
  do i_range = 1, soft_line
      if (n .ge. soft_atom_range1(i_range).and.
n.le.soft_atom_range2(i_range)) then
         lig_type_ct(nt) = lig_type_ct(nt) + 1
      endif
  enddo
end do

...skipping...

    ! Note the 2 * pi below not 4 * pi---since we do each i, j pair 2x.
    ene_vdw_longrange_factor = ene_vdw_longrange_factor + 2.d0 * PI *
sume * (vdw_type_count(i)*vdw_type_count(j) &
    -(1-
vdw_longrange_lambda)*(lig_type_ct(i)*vdw_type_count(j)+(vdw_type_count
(i)-lig_type_ct(i))*lig_type_ct(j)))

     vir_vdw_longrange_factor = vir_vdw_longrange_factor + 2.d0 * PI *
sumv * (vdw_type_count(i)*vdw_type_count(j) &
    -(1-
vdw_longrange_lambda)*(lig_type_ct(i)*vdw_type_count(j)+(vdw_type_count
(i)-lig_type_ct(i))*lig_type_ct(j)))
   end do
 end do

 return

end subroutine am_vdw_longrange_factor
```

References

Allinger, N. L., Yuh, Y. H., & Lii, J.-H. (1989). Molecular Mechanics. The MM3 Force Field for Hydrocarbons. *Journal of the American Chemical Society, 111*(23), 8551-8566.

Aqvist, J. (1996). Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *Journal of Computational Chemistry, 17*(14), 1587-1597.

Aqvist, J., Medina, C., & Samuelsson, J. E. (1994). NEW METHOD FOR PREDICTING BINDING-AFFINITY IN COMPUTER-AIDED DRUG DESIGN. *Protein Engineering, 7*(3), 385-391.

Asthagiri, D., Pratt, L. R., Paulaitis, M. E., & Rempe, S. B. (2004). Hydration structure and free energy of biomolecularly specific aqueous dications, including Zn2+ and first transition row metals. *Journal of the American Chemical Society, 126*(4), 1285-1289.

Bennett, C. H. (1976). Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *Journal of Computational Physics, 22*(2), 245-268.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984a). Molecular Dynamics with Coupling to an External Bath. *Journal of Chemical Physics, 81*(8), 3684-3690.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984b). Molecular Dynamics with Coupling to an External Bath. *Journal of Chemical Physics, 81*, 3684-3690.

Berne, B. J., & Straub, J. E. (1997). Novel methods of sampling phase space in the simulation of biological systems. *Current Opinion in Structural Biology, 7*(2), 181-189.

Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R., & Vangunsteren, W. F. (1994). Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. *Chemical Physics Letters, 222*(6), 529-539.

Beutler, T. C., & Vangunsteren, W. F. (1994). THE COMPUTATION OF A POTENTIAL OF MEAN FORCE - CHOICE OF THE BIASING POTENTIAL IN THE UMBRELLA SAMPLING TECHNIQUE. *Journal of Chemical Physics, 100*(2), 1492-1497.

Boresch, S., Tettinger, F., Leitgeb, M., & Karplus, M. (2003). Absolute binding free energies: A quantitative approach for their calculation. *Journal of Physical Chemistry B, 107*(35), 9535-9551.

Brandsdal, B. O., Osterberg, F., Almlof, M., Feierberg, I., Luzhkov, V. B., & Aqvist, J. (2003a). Free energy calculations and ligand binding. *Protein Simulations, 66*, 123.

Brandsdal, B. O., Osterberg, F., Almlof, M., Feierberg, I., Luzhkov, V. B., & Aqvist, J. (2003b). Free energy calculations and ligand binding. *Protein Simulations, 66*, 123-+.

Burykin, A., Schutz, C. N., Villa, J., & Warshel, A. (2002). Simulations of ion current in realistic models of ion channels: The KcsA potassium channel. *Proteins-Structure Function and Genetics, 47*(3), 265-280.

Caminiti, R., Licheri, G., Piccaluga, G., & Pinna, G. (1977). X-RAY-DIFFRACTION STUDY OF A 3-ION AQUEOUS-SOLUTION. *Chemical Physics Letters, 47*(2), 275-278.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry, 26*(16), 1668-1688.

