The Dissertation Committee for Yue Shi Certifies that this is the approved version of the following dissertation:

Modeling the Interaction and Energetics of Biological Molecules with a Polarizable Force Field

Committee:

Pengyu Ren, Supervisor

Ron Elber, Co-Supervisor

Aaron Baker

Kevin Dalby

Stephen Martin

Modeling the Interaction and Energetics of Biological Molecules with a Polarizable Force Field

by

Yue Shi, B.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

May 2013

Dedication

To my family

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Pengyu Ren, for his continuous support of my study and research. I could not complete this thesis without his guidance and encouragement. He is the greatest mentor I can imagine.

I would like to thank my committee members Dr. Ron Elber, Dr. Kevin Dalby, Dr. Stephen Martin, and Dr. Aaron Baker for their guidance and acknowledgement of the value of my work. Dr. Elber has lectured many biophysics and statistical mechanics courses in which I learned important concepts and principles in my later work. He also kindly offered a lot of insightful advices on my research. Our lab has been working with Dr. Dalby and Dr. Martin closely. I am lucky to have the opportunity to get involved in these collaborations and I am excited to see how the computational techniques are applied to solve the practical problems. Dr. Baker is a great professor, and I thank him for giving helpful suggestions on my dissertation.

I would like to give my thanks to Dr. Johnny Wu, Dr. John Fonner, Dr. Qiantao Wang and Dr. Tianyi Yang for their great help in my research. I would especially like to thank Dr. Chunli Yan, Dr. Michael Schnieders, Dr. Jiajing Zhang and Dr. Dian Jiao for many helpful technical discussions, valuable suggestions and comments about this dissertation. I want to thank my labmates in Computational Biomolecular Engineering Lab, Zhen Xia, Xiaojia Mu, Danny Dykstra, and David Bell for their assistance and advices. My studies at the University of Texas at Austin were enriched immeasurably by the friendship and technical knowledge of my colleagues.

Finally, I would like to express my deepest gratitude to my family for their constant love, understanding, encouragement, and support.

Modeling the Interaction and Energetics of Biological Molecules with a Polarizable Force Field

Yue Shi, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Pengyu Ren

Accurate prediction of protein-ligand binding affinity is essential to computational drug discovery. Current approaches are limited by the accuracy of the underlying potential energy model that describes atomic interactions. A more rigorous physical model is critical for evaluating molecular interactions to chemical accuracy. The objective of this thesis research is to develop a polarizable force field with an accurate representation of electrostatic interactions, and apply this model to protein-ligand recognition and to ultimately solve practical problems in computer aided drug discovery. By calculating the hydration free energies of a series of organic small molecules, an optimal protocol is established to develop the electrostatic parameters from quantum mechanics calculations. Next, the systematical development and parameterization procedure of AMOEBA protein force field is presented. The derived force field has gone through extensive validations in both gas phase and condensed phase. The last part of the thesis involves the application of AMOEBA to study protein-ligand interactions. The binding free energies of benzamidine analogs to trypsin using molecular dynamics alchemical perturbation are calculated with encouraging accuracy. AMOEBA is also used

to study the thermodynamic effect of constraining and hydrophobicity on binding energetics between phosphotyrosine(pY)-containing tripeptides and the SH2 domain of growth receptor binding protein 2 (Grb2). The underlying mechanism of an "entropic paradox" associated with ligand preorganization is explored.

Table of Contents

List of Tables

List of Figures

1 Introduction

Molecular modeling is one of the fastest growing fields in science. It combines physical and biological principles with computational techniques to simulate and understand the behavior of molecules. Common computational approaches include *ab initio* or semi-empirical quantum mechanics, empirical molecular mechanics, molecular dynamics (MD) and Monte Carlo simulations, free energy and solvation methods, structure activity relationships, chemical or biochemical information and databases, and so on. These computational techniques have been widely used in a variety of studies ranging from drug discovery, structural modeling of biomolecules to material assemblies. Combining with experimental methods, such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, molecular modeling serves as a powerful tool to understand the molecular structure, interaction, energetics, and therefore provide insights into the underlying mechanism.

Depending on the size of the systems, molecules can be represented at different scales, from highly accurate to very approximate. Quantum mechanics (QM) methods, providing the highest accuracy when used with proper theory and basis sets, are based on the solution of the Schrödinger equation which describes the motions of the electrons and nuclei in a molecular system from first principle. Molecular mechanics (MM) is a more popular and feasible tool for handling systems with significant larger number of atoms with reasonable efficiency. In MM models, the potential energy of a given conformation is represented by the sum of a series of individual energy terms including bond

stretching, angle bending, dihedral angle rotation and long range interactions. For mesosacle simulations, coarse-grained model is a good choice.

Drug development relies on identifying lead compounds with high affinity for specific targets such as receptors, enzymes, hormones and ion channels. Structure-based computer modeling of ligand-protein interactions is now a core component of modern drug discovery. Although various studies suggest that the calculations of the binding free energies based on alchemical transformation have shown reasonable agreement with experimental data, chemical accuracy can hardly be achieved.[1, 2] The underlying physical model is a major bottleneck.[1, 3] While they are widely used in many areas of biological and materials sciences, several aspects of classical force fields require closer examination, especially the fixed atomic-charge based electrostatic model. For ligand-protein binding systems, where ligand may experience significant change in environment, it is particularly important to have a more rigorous description in electrostatic interactions and take polarization effect into account.

1.1 Polarization Effects in Molecular Mechanical Force Fields

Molecular mechanics based modeling has been widely used in the study of chemical and biological systems. The classical potential energy functions and their parameters are referred to as force fields. Empirical force fields for biomolecules emerged in the early 1970's[4, 5], followed by the first molecular dynamics simulations of the bovine pancreatic trypsin inhibitors (BPTI).[6-8] Over the past 30 years, a great number of empirical molecular mechanics force fields, including AMBER,[9] CHARMM,[10] GROMOS,[11] OPLS,[12] and many others, have been developed. These force fields share similar functional forms, including valence interactions

represented by harmonic oscillators, point dispersion-repulsion for van der Waals (vdW) interactions, and an electrostatic contribution based on fixed atomic partial charges. This generation of molecular mechanics force fields has been widely used in the study of molecular structures, dynamics, interactions, design and engineering. We refer interested readers to some recent reviews for detailed discussions.[13, 14]

Although the fixed charge force fields enjoyed great success in many areas there remains much room for improvement. In fixed charge based electrostatic models, the atomic partial charges are meant to be "pre-polarized" for condensed phases in an averaged fashion, typically achieved by the fortuitous overestimation of electrostatic charges by low-level ab initio quantum mechanics. Such models thus lack the ability to describe the variation in electrostatics due to many-body polarization effects, which have been shown to be a significant component of intermolecular forces.[14-16] With the rapid growth of computational resources, there has been increasing effort to explicitly incorporate many-body induction into molecular mechanics to improve the accuracy of molecular modeling.