Case, D. A., Darden, T. A., Cheatham, I., T. E., , Simmerling, C. L., Wang, J., Duke, R. E., et al. (2006). *AMBER 9, University of California, San Francisco.*

Chaturvedi, U. C., & Shrivastava, R. (2005). Interaction of viral proteins with metal ions: role in maintaining the structure and functions of viruses. *Fems Immunology and Medical Microbiology, 43*(2), 105-114.

Cheng, X. L., Cui, G. L., Hornak, V., & Sinnnerling, C. (2005). Modified replica exchange simulation methods for local structure refinement. *Journal of Physical Chemistry B, 109*(16), 8220-8230.

Chipot, C., & Pohorille, A. (2007). *Free Energy Calculations. Theory and Applications in Chemistry and Biology*: Springer.

Cieplak, P., Dupradeau, F. Y., Duan, Y., & Wang, J. M. (2009). Polarization effects in molecular mechanical force fields. *Journal of Physics-Condensed Matter, 21*(33).

Dang, L. X. (2002). Computational study of ion binding to the liquid interface of water. *Journal of Physical Chemistry B, 106*(40), 10388-10394.

Darden, T. (2008). Extensions of the Ewald method for Coulomb interactions in crystals In U. Shmueli (Ed.), *International Tables for Crystallography, Vol. B, Ed. 3*: Springer-Verlag

Deng, Y. Q., & Roux, B. (2006). Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *Journal of Chemical Theory and Computation, 2*(5), 1255-1273.

Eisenberg, D., & McLachlan, A. D. (1986). SOLVATION ENERGY IN PROTEIN FOLDING AND BINDING. *Nature, 319*(6050), 199-203.

Essex, J. W., Severance, D. L., TiradoRives, J., & Jorgensen, W. L. (1997). Monte Carlo simulations for proteins: Binding affinities for trypsin-benzamidine complexes via free-energy perturbations. *Journal of Physical Chemistry B, 101*(46), 9663-9669.

Florian, J., Goodman, M. F., & Warshel, A. (2002). Theoretical investigation of the binding free energies and key substrate-recognition components of the replication fidelity of human DNA polymerase beta. *Journal of Physical Chemistry B, 106*(22), 5739-5753.

Frisch, M. J. e. a. (2003). Gaussian 03. *(Pittsburgh, PA: Gaussian Inc.)*.

Fujitani, H., Tanida, Y., Ito, M., Jayachandran, G., Snow, C. D., Shirts, M. R., et al. (2005). Direct calculation of the binding free energies of FKBP ligands. *Journal of Chemical Physics, 123*(8), 5.

Fujitani, H., Tanida, Y., Ito, M., Shirts, R., Snow, C. D., Jayachandran, G., et al. (2005). Binding energy calculation for FKBP receptor and ligands by generalized bar method. *Abstracts of Papers of the American Chemical Society, 229*, U757-U757.

Fulton, J. L., Heald, S. M., Badyal, Y. S., & Simonson, J. M. (2003). Understanding the effects of concentration on the solvation structure of Ca2+ in aqueous solution. I: The perspective on local structure from EXAFS and XANES. *Journal of Physical Chemistry A, 107*(23), 4688-4696.

Geerke, D. P., van Gunsteren, W. F., & Sk. (2007). Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *Journal of Physical Chemistry B, 111*(23), 6425-6436.

Gilson, M. K., Given, J. A., Bush, B. L., & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophysical Journal, 72*(3), 1047-1069.

Gilson, M. K., & Zhou, H. X. (2007). Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure, 36*, 21-42.

Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie-International Edition, 41*(15), 2645-2676.

Grater, F., Schwarzl, S. M., Dejaegere, A., Fischer, S., & Smith, J. C. (2005). Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *Journal of Physical Chemistry B, 109*(20), 10474-10483.

Grossfield, A. (2005). Dependence of ion hydration on the sign of the ion's charge. *Journal of Chemical Physics, 122*(2), -.

Grossfield, A., Ren, P. Y., & Ponder, J. W. (2003). Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society, 125*(50), 15671-15682.

Hamelberg, D., & McCammon, J. A. (2004). Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *Journal of the American Chemical Society, 126*(24), 7683-7689.

Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics, 120*(24), 11919-11929.