Classical electrostatics models that take into account polarization appeared as early as the 1950s. Barker in his 1953 paper "Statistical Mechanics of Interacting Dipoles" discussed the electrostatic energy of molecules in terms of "permanent and induced dipoles".[17] Currently, polarizable models generally fall into three categories, those based on induced point dipoles,[13, 18-27] the classical Drude oscillators,[28-30] or fluctuating charges.[31-33] More sophisticated force fields that are "electronic structure-based" [34] or use "machine learning methods"[35] also exist, but incur higher computational costs. Compared to fixed charge models, the polarizable models are still in

a relatively early stage. Only in the past decade or so, has there been a systematic effort to develop general polarizable force fields for molecular modeling. A number of reviews have been published to discuss various aspects of polarizable force fields and their development.[13, 36-42] Discussions of the advantages and disadvantages of each model and their applications will be presented in the following sections.

1.2 MODELING POLARIZATION EFFECT

1.2.1 Induced Dipole Models

To describe electrostatic interactions involving polarization, we will consider a system consisting of a collection of charge distribution sites located at lone-pair positions, atomic centers and/or molecular centers, depending on the resolution of the model. The total charge distribution at site *i* is the sum of permanent and induced charge

$$\mathrm{M_i} = \mathrm{M_i^0} + \mathrm{M_i^{ind}} \tag{1.1}$$

where ***M*** represents the charge distribution. This distribution can be a simple point charge, a point multipole expansion with charge, dipole, quadrupole and/or higher order moments, or a continuous charge distribution. While the principles described below are not limited to any particular representation of charge distribution, we will use point multipoles for convenience.

The electrostatic interaction energy between two charge sites *i* and *j* is given by

$$\mathrm{U_{ele}} = \frac{1}{2}\sum_{\mathrm{i}}\sum_{\mathrm{j}\neq\mathrm{i}} \mathrm{M_i^t T_{ij} M_j} \tag{1.2}$$

where ***T*** is the interaction operator and is a function of the distance between *i* and *j*. In the case of point charge interactions, ***T*** simply equals 1/r. The work (positive energy) needed

to polarize a charge distribution also has a quadratic dependence on the induced charge distribution:

$$U_{work} = \frac{1}{2}\sum_i \left(M_i^{ind}\right)^t \alpha_i^{-1} M_i^{ind} \quad (1.3)$$

where $\boldsymbol{\alpha}$ is the polarizability of site *i* that includes all orders of polarizability including dipole polarizability.[43] Although $\boldsymbol{\alpha}$ is in generally treated as an isotropic quantity, as in the Applequist scheme [43], *ab initio* anisotropic polarizability tensors can be derived from quantum mechanical calculations.[44, 45]

The total energy is the sum of the electrostatic energy and work spent

$$U_{ele} = \frac{1}{2}\sum_i \sum_{j\neq i} M_i^t T_{ij} M_j + \frac{1}{2}\sum_i \left(M_i^{ind}\right)^t \alpha_i^{-1} M_i^{ind} \quad (1.4)$$

The values of the induced moments minimize the total energy, by satisfying

$$\frac{\partial U_{ele}}{\partial M_i^{ind}} = \sum_{j\neq i} T_{ij} M_j + \alpha_i^{-1} M_i^{ind} = 0 \quad (1.5)$$

As a result

$$M_i^{ind} = \alpha_i^{-1} \sum_{j\neq i} T_{ij}\left(M_j^0 + M_j^{ind}\right) \quad (1.6)$$

The above equation can be solved iteratively to obtain the induced dipoles. The self-consistent calculation is computational expensive, however can be accelerated with predictors and non-stationary iterative methods.[4]

Substituting $\alpha_i^{-1} M_i^{ind}$ from Eq (1.5) into Eq (1.4), the final electrostatic energy becomes

$$U_{ele} = \frac{1}{2}\sum_i \sum_{j\neq i} \left(\mathbf{M_i^0}\right)^{\boldsymbol{t}} \mathbf{T}_{ij} \mathbf{M}_j^0 + \frac{1}{2}\sum_i \sum_{j\neq i} \left(\mathbf{M}_i^{ind}\right)^t \mathbf{T}_{ij} \mathbf{M}_j^0 \quad (1.7)$$

where the first term is the permanent electrostatic energy and the second term is the polarization energy.

1.2.2 Classic Drude Oscillators

In the Drude oscillator model, the polarization effect is described by a point charge (the Drude oscillator) attached to each non-hydrogen atom via a harmonic spring. The point charge can move relative to the attachment site in response to the electrostatic environment. The electrostatic energy is the sum of the pairwise interactions between atomic charges and the partial charge of the Drude particles.

$$\mathrm{E_{ele}} = \sum_{\mathrm{A<}B}^{\mathrm{N}} \frac{\mathrm{q_c(A)q_c(B)}}{|\mathrm{r_c(A)-r_c(B)}|} + \sum_{\mathrm{A<}B}^{\mathrm{N,N_D}} \frac{\mathrm{q_D(A)q_c(B)}}{|\mathrm{r_D(A)-r_c(B)}|} + \sum_{\mathrm{A<}B}^{\mathrm{N_D}} \frac{\mathrm{q_D(A)q_D(B)}}{|\mathrm{r_D(A)-r_D(B)}|} + \frac{1}{2}\sum_{\mathrm{A}}^{\mathrm{N_D}} \mathrm{k_D(r_D(A)-r_c(B))^2} \quad (1.8)$$

where N_D and N are the number of Drude particles and non-hydrogen atoms, q_D and q_C are the charges on the Drude particle and its parent atom, respectively; r_D and r_C are their respective positions, and k_D is the force constant of the harmonic spring between the Drude oscillator and its parent atom. The last term in the above equation accounts for the cost of polarizing the Drude particles.

The atomic polarizability (α) is a function of both the partial charge on the Drude particle and the force constant of the spring

$$\alpha = \frac{q_D^2(A)}{k_D} \quad (1.9)$$

Both the induced-dipole and Drude oscillator approaches benefit from short-range Thole damping to avoid a polarization catastrophe and to produce an anisotropic molecular polarization response.[46]

1.2.3 Fluctuating Charges

The formalism of the fluctuating charge model is based on the charge equilibration (CHEQ) method,[47] in which the chemical potential is equilibrated via the

redistribution of charge density. The charge-dependent energy for a system of M molecules containing N_i atoms per molecule is expressed as

$$\mathrm{E_{CHEQ}}(R,Q) = \sum_{\mathrm{i}=1}^{M}\sum_{\alpha=1}^{\mathrm{N}} \chi_{i\alpha} Q_{i\alpha} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\alpha=1}^{N_i}\sum_{\beta=1}^{N_j} J_{i\alpha i\beta} Q_{i\alpha} Q_{j\beta} + \frac{1}{2}\sum_{i=1}^{MN'}\sum_{j=1}^{MN'} \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{j=1}^{M} \lambda_i \left(\sum_{i=1}^{N} Q_{ij} - Q_j^{Total}\right) \quad (1.10)$$

where Q_i is the partial charge on atomic site i. The χ describes the atomic electronegativity that controls the directionality of electron flow and J is the atomic hardness that represents the resistance to electron flow or from the atom. These parameters are optimized to reproduce molecular dipoles and the molecular polarization response. The charge degrees of freedom are typically propagated via an extended Lagrangian formulation:[48]

$$\mathrm{L} = \sum_{\mathrm{i}=1}^{\mathrm{M}}\sum_{\alpha=1}^{\mathrm{N}} \frac{1}{2}\mathrm{m_{i\alpha}}\left(\frac{\mathrm{dr_{i\alpha}}}{\mathrm{dt}}\right)^2 + \sum_{\mathrm{i}=1}^{\mathrm{M}}\sum_{\alpha=1}^{\mathrm{N}} \frac{1}{2}\mathrm{m_{Q,i\alpha}}\left(\frac{\mathrm{dQ_{i\alpha}}}{\mathrm{dt}}\right)^2 - \mathrm{E(Q,r)} - \sum_{\mathrm{i}=1}^{\mathrm{M}} \lambda_{\mathrm{i}} \sum_{\alpha=1}^{\mathrm{N}} \mathrm{Q_{i\alpha}} \quad (1.11)$$

where the first two terms represent the nuclear and charge kinetic energies, the third term is the potential energy, and the fourth term is the molecular charge neutrality constraint enforced on each molecule i via a Lagrange multiplier λ_i. The extended Lagrangian approach can also be applied to the induced dipole and Drude oscillator models described earlier. While the extended Lagrangian seems to be more efficient than the iterative method, fictitious masses and smaller time-steps are required to minimize the coupling between the polarization and atomic degrees of freedom, which can never be completely eliminated.[4]

A few general force fields have been developed based on these formulas to explicitly treat the polarization effect. In the following sections, we will discuss development highlights for some of the representative force fields.

1.3 RECENT DEVELOPMENT

1.3.1 AMOEBA

The AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) force field, developed by Ponder, Ren and co-workers,[19, 22, 39] utilizes atomic multipoles to represent permanent electrostatics and induced atomic dipoles for many-body polarization. The valence interactions include bond, angle, torsion and out-of-plane contributions using typical molecular mechanics functional forms. The van der Waals interaction is described by a buffered-14-7 function. The atomic multipole moments that consist of charge, dipole and quadrupole moments, which are derived from the *ab initio* quantum mechanical calculations using procedures such as Stone's Distributed Multipole Analysis (DMA).[49-51] The higher order moments make possible anisotropic representations of the electrostatic potential outside atoms and molecules. The polarization effect is explicitly taken into account via atomic dipole induction. The combination of permanent atomic multipoles and induced dipoles enables AMOEBA to accurately capture electrostatic interactions in both gas and condensed phase. The vdW parameters of AMOEBA are optimized simultaneously optimized against both *ab initio* gas-phase data and condensed-phase experimental properties.

In the past decade, AMOEBA has been applied to the study of water model,[19] monovalent and divalent ions,[52-54] small molecules,[55, 56] peptides[22, 57] and proteins.[58-60] AMOEBA demonstrated that a polarizable force field is able to perform well in both gas and solution phases with a single set of parameters. In addition, AMOEBA is the first general-purpose polarizable force field that has been utilized in molecular dynamics simulations of protein-ligand binding and calculation of absolute and relative binding free energies.[59-63] The calculation of binding free energies between

trypsin and benzamidine derivatives suggests significant non-additive electrostatic interactions as the ligand desolvates from water and enters the protein pocket. (see Section 4.4 for further discussion) Recently, AMOEBA has been extended to the biomolecular X-ray crystallography refinement[64, 65] and consistent prediction of the structure, thermodynamic stability and solubility of organic crystals[66] with encouraging success.

To date, AMOEBA has been implemented in several widely used software packages, including TINKER,[67] OpenMM,[68] Amber,[69] and Force Field X.[70] The AMOEBA polarizable force field was first implemented within the FORTRAN based TINKER software package[71] using PME for long-range electrostatics. Implementation of the polarizable-multipole Poisson-Boltzmann,[72] which depends on APBS,[73] and generalized Kirkwood[74] continuum electrostatics models can also be found in TINKER, which is now being parallelized using OpenMP. The algorithms in TINKER are also available from within CHARMM using the MSCALE interface.[75, 76] Alternative FORTRAN implementations of AMOEBA using Particle Mesh Ewald (PME) are available in the Sander and PMEMD molecular dynamics engines of the AMBER,[69] with the latter parallelized using MPI. The PME treatment of AMOEBA electrostatics has recently been extended within the Java Runtime Environment (JRE) program *Force Field X* by incorporation of explicit support for crystal space group symmetry,[64] parallelization for heterogeneous computer hardware environments[64] and support for advanced free energy methods such as the Orthogonal Space Random Walk (OSRW) strategy.[66, 77] These advancements are critical for applications such as AMOEBA-assisted biomolecular X-ray refinement,[64, 78] efficient computation of

protein-ligand binding affinity,[58, 62] and prediction of the structure, stability and solubility of organic crystals.[66] Finally, the OpenMM software is working toward a general implementation of AMOEBA using the CUDA GPU programming language.[79]

1.3.2 SIBFA

The SIBFA (Sum of Interactions Between Fragments *Ab initio* computed) force field for small molecules and flexible proteins, developed by Gresh, Piquemal *et. al*,[80-84] is one of the most sophisticated polarizable force fields that incorporate polarization, electrostatic penetration [85] and charge-transfer effects.[86]

The polarization is treated with induced dipole model, with the distributed anisotropic polarizabilities tensors[45] placed on the bond centers and on the heteroatom lone pairs. Quadrupolar polarizabilities are used to treat metal centers. The force field is designed to enable the simultaneous and reliable computations of both intermolecular and conformational energies governing the binding specificities of biologically and pharmacologically relevant molecules. Similar to AMOEBA, permanent multipoles are used for permanent electrostatics. A flexible molecule is modeled by combining the constitutive rigid fragments. SIBFA is formulated on the basis of quantum chemistry and calibrated on energy decomposition analysis, as oppose to AMOEBA which relies more on condensed-phase experimental data. It aims to produce accurate interaction energy comparable with *ab inito* results. Currently the analytical gradients for charge-transfer energy and solvation contribution are not yet available although molecular dynamics simulations with a simplified SIBFA potential have been attempted and will be reported in the near future.