Hansmann, U. H. E. (1997). Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters, 281*(1-3), 140-150.

Hensen, C., Hermann, J. C., Nam, K. H., Ma, S. H., Gao, J. L., & Holtje, H. D. (2004). A combined QM/MM approach to protein-ligand interactions: Polarization effects of

the HIV-1 protease on selected high affinity inhibitors. *Journal of Medicinal Chemistry, 47*(27), 6673-6680.
Hermans, J., & Wang, L. (1997). Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *Journal of the American Chemical Society, 119*(11), 2707-2714.
Hewish, N. A., Neilson, G. W., & Enderby, J. E. (1982). ENVIRONMENT OF CA-2+ IONS IN AQUEOUS SOLVENT. *Nature, 297*(5862), 138-139.
Hofer, T. S., Tran, H. T., Schwenk, C. F., & Rode, B. M. (2004). Characterization of dynamics and reactivities of solvated ions by ab initio simulations. *Journal of Computational Chemistry, 25*(2), 211-217.
Hurst, T., Heritage, T. W., & Clark, R. D. (1998). HQSAR - Predicting biological activity using molecular holograms. *Abstracts of Papers of the American Chemical Society, 215*, 038-COMP.
Illingworth, C. J., & Domene, C. (2009). Many-body effects and simulations of potassium channels. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences, 465*(2106), 1701-1716.
Ivano Bertini, H. S., Astrid Sigel. (2001). *Handbook on Metalloproteins* (1 ed.): CRC.
Jager, R., & Kast, S. M. (2001). Fast prediction of hydration free energies from molecular interaction fields. *Journal of Molecular Graphics & Modelling, 20*(2), 123-131.
Jalilehvand, F., Spangberg, D., Lindqvist-Reis, P., Hermansson, K., Persson, I., & Sandstrom, M. (2001). Hydration of the calcium ion. An EXAFS, large-angle X-ray scattering, and molecular dynamics simulation study. *Journal of the American Chemical Society, 123*(3), 431-441.
Jiang, H., Jordan, K. D., & Taylor, C. E. (2007). Molecular dynamics simulations of methane hydrate using polarizable force fields. *Journal of Physical Chemistry B, 111*(23), 6486-6492.
Jiang, H., Jordan, K. D., Taylor, C. E., & Sk. (2007). Molecular dynamics simulations of methane hydrate using polarizable force fields. *Journal of Physical Chemistry B, 111*(23), 6486-6492.
Jiao, D., Golubkov, P. A., Darden, T. A., & Ren, P. (2008). Calculation of protein-ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences of the United States of America, 105*(17), 6290-6295.
Jiao, D., King, C., Grossfield, A., Darden, T. A., & Ren, P. Y. (2006). Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. *Journal of Physical Chemistry B, 110*(37), 18553-18559.
Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science, 303*(5665), 1813-1818.
Jorgensen, W. L., & Ei. (2004). The many roles of computation in drug discovery. *Science, 303*(5665), 1813-1818.
Jorgensen, W. L., & Ravimohan, C. (1985). Monte Carlo Simulation of Differences in Free Energies of Hydration. *Journal of Chemical Physics, 83*, 3050-3054.

Kaminski, G., Duffy, E. M., Matsui, T., & Jorgensen, W. L. (1994). FREE-ENERGIES OF HYDRATION AND PURE LIQUID PROPERTIES OF HYDROCARBONS FROM THE OPLS ALL-ATOM MODEL. *Journal of Physical Chemistry, 98*(49), 13077-13082.

Karplus, M. (2003). Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers, 68*(3), 350-358.

Kato, M., Braun-Sand, S., & Warshel, A. (2008). Challenges and Progresses in Calculations of Binding Free Energies - What Does it Take to Quantify Electrostatic Constributions to Protein-Ligand Interactions. In R. Stroud & J. Finer-Moore (Eds.), *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions* (pp. 268-290): RSC.

Kato, M., & Warshel, A. (2005). Through the channel and around the channel: Validating and comparing microscopic approaches for the evaluation of free energy profiles for ion penetration through ion channels. *Journal of Physical Chemistry B, 109*(41), 19516-19522.

Katz, B. A., Elrod, K., Luong, C., Rice, M. J., Mackman, R. L., Sprengeler, P. A., et al. (2001). A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *Journal of Molecular Biology, 307*(5), 1451-1486.