The development of SIBFA emphasizes on separability, anisotropy, nonadditivity and transferability. SIBFA has been validated on a wide range of molecular systems from water clusters[87] to large complexes such as metalloenzymes encompassing Zn(II).[88-93] It has been applied to investigate molecular recognition problems, including the binding of nucleic acids to metal ions,[94-96] the prediction of oligopeptide conformations,[97, 98] and ligand-protein binding.[99] Most of the SIBFA calculations closely reproduced the quantum chemistry results, both the interaction energy and the decomposed terms. At the same time, electrostatic parameters demonstrate its transferability between similar molecules.

As an alternative to distributed point multipole electrostatic representation, a Gaussian based electrostatic model (GEM) has been explored.[100] GEM computes the molecular interaction energies using a similar approach to SIBFA but replacing distributed multipoles by electron densities.[101] GEM is shown to better capture the short-range effects on intermolecular interaction energies, and naturally includes the penetration effect. Calculations on a few simple systems like water clusters[101] have demonstrated GEM's capability to reproduce quantum chemistry results. Furthermore, implementation of PME for GEM in PBC showed reasonable computational efficiency thanks to the use of Hermite Gaussian functions. [102] Therefore, replacing SIBFA's distributed multipoles with the GEM continuous electrostatic model will be a future direction of methodology development.[100]

1.3.1 NEMO

NEMO (Non-Empirical Molecular Orbital) is a polarizable potential developed by Karlström and co-workers.[103-105] The NEMO potential energy function is composed

of electrostatics, induction, dispersion and repulsion terms. The induction is modeled using induced point–dipole moments with recent addition of induced point–quadrupole moments.[26] The electrostatics, previously represented by atomic charges and dipoles, has also been extended to include atomic quadrupole moments. These extensions showed notable improvement on formaldehyde. The atomic multipole moments are now obtained from ab *initio* calculation using a LoProp procedure.[106] The LoProp is claimed to provide atomic multipoles and atomic polarizabilities that are less sensitive to basis sets than other methods such as DMA. Also, NEMO is the only force field that explores the possibility of including interactions between permanent multipoles and higher-order induced multipoles involving higher-order hyperpolarizabilities.[26]

NEMO has demonstrated its strength in accurate description of inter and intramolecular interactions in small systems, including glycine dipeptide conformation profiles,[107] ion-water droplets,[108] and urea transition from nonplanar to planar conformation in water.[109] Its application to biomacromolecules is no yet available.

1.3.4 CHARMM-Drude

In addition to the induced dipole model, the classical Drude oscillator model is another popular approach to model polarization effect.[41, 110] Roux, MacKerell and colleagues have been developing a polarizable CHARMM force field basing on this approach. [29, 30, 111-119] Unlike the induced dipole model, which treats the polarization response using point dipoles, the Drude model represents the polarizable centers by a pair of point charges. For each non-hydrogen atom, a point partial charge is tethered via a harmonic spring. This point charge (the Drude oscillator) can react to the electrostatic environment and cause the displacement of the local electron density. The

atomic polarizability is dependent on both the charge on the Drude particle and the harmonic force constant. In MD simulations, the extended Lagrangian, rather than the self-consistent iteration is used to evaluate the polarization response, by allowing the Drude particles to move dynamically and experience nonzero force. Small fictitious masses are assigned to each Drude particle and separate low temperature thermostats are applied to the Drude particle degrees of freedom.[120] In case of energy minimization, self-consistent iteration will be required to solve for the polarization.

Determination of electrostatic parameters for the Drude oscillator is not as straightforward as induced dipole models. Masses assigned to the Drude particles are chosen empirically. The determination of atomic charges and polarizabilities requires a series of calculations of perturbed ESP maps. To date, this force field has been for parameterized for water[29, 30], a series of organic molecules including alkanes,[112] alcohols,[113] aromatics,[114] ethers,[115, 116] amides,[111] sulfurs,[117] and ions.[121, 122] Attempt has also been made to combine the Drude based polarizable force field with quantum mechanics in QM/MM methods.[123] It was noted that pair-specific vdW parameters are needed to obtain accurate hydration free energies of small molecules using the polarizable force field. This is likely due to the problematic combining rules used to compute the vdW interactions between unlike atoms.

The Drude model has been implemented in CHARMM[75, 124] and recently in NAMD package,[125] in which the computational cost is about 1.2 to 1.8 times that of fixed-charge CHARMM.[126]

1.3.4 CHARMM-FQ

The fluctuating charge model, also known as charge equilibration or electronegativity equalization model, is an empirical approach for calculating the charge distributions in molecules. In the fluctuating charge formalism, the partial charge on each atom is allowed the change to adapt to different electrostatic environments. The variable partial charges are computed by minimizing the electrostatic energy for a given molecular geometry. Compared with the induced dipole and Drude models, the fluctuating charge models are minimally parameterized and easier to implement because the polarizability is induced without introducing new interactions beyond the point charges. Either extended Lagrangian or self-consistent iteration can be used to compute the fluctuating charges in the MD simulations, with similar advantages and disadvantages discussed above.

The CHARMM-FQ force field,[127, 128] developed by Patel, Brooks, and coworkers, has been parameterized for small molecules,[32] proteins,[32, 129] lipids, lipid bilayers,[115, 130] and carbohyrates.[127] In addition to biophysical studies, the force field has been applied to investigate liquid–vapor interfaces.[131] On the other hand, there are some known limitations for the fluctuating charge models. Such models allow artificial charge transfer between widely separated atoms, which however can be controlled with addition constraints. Also the intramolecular charge-flow is limited by the chemical connectivity. Thus it is difficult to capture the out of plane polarization in molecules such as aromatic benzenes with additional charge sites. The CHARMM-FQ force field has been implemented in CHARMM software package.[75]

1.3.5 X-Pol

Gao and coworkers proposed the X-Pol framework by combining the fragment-based electronic structure theory with molecular mechanical force field.[34, 132] Different from the traditional force field conception, X-Pol does not need the bond stretching, angle, and torsion terms as they are explicitly represented by quantum mechanics. The polarization and charge transfer between fragments are also evaluated quantum mechanically with the electronic structure theory.[132] Furthermore, X-Pol can be used to model chemical reactions.