Katz, B. A., Finermoore, J., Mortezaei, R., Rich, D. H., & Stroud, R. M. (1995). Episelection - Novel K-I-Similar-to Nanomolar Inhibitors of Serine Proteases Selected by Binding or Chemistry on an Enzyme Surface. *Biochemistry, 34*(26), 8264-8280.

Katz, B. A., Mackman, R., Luong, C., Radika, K., Martelli, A., Sprengeler, P. A., et al. (2000). Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chemistry & Biology, 7*(4), 299-312.

Khoruzhii, O., Donchev, A. G., Galkin, N., Illarionov, A., Olevanov, M., Ozrin, V., et al. (2008). Application of a polarizable force field to calculations of relative protein-ligand binding affinities. *Proceedings of the National Academy of Sciences of the United States of America, 105*(30), 10378-10383.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research, 33*(12), 889-897.

Kravtsov, A. A., Karpov, P. V., Baskin, II, Palyulin, V. A., & Zefirov, N. S. (2007). "Bimolecular" QSPR: Estimation of the solvation free energy of organic molecules in different solvents. *Doklady Chemistry, 414*, 128-131.

Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., & Rosenberg, J. M. (1992). THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD. *Journal of Computational Chemistry, 13*(8), 1011-1021.

Lamb, M. L., Jorgensen, W. L., & Yw. (1997). Computational approaches to molecular recognition. *Current Opinion in Chemical Biology, 1*(4), 449-457.

Lee, F. S., Chu, Z. T., Bolger, M. B., & Warshel, A. (1992). CALCULATIONS OF ANTIBODY ANTIGEN INTERACTIONS - MICROSCOPIC AND SEMIMICROSCOPIC EVALUATION OF THE FREE-ENERGIES OF BINDING OF PHOSPHORYLCHOLINE ANALOGS TO MCPC603. *Protein Engineering, 5*(3), 215-228.

Lei, H. X., & Duan, Y. (2007). Improved sampling methods for molecular simulation. *Current Opinion in Structural Biology, 17*(2), 187-191.

Leiros, H. K. S., Brandsdal, B. O., Andersen, O. A., Os, V., Leiros, I., Helland, R., et al. (2004). Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. *Protein Science, 13*(4), 1056-1070.

Liang, T., Walsh, T. R., & Kn. (2007). Simulation of the hydration structure of glycyl-alanine. *Molecular Simulation, 33*(4-5), 337-342.

Liang, T., Walsh, T. R., & Mu. (2006). Molecular dynamics simulations of peptide carboxylate hydration. *Physical Chemistry Chemical Physics, 8*(38), 4410-4419.

Lightstone, F. C., Schwegler, E., Allesch, M., Gygi, F., & Galli, G. (2005). A first-principles molecular dynamics study of calcium in water. *Chemphyschem, 6*(9), 1745-1749.

Lightstone, F. C., Schwegler, E., Hood, R. Q., Gygi, F., & Galli, G. (2001). A first principles molecular dynamics simulation of the hydrated magnesium ion. *Chemical Physics Letters, 343*(5-6), 549-555.

Liu, P., Kim, B., Friesner, R. A., & Berne, B. J. (2005). Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America, 102*(39), 13749-13754.

Liu, W. B., Sakane, S., Wood, R. H., & Doren, D. J. (2002). The hydration free energy of aqueous Na+ and Cl- at high temperatures predicted by ab initio/classical free energy perturbation: 973 K with 0.535 g/cm(3) and 573 K with 0.725 g/cm(3). *Journal of Physical Chemistry A, 106*(7), 1409-1418.

Liwo, A., Czaplewski, C., Oldziej, S., & Scheraga, H. A. (2008). Computational techniques for efficient conformational sampling of proteins. *Current Opinion in Structural Biology, 18*(2), 134-139.

Lopes, P. E. M., Roux, B., & MacKerell, A. D. (2009). Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theoretical Chemistry Accounts, 124*(1-2), 11-28.

Lyubartsev, A. P., Laasonen, K., & Laaksonen, A. (2001). Hydration of Li+ ion. An ab initio molecular dynamics simulation. *Journal of Chemical Physics, 114*(7), 3120-3126.

Mares-Guia, M. (1965). Studies on Active Center of Trypsin - Binding of Amidines and Guanidines as Models of Substrate Side Chain. *Journal of Biological Chemistry, 240*, 1579.

Maresguia, M., Nelson, D. L., & Rogana, E. (1977). ELECTRONIC EFFECTS IN INTERACTION OF PARA-SUBSTITUTED BENZAMIDINES WITH TRYPSIN - INVOLVEMENT OF PI-ELECTRONIC DENSITY AT CENTRAL ATOM OF SUBSTITUENT IN BINDING. *Journal of the American Chemical Society, 99*(7), 2331-2336.

Martin, R. L., Hay, P. J., & Pratt, L. R. (1998). Hydrolysis of ferric ion in water and conformational equilibrium. *Journal of Physical Chemistry A, 102*(20), 3565-3573.

Martinez, J. M., Pappalardo, R. R., & Marcos, E. S. (1999). First-principles ion-water interaction potentials for highly charged monatomic cations. Computer simulations of Al3+, Mg2+, and Be2+ in water. *Journal of the American Chemical Society, 121*(13), 3175-3184.

Masia, M., Probst, M., & Rey, R. (2004). On the performance of molecular polarization methods. I. Water and carbon tetrachloride close to a point charge. *Journal of Chemical Physics, 121*(15), 7362-7378.

Megyes, T., Grosz, T., Radnai, T., Bako, I., & Palinkas, G. (2004). Solvation of calcium ion in polar solvents: An X-ray diffraction and ab initio study. *Journal of Physical Chemistry A, 108*(35), 7261-7271.

Miller, W. H., Hernandez, R., Handy, N. C., Jayatilaka, D., & Willetts, A. (1990). Abinitio Calculation of Anharmonic Constants for a Transition-State, with Application to Semiclassical Transition-State Tunneling Probabilities. *Chemical Physics Letters, 172*(1), 62-68.

Mobley, D. L., Bayly, C. I., Cooper, M. D., Shirts, M. R., & Dill, K. A. (2009). Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *Journal of Chemical Theory and Computation, 5*(2), 350-358.

Ota, N., Stroupe, C., Ferreira-da-Silva, J. M. S., Shah, S. A., Mares-Guia, M., & Brunger, A. T. (1999). Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: Relative free energy of binding of trypsin to benzamidine and benzylamine. *Proteins-Structure Function and Genetics, 37*(4), 641-653.

P. Hohenberg, W. K. (1964). Inhomogeneous Electron Gas. *Physical Review, 136*(3B), 8.

Peters, M. B., & Merz, K. M. (2006). Semiempirical comparative binding energy analysis (SE-COMBINE) of a series of trypsin inhibitors. *Journal of Chemical Theory and Computation, 2*(2), 383-399.

Ponder, J. W. (2001). TINKER: Software Tools for Molecular Design, Version 3.9;. Saint Louis, MO.

Ponder, J. W. (2004). TINKER: Software Tools for Molecular Design, version 4.2; Saint Louis, MO, http://dasher.wustl.edu/tinker/.

Ponder, J. W. (2006). TINKER: Software Tools for Molecular Design *Washington University Medical School* *http://dasher.wustl.edu/tinker/*.

Ponder, J. W., & Case, D. A. (2003a). Force fields for protein simulations. In *Protein Simulations* (Vol. 66, pp. 27-+). San Diego: Academic Press Inc.

Ponder, J. W., & Case, D. A. (2003b). Force fields for protein simulations. *Advances in Protein Chemistry, 66*, 27-85.

Radmer, R. J., & Kollman, P. A. (1998). The application of three approximate free energy calculations methods to structure based ligand design: Trypsin and its complex with inhibitors. *Journal of Computer-Aided Molecular Design, 12*(3), 215-227.

Rempe, S. B., Asthagiri, D., & Pratt, L. R. (2004a). *Inner shell definition and absolute hydration free energy of K+(aq) on the basis of quasi-chemical theory and ab initio molecular dynamics*.

Rempe, S. B., Asthagiri, D., & Pratt, L. R. (2004b). Inner shell definition and absolute hydration free energy of K+(aq) on the basis of quasi-chemical theory and ab initio molecular dynamics. *Physical Chemistry Chemical Physics, 6*(8), 1966-1969.