In X-Pol, large molecular systems are divided into small fragments. Electrostatic interactions within the fragments are treated based on the electronic structure theory. Between fragments, the electrostatic interactions are described by the combined quantum mechanical and molecular mechanical (QM/MM) approach. Also, a vdW term is added to the interfragment interaction due to the omitted electron correlation and exchange repulsion. A double self-consistent-field (DSCF) procedure was applied to converge the total electronic energy of the system as well as within the fragments, which includes the mutual polarization effect.

The X-Pol potential has been applied to MD simulations of liquid water,[133] liquid hydrogen fluoride,[134] and covalently bonded fragments.[135, 136] Recently, this model was demonstrated in a molecular dynamics simulation of solvated protein.[137] As expected the computational efficiency of the X-Pol is in between those of simple classical force fields and full ab initio method. For the solvated trypsin, it took 62.6 h to run 5 ps on a single 1.5 GHz IBM Power4 processor. The parallel version of the X-Pol is under development.

1.3.4 PFF

Kaminski *et al* developed a polarizable protein force field (PFF) based on *ab initio* quantum chemistry.[138, 139] The electrostatic interaction is modeled with induced dipoles and permanent point charges. Except the dispersion parameter, all the other parameters, including the electrostatic charges and polarizabilties, are obtained by fitting to quantum chemical binding energy calculations for homodimers. The dispersion parameters were later refined by fitting to the experimental densities of organic liquids that relate to fragments through condensed-phase simulations.[20] Using the PFF model, gas-phase many-body effects, as well as the conformational energies can be well captured.[139] At the same time, MD simulations for real proteins were performed with reasonable accuracy and computational cost.[20, 140]

To reduce the computational cost, a POSSIM (Polarizable Simulations with Second-order Interaction Model) force field was later proposed, and the calculation of induced dipoles stops after one iteration.[141, 142] By using this formalism, the computational efficiency can be improved by almost an order of magnitude. As the analytical gradients (forces) are unavailable, the Monte-Carlo technique is used in condensed-phase simulations. POSSIM have been validated on selected small model systems, showing good agreement with *ab initio* quantum chemical and experimental data. Recently, the parameters of alanine and protein backbone have been reported.[143]

1.4 Overview of Thesis Work

In this dissertation, the development and application of a polarizable atomic model, AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications), is presented. By calculating the hydration free energies of a series of small

molecules, an optimized protocol was established to parameterize the electrostatic interactions. Next, the systematical development and parameterization procedure of AMOEBA protein force field was reported. To access the quality the force field, the binding free energies of benzamidine analogs to trypsin using molecular dynamics alchemical perturbation were calculated. Last, we applied AMOEBA to study the thermodynamic effect of constraining and hydrophobicity on binding energetics between phosphotyrosine(pY)-containing tripeptides and the SH2 domain of growth receptor binding protein 2 (Grb2).

2 Investigation of Multipole Electrostatics in Hydration Free Energy Calculations

2.1 INTRODUCTION

Hydration of small molecules is an important phenomenon in many chemical and biochemical processes. The ability to accurately calculate the hydration free energy is critical in the force field development and the application of molecular modeling to molecular design and drug discovery. For example, hydration free energy is one of the components in determining the binding affinity of a ligand to its receptor.[144] Since hydration free energy is a sensitive measure of the interaction between a solute and water, it has been commonly used to assess the accuracy of physical models, such as the quality of partial charges and implicit solvent models, by comparing with the experimental hydration free energies of a wide range of organic molecules.[145-149]

Solvent effects can be computationally investigated with implicit and explicit methods.[150] The implicit solvent approaches including Poisson Boltzmann (PB) and Generalized Born (GB) methods. Studies with implicit models usually focus on the evaluation of charge parameters,[149] and improvement of the polar/nonpolar solvation models.[150-154] Although implicit-solvent methods are computationally efficient, there are still notable limitations. The continuum approximation ignores finite size effect of water as well as tightly bound individual water molecules.[155] It is unable to distinguish positively and negatively charged molecules of the same size[156] unless they are specially treated.[157] Extensive parameterizations against experimental data and explicit-solvent simulations are necessary.[151, 152, 158] The alternative in treating solvent is through explicit representation of solvent molecules. With recent computational and methodological advancement, alchemical approaches such as

thermodynamic integration (TI)[159, 160] and free energy perturbation (FEP)[161, 162] are increasingly used to compute precise hydration free energies of amino acid side chain analogs and small molecules in explicit solvent.[147, 148, 150, 163-166] In recent studies that cover a wide range of organic molecules,[147, 148, 162] the reported root mean square error (RMSE) of predicted hydration free energies from experimental values is slightly over 1 kcal/mol using fixed-charge models, which is somewhat similar to that using implicit solvent approach.[149] It was suggested that treatment of polar and poly-functional molecules need to be improved.[147] Hydration free energy calculations have also been utilized in the parameterization of a Drude oscillator based polarizable force field.[167]

In an effort to develop accurate atomic force field for molecular interactions, polarizable multipole has been introduced in AMOEBA to account for atomic charge distribution. While it is the advantage of a polarizable force field that the charge distribution can be derived from high-level gas-phase QM calculations of model compounds, there are various approaches available to obtain the atomic multipole expansion, such as distributed multipole analysis (DMA),[168, 169] atoms-in-molecules (AIM),[170] cumulative atomic multipole moments (CAMM),[171, 172] and electrostatic potential fitting.[173] Previously we have shown that DMA derived multipoles perform well in modeling both the inter- and intramolecular electrostatic interactions.[22] In the original DMA the multipole moments of the associated charge distribution are represented by a multipole expansion at a set of atoms (or selected sites). This procedure is exact and efficient, and gives an excellent representation of the molecular charge distribution with up to quadrupole moments at each atom.[169] For large basis sets with

diffuse functions, however, DMA yields seemingly “unphysical” atomic multipoles. In addition, the multipole values vary significantly with the size of the basis sets used.[51] Recently, a modified DMA procedure[51] was proposed to overcome these problems by using numerical quadrature for the diffuse functions. While the new DMA produces atomic multipoles converge with improved basis-sets, the magnitudes of monopole, dipole and quadrupoles seem very different from those given by original DMA. At very short range, such as hydrogen-bonding distance, the different distribution of multipoles will lead to different interaction energies, even though the molecular moments are always well reproduced.