Rempe, S. B., & Pratt, L. R. (2001). *The hydration number of Na+ in liquid water*.

Ren, P., Ponder, JW. (2003). Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B, 107* 5933-5947.

Ren, P. Y., & Ponder, J. W. (2002). Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry, 23*(16), 1497-1506.

Ren, P. Y., & Ponder, J. W. (2003). Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B, 107*(24), 5933-5947.

Ren, P. Y., & Ponder, J. W. (2004). Temperature and pressure dependence of AMOEBA water model. *Journal of Physical Chemistry B, 108*(35), 13427-13437.

Ren, P. Y., Ponder, J. W., & Mz. (2002). Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry, 23*(16), 1497-1506.

Resat, H., Marrone, T. J., & McCammon, J. A. (1997). Enzyme-inhibitor association thermodynamics: Explicit and continuum solvent studies. *Biophysical Journal, 72*(2), 522-532.

Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annual Review of Biophysics and Biomolecular Structure, 28*, 155-179.

Sagui, C., Pedersen, L. G., & Darden, T. A. (2004). Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *Journal of Chemical Physics, 120*(1), 73-87.

Sagui, C., Pomorski, P., Darden, T. A., & Roland, C. (2004). Ab initio calculation of electrostatic multipoles with Wannier functions for large-scale biomolecular simulations. *Journal of Chemical Physics, 120*(9), 4530-4544.

Sandberg, L., Casemyr, R., & Edholm, O. (2002). Calculated hydration free energies of small organic molecules using a nonlinear dielectric continuum model. *Journal of Physical Chemistry B, 106*(32), 7889-7897.

Schmid, R., Miah, A. M., & Sapunov, V. N. (2000). A new table of the thermodynamic quantities of ionic hydration: values and some applications (enthalpy-entropy compensation and Born radii). *Physical Chemistry Chemical Physics, 2*(1), 97-102.

Schnieders, M. J., Baker, N. A., Ren, P. Y., & Ponder, J. W. (2007). Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. *Journal of Chemical Physics, 126*(12), 124114.

Schwarzl, S. M., Tschopp, T. B., Smith, J. C., & Fischer, S. (2002). Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? *Journal of Computational Chemistry, 23*(12), 1143-1149.

Schwarzl, S. M., Tschopp, T. B., Smith, J. C., Fischer, S., & Er. (2002). Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? *Journal of Computational Chemistry, 23*(12), 1143-1149.

Schwenk, C. F., Loeffler, H. H., & Rode, B. M. (2001). Molecular dynamics simulations of Ca2+ in water: Comparison of a classical simulation including three-body corrections and Born-Oppenheimer ab initio and density functional theory quantum mechanical/molecular mechanics simulations. *Journal of Chemical Physics, 115*(23), 10808-10813.

Sham, Y. Y., Chu, Z. T., Tao, H., & Warshel, A. (2000). Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins-Structure Function and Bioinformatics, 39*(4), 393-407.

Shirts, M. R., & Pande, V. S. (2005). Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *Journal of Chemical Physics, 122*(14), 16.

Shivakumar, D., Deng, Y. Q., & Roux, B. (2009). Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *Journal of Chemical Theory and Computation, 5*(4), 919-930.

Sitkoff, D., Sharp, K. A., & Honig, B. (1994). ACCURATE CALCULATION OF HYDRATION FREE-ENERGIES USING MACROSCOPIC SOLVENT MODELS. *Journal of Physical Chemistry, 98*(7), 1978-1988.

Spangberg, D., & Hermansson, K. (2004). Many-body potentials for aqueous Li+, Na+, Mg2+, and Al3+: Comparison of effective three-body potentials and polarizable models. *Journal of Chemical Physics, 120*(10), 4829-4843.

Stone, A. J. (2005). Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation, 1*(6), 1128-1132.

Talhout, R., & Engberts, J. B. F. N. (2001). Thermodynamic analysis of binding of p-substituted benzamidines to trypsin. *European Journal of Biochemistry, 268*(6), 1554-1560.

Talhout, R., & Engberts, J. B. F. N. (2004). Probing the effect of the amidinium group and the phenyl ring on the thermodynamics of binding of benzamidinium chloride to trypsin. *Organic & Biomolecular Chemistry, 2*(21), 3071-3074.