In this chapter, we applied a polarizable multipole-based electrostatic model to calculate hydration free energies of small ligands. The main purpose is to systematically investigate different approaches for derivation of atomic multipoles and the effect of basis sets on the hydration free energy. Beside the new DMA procedure, we evaluated an alternative approach to deal with the diffuse function issue in the original DMA: the atomic multipoles are derived from the original DMA with small basis set and then optimized by fitting to the electrostatic potential around the molecule. In addition, we examined the treatment of long-range correction to vdW interactions with finite cutoffs in the hydration free energy calculations. Finally, we explored the possibility of combining solutes with polarizable multipoles and solvent calculated through traditional molecular mechanics (PM/MM) in a hybrid model.

2.2 COMPUTATIONAL METHODS

2.2.1 Molecular Systems

The free energy of hydration was computed using molecular dynamics and Bennett Acceptance Ratio (BAR) method. Seven organic compounds were investigated: ethylbenzene, *p*-cresol, isopropanol, imidazole, methylethylsulfide, acetic acid and ethanol. This set of molecules represents the common chemical functional groups in bimolecular systems and drug-like compounds, including alkyl, benzyl, phenol, hydroxyl, imidazole, carboxyl, and sulfide groups. The solvent was modeled using both the AMOEBA[19] polarizable water and a TIP3P-like fixed-charge water model that we developed here to use with polarizable solutes.

2.2.2 The Polarizable Multipole Force Field

The molecular dynamics simulations for various systems have been performed using AMOEBA polarizable force field.[19, 22, 58, 62, 174, 175] In AMOEBA force field, each atom possesses permanent charge, dipole and quadrupole moments. Moreover, the electronic polarization effects are included, using a self-consistent dipole induction procedure.[22] Repulsion-dispersion interactions between pairs of non-bonded atoms are represented by a buffered 14-7 potential.[176] Parameters for all the organic compounds were taken from the AMOEBA force field for small molecules (available in TINKER 5.1 molecular modeling package[177]), except atomic multipoles which have been varied in the current study. New atomic multipole parameters of these molecules were derived from *ab initio* calculations with different basis sets using the GDMA program.[178] Early version of GDMA (v1.x) offers the original DMA while the latest GDMA (v2.2) now implements the modified DMA to address diffuse function and basis set dependence

issues. In GMDA v2.2, by setting “Switch 0” and “Radius H 0.65”, one can also access the original DMA procedure. The default behavior is to set “Switch 4”, which invokes the modified DMA protocol. The structure of each molecule was optimized quantum mechanically at the level of HF/6-31G* using Gaussian 03.[179] Single point calculations were subsequently performed on the optimized geometry using MP2 method with four basis sets: (i) 6-311G**, (ii) 6-311++G(2d,2p), (iii) cc-pVTZ, and (iv) aug-cc-pVTZ, for comparison.

We tested two procedures to obtain the atomic multipoles for each basis set. One is to use the original DMA to derive the permanent atomic multipole using the 6-31G* basis set, and then optimize the multipoles to the electrostatic potential (ESP) derived from the above four basis sets. We refer to this procedure as “original-fit” in the discussion below. The optimization was done by using the POTENTIAL program in TINKER; the atomic monopoles were fixed at those from 6-31G* and only dipole and quadrupole moments were allowed to vary. This ESP fitting procedure also gave us a consistent set of multipoles across the basis sets by perturbing the dipole and quadrupole moments from a lower basis set. In the second approach, which we call “DMA2”, the new DMA procedure was used to compute the permanent multipole moments directly at each basis set. Multipoles for the same atom type were then averaged and optimized to the QM electrostatic potential (ESP) from the same basis set. GDMA 2.2 software package was used. The hydrogen atomic radius ratio was set to 0.31 in both original and modified DMA procedures. Also in both procedures, the atomic multipoles were optimized to the QM electrostatic potential until the root mean square gradient difference was smaller than 0.1, the grid offset 1.0 Å from the vdW surface of each atom. For

alcohols, the atomic quadrupole components of the hydroxyl group (O and H atoms) were reduced by a constant factor similar to that previously applied to water.[19] This reduction led to better agreement with both *ab initio* water dimer geometry and experimental liquid properties.

2.2.3 Hydration Free Energy Caculation

Alchemical transformation was performed to calculation the hydration free energy of the small organic compounds (**Figure 2.1**). The hydration free energy of a compound was calculated from:

$$\Delta A_{\text{hydration}} = -\Delta A_{\text{discharging(aq.)}} - \Delta A_{\text{decoupling(aq.)}} - \Delta A_{\text{recharging(vac.)}} \tag{2.1}$$

where $\Delta A_{recharging(vac.)}$ is the free energy change due to growing the intramolecular electrostatic interactions in vacuum; $\Delta A_{discharging(aq.)}$ results from annihilating electrostatic interactions both within the solute itself and between solute and solvent. $\Delta A_{decoupling(aq.)}$ represents the free energy change by turning off vdW interactions between the solute and its environment using a soft-core buffered 14-7 potential:

$$U_{ij} = \lambda^5 \varepsilon_{ij} \frac{1.07^7}{\left[0.7(1-\lambda)^2 + (\rho + 0.07)^7\right]} \left(\frac{1.12}{0.7(1-\lambda)^2 + \rho^7 + 0.12} - 2 \right) \tag{2.2}$$

where ε is the well depth and λ is the scaling factor. By varying λ from 1 to 0, the vdW interactions between ligand and its environment are gradually turned off. $\rho_{ij}=R_{ij}/R_{ij}^0$ with R_{ij} as the actual separation between atoms *i* and *j* and R_{ij}^0 the minimum energy distance parameter.

The scaling factor λ is also introduced to the long-range correction (LRC):

$$U_{LRC} = \frac{1}{2}\sum_i \sum_j \left\{ (N_i \times N_j) - (1-\lambda^5)\left[L_i \times N_j + (N_i - L_i) \times L_j \right] \right\} \frac{4\pi}{V} \int_{r_c}^{\infty} U_{ij}(r) r^2 dr \quad (2.3)$$

where r_c is the cutoff radius, V is the system volume, U is the soft-core potential, N_i and N_j are the total number of atoms of type i and j in the system, L_i and L_j are the number of soft-core interaction sites of atom type i and j in the solute being annihilated. In this way, U_{LRC} is scaled to the same extent as soft-core interactions within the cutoff radius. When λ=0, the long-range vdW interactions between ligand and its environment is removed. Polynomial tapering function applied from 0.9 r_c to r_c reduces vdW interactions to zero at the distances beyond r_c and maintains smooth atomic forces. The reduced energy and virial value in the tapering range are also included in the LRC correction.

To compute the free energy changes between neighboring states (λ_i and λ_{i+1}), the Bennett Acceptance Ratio (BAR) method[180] was utilized:

$$\Delta A(j)_{\lambda_i \to \lambda_{i+1}} = RT \ln \frac{<1/[1+\exp((U_{\lambda_i} - U_{\lambda_{i+1}} + C)/RT)]>_{\lambda_{i+1}}}{<1/[1+\exp((U_{\lambda_{i+1}} - U_{\lambda_i} + C)/RT)]>_{\lambda_i}} + C \quad (2.4)$$

where C is given by:

$$C = \Delta A(j-1)_{\lambda i \to \lambda i+1} \quad (2.5)$$

and j is the iteration index. Here, $U_{\lambda i}$ is the potential energy of the system evaluated using the parameters from λ_i, and <> is the ensemble average. ΔA is solved iteratively until the value of (ΔA(j) – ΔA(j-1)) < 0.01 kcal/mol. The statistical error in the free energy change between two steps was computed from Eq. 10 in Reference [180]. The total statistical error in the solvation free energy in bulk water was computed as the sum of the errors from individual perturbation steps.

2.3 COMPUTATIONAL DETAILS

All liquid simulations were performed using PMEMD and SANDER in AMBER 9.[69] TINKER was used to prepare the initial systems and for the gas phase re-charging simulations in vacuum. An automation script to set up the system with all perturbation steps as well as the post free energy analysis procedures is available online.

To carry out the perturbations in bulk water, the solute molecule was placed at the origin of a pre-equilibrated periodic box of solvent containing 800 water molecules in a cube with a 28.78 Å dimension on each side. The system was then equilibrated for 50 ps using NPT ensemble at 298 K. The last frame of the simulation was used as the starting point for all the intermediate states λ_i. The electrostatic interactions were decoupled in 11 steps by scaling down solute atomic multipoles and polarizabilities linearly with λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0). For vdW interactions, we compared different scaling protocols (see Results and Discussion). For each value of λ, 500 ps or 1 ns of constant volume molecular dynamics was performed, with the density fixed at the NPT-average and a time step of 1 fs. The temperature was maintained at 298 K using Berendsen thermostat[181]. The vdW cutoff was set to both 9 Å and 12 Å with and without long-range correction to evaluate the cutoff and LRC effects. Long-range electrostatics for all the systems were treated using Particle Mesh Ewald (PME) summation.[182-184] The PME calculation used a 36 × 36 × 36 grid and fifth-order B-spline interpolation. The induced dipoles were iterated until the root-mean-square change was below 0.01 D/atom. Atomic coordinates of the simulation system were saved every 500 fs. The first 100 ps simulation trajectories were ignored in the free energy analysis. A tighter induced dipole convergence of 10^{-5} D/atom was used in the energy calculation for the free energy analysis.

Gas phase simulations were run on the single solute molecule only for 50 ps with a time step of 0.1 ps. The temperature was maintained at 298 K using stochastic thermostat. The induced dipoles were converged to 10^{-5}D. Atomic coordinates were collected every 100 fs. Post free energy analyses were performed on all 500 configurations. BAR was used to evaluate the free energy change between adjacent states.

2.4 Results and Discussion

2.4.1 Hydration Free Energy Calculation Protocol

In alchemical free energy calculations, a sufficient number of "small" perturbation steps would ensure adequate overlaps between adjacent states. However the computational cost is also an important consideration especially given the AMOEBA polarizable potential is more costly than fixed charge models. In the approach adopted in this study, the electrostatic and vdW interactions between solute and solvent were decoupled sequentially. The sampling for vdW decoupling was the most challenging step since solute and solvent molecules begin to overlap. We thus first determined an optimal distribution of intermediate states and simulation length in the vdW decoupling simulations.

Hydration free energy of ethylbenzene has been calculated using different simulation protocols with a different number of intermediate states. **Figure 2** shows the hydration free energy results from three vdW decoupling schedules: (a). λ = (1.0, 0.95, 0.9, 0.85, 0.8, 0.775, 0.75, 0.725, 0.7, 0.675, 0.65, 0.625, 0.6, 0.55, 0.5, 0.4, 0.2, 0.0); (b). λ = (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0); (c). λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.0). Simulations for all intermediate states were performed for 1ns with a

12 Å cutoff, using AMOEBA force field. The results here include the electrostatic component of the hydration free energies and the long-range correction, however both contributions are constants and do not affect the comparisons we made. It can be seen that for the decoupling paths with 18 steps (a) and 11 steps (b), hydration free energies converge after 500 ps. Standard deviations of the hydration free energy value calculated from 500 ps to 1 ns (with a 100 ps interval) simulations are 0.04 kcal/mol for protocol (a), and 0.05 kcal/mol for protocol (b). However in the 9-step (c) vdW decoupling, the hydration free energy does not show a sign of convergence even at 1ns.

We have summarized the vdW decoupling free energies calculated using the three schedules in **Table 2.1**. By further examining the free energy change at the intermediate states, it is clear that the difference between the hydration free energy from the 9-step protocol (c) and the results of the other two mostly occurs when λ is varied from 0.8 to 0.6. The free energy change of this perturbation is -4.30 and -4.32 kcal/mol calculated from protocols (a) and (b), but -3.77 kcal/mol using protocol (c). A total of 9 steps (including the boundary states) were used in protocol (a) to transform λ from 0.8 to 0.6, 5 steps in protocol (b), and 3 steps in protocol (c). The more than half a kcal/mol difference indicates the linear schedule of protocol (c) is not sufficient. In addition, the simulation results remain not converged by extending the simulation time from 500 ps to 1 ns. The difficulty in calculating the vdW decoupling free energy at λ near 0.5 is a result of large fluctuation in the soft-core vdW energy. On the other hand, all three protocols gave almost identical free energies when λ is varied from 1.0 to 0.8. The soft-core vdW potential is also greatly "smoothed" at the end point when λ approaches zero, and the contribution between λ=0.4 and 0 is nearly zero. The statistical errors in **Figure 2**,

calculated using Bennett's formula, also show a trend of smaller statistical error with more intermediate states. Note that an excessive number of states could also lead to a large statistical error. At 500 ps, the statistical error of the calculated hydration free energies is ±0.05 kcal/mol, ±0.15 kcal/mol and ±0.25 kcal/mol for protocol (a), (b) and (c), respectively. The error due to the electrostatic decoupling is negligible and again does not affect the comparison here. With both computation expense and precision in consideration, 11 steps of 500 ps simulations were used to decouple the vdW interactions in the remaining hydration free energy calculations.