Talhout, R., Villa, A., Mark, A. E., & Engberts, J. B. F. N. (2003). Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *Journal of the American Chemical Society, 125*(35), 10570-10579.

Thole, B. T., & Md. (1981). MOLECULAR POLARIZABILITIES CALCULATED WITH A MODIFIED DIPOLE INTERACTION. *Chemical Physics, 59*(3), 341-350.

Torrie, G. M., & Valleau, J. P. (1977). NON-PHYSICAL SAMPLING DISTRIBUTIONS IN MONTE-CARLO FREE-ENERGY ESTIMATION - UMBRELLA SAMPLING. *Journal of Computational Physics, 23*(2), 187-199.

Toukmaji, A., Sagui, C., Board, J., & Darden, T. (2000). Efficient Particle-Mesh Ewald Based Approach to Fixed and Induced Dipolar Interactions. *Journal of Chemical Physics, 113*(24), 10913-10927.

Trylska, J., Tozzini, V., & McCammon, J. A. (2005). Exploring global motions and correlations in the ribosome. *Biophysical Journal, 89*(3), 1455-1463.

Tuma, L., Jenicek, D., Jungwirth, P., & Cx. (2005). Propensity of heavier halides for the water/vapor interface revisited using the Amoeba force field. *Chemical Physics Letters, 411*(1-3), 70-74.

Viswanadhan, V. N., Ghose, A. K., Singh, U. C., & Wendoloski, J. J. (1999). *Prediction of solvation free energies of small organic molecules: Additive-constitutive models based on molecular fingerprints and atomic constants*.

Viswanadhan, V. N., Ghose, A. K., & Wendoloski, J. J. (2000). Estimating aqueous solvation and lipophilicity of small organic molecules: A comparative overview of atom/group contribution methods. *Perspectives in Drug Discovery and Design, 19*(1), 85-98.

Wang, J. Y., Deng, Y. Q., & Roux, B. (2006). Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical Journal, 91*(8), 2798-2814.

Wang, J. Y., & Roux, B. (2005). Absolute free energy calculations for the interactions of FKBP with ligands using the GSBP method. *Biophysical Journal, 88*(1), 517a-517a.

Wang, W., Wang, J., & Kollman, P. A. (1999). What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins-Structure Function and Genetics, 34*(3), 395-402.

Warshel, A., Kato, M., & Pisliakov, A. V. (2007). Polarizable force fields: History, test cases, and prospects. *Journal of Chemical Theory and Computation, 3*(6), 2034-2045.

Warshel, A., & Levitt, M. (1976). THEORETICAL STUDIES OF ENZYMIC REACTIONS - DIELECTRIC, ELECTROSTATIC AND STERIC STABILIZATION OF CARBONIUM-ION IN REACTION OF LYSOZYME. *Journal of Molecular Biology, 103*(2), 227-249.

Warshel, A., Sharma, P. K., Kato, M., & Parson, W. W. (2006). Modeling electrostatic effects in proteins. *Biochimica Et Biophysica Acta-Proteins and Proteomics, 1764*(11), 1647-1676.

Warshel, A., Sussman, F., & Hwang, J. K. (1988). EVALUATION OF CATALYTIC FREE-ENERGIES IN GENETICALLY MODIFIED PROTEINS. *Journal of Molecular Biology, 201*(1), 139-159.

Woo, H. J., & Roux, B. (2005). Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America, 102*(19), 6825-6830.

Yague, J. I., Mohammed, A. M., Loeffler, H. H., & Rode, B. M. (2003). MD and MC simulations of hydrated manganous ion including three-body effects. *Journal of Molecular Structure-Theochem, 620*(1), 15-20.

VITA

Dian Jiao attended the Southeast University in China in 1998. He graduated in 2002 with the degree of Bachelor of Science from the Biomedical Engineering department. He joined the graduate school of Southeast University in 2002 and was granted the degree of Master of Science in 2005. In September 2005, he entered the Biomedical Engineering department at the University of Texas at Austin.

Permanent Address: 129 Gaotang Rd #41 Apt 101, Ningbo, China 315000.

This manuscript was typed by the author.