2.4.2 van der Waals Cutoff and Long-range Correction

Previous works showed that long-range vdW interactions between ligands and proteins can contribute to the binding affinity by more than 1 or 2 kcal/mol, with cutoff at 7.0 to 9.0 Å[185]. The long range vdW interactions beyond a "large" cutoff should be negligible. In the current study, we are computing the hydration free energy values for relatively small solute molecules. To evaluate the effect of vdW cutoffs and long-range correction, we calculated the hydration free energy values of six organic compounds using different cutoffs with and without LRC. All the other setups of the six systems remained the same in the calculations, including box sizes, atomic multipoles (derived at MP2/6-311++G(2d,2p) level using the DMA2 method). **Table 2.2** shows the hydration free energy results using either a 9 Å or a 12 Å cutoff for vdW interactions, with and without LRC in the vdW decoupling. Note that the LRC is a constant in the NVT simulations and thus contribute no forces on each atom at any given time. Therefore the LRC has no effect on the simulation trajectories or the electrostatic decupling free energy at the same vdW cutoff, but it does affect the vdW decoupling free energy when the

solute atoms are "disappeared" in the solvent. The root mean square difference (RMSD) in the hydration free energies between the 9 Å and 12 Å cutoff simulations is 0.33 kcal/mol without the LRC. When the LRC is included in the vdW decoupling free energy calculation, the difference is reduced to 0.17 kcal/mol. Since the electrostatic component of hydration free energy is not affected by the LRC, this error reduction was exclusively in the vdW component. For smaller molecules such as acetic acid, the results from 9 or 12 Å cutoff are essentially identical with the LRC. LRC is clearly making the results at different cutoffs more consistent. However a 9 Å with LRC can still lead to non-negligible errors for large molecules such as p-cresol. With regard to the computation cost, simulations with 12 Å cutoff are about 1.25 times slower than those using 9 Å cutoff. In our study, we chose the 12 Å cutoff, to get the most rigorous hydration free energy results.

When a 12 Å cutoff is used for the vdW interaction, the inclusion of LRC lowered the hydration free energy by 0.14 kcal/mol (RMSD). For simulations using 9 Å cutoffs the contribution of LRC was -0.33 kcal/mol. The contribution of LRC to hydration free energy is more negative than those without LRC as the vdW interactions beyond the cutoff radii are always favorable. In the remaining study, a 12 Å cutoff has been applied to the vdW interactions, with LRC included. As the LRC to hydration free energy is already comparable to the magnitude of the statistical error in our hydration free energy calculations at the 12 Å cutoff, we believe the error (comparing to an infinite cutoff) should be negligible. However, it should be noted that the LRC contribution is roughly proportional to number of the atoms of the solute molecule. In the system studied here, the average atomic LRC contribution during vdW decoupling is ~0.01 kcal/mol per atom

at a 12 Å cutoff; for a molecule with 100 atoms, there would be an error of ~1 kcal/mol if the LRC decoupling is ignored. Similarly, for simulations with a 9 Å cutoff, every 37 atoms in a drug molecule will lead to an error around 1 kcal/mol if the LRC decoupling is ignored. Therefore, when a 9 Å cutoff, which is frequently used in molecular simulations, is employed, LRC decoupling term is highly recommended. In addition, it should be kept in mind that the correction is based on the assumption of isotropic environment which is appropriate for solvation in a homogenous solvent.

2.4.3 Electrostatic Multipoles from Different Methods and Basis Sets

We investigated different approaches to derive atomic multipoles and the effect of basis sets on hydration free energy calculations. The vdW and remaining parameters are transferred from AMOEBA. All the simulations in this section employed a 12 Å cutoff for vdW plus LRC, so that the vdW contributions to the hydration free energy are exactly the same in all comparisons.

Atomic multipoles for the small molecules were derived from QM single point calculations using the original DMA with ESP fitting (original-fit) or the new DMA2 method (see Computational Methods). **Table 2.3** compares the hydration free energy of p-cresol computed from atomic multipoles using four different basis sets (6-311G**, 6-311++G(2d,2p), cc-pVTZ, and aug-cc-pVTZ) at the MP2 level. It can be seen that with 6-311++G(2d,2p) and cc-pVTZ basis sets, hydration free energy difference between using original-fit and DMA2 is within 0.1 kcal/mol. With aug-cc-pVTZ and 6-311G**, the difference is around 0.4 kcal/mol. The overall RMSEs are 0.69 and 0.66 kcal/mol for original-fit and DMA2 respectively. Thus for small molecules, the atomic multipoles

from the new DMA2 procedure and old ESP fitting give very similar intermolecular interaction energies as reflected in the hydration free energy.

Consistent with our previous study, we found that it is necessary to scale down the atomic quadrupole moments of the hydroxyl group for alcohols. According to the work by Ren and Ponder,[19] the quadrupole moments of AMOEBA water molecules were reduced to 73% of the QM DMA values, which led to a reduction of the water-water flap angle to 57 ° and better reproduced a series of *ab initio* and experimental properties. From **Table 2.4**, we can see that hydration free energy values of isopropanol without scaling poorly agree with experiment (RMSE of 1.89 kcal/mol). All basis sets overestimate the hydration free energy values by more than 1 kcal/mol. However when quadrupoles of hydroxyl groups are scaled to 70% of the DMA2 values, the RMSE significantly reduces to 0.62 kcal/mol. Same quadrupole parameters obtained with original-fit needs to be scaled to 60% to achieve similar RMSE (0.58 kcal/mol). This is due to the different distribution of multipole moments given by the new and original DMA methods, even though both theoretically produce the same molecule moments and ESP. To verify the transferability, the same scaling factors were applied to the hydroxyl atomic quadrupoles in ethanol. Indeed, with the scaling, the hydration free energies of ethanol show a satisfactory agreement with experimental data using DMA2 (RMSE=0.61 kcal/mol) and original-fit (RMSE=0.70 kcal/mol). Acetic acid and p-cresol also have the hydroxyl group, but no scaling is required since their hydroxyl groups are considered part of the larger functional groups and the scaling of the quadrupoles have little effect on the gas-phase dimer or hydration properties.

Table 2.5 illustrates the effects of basis sets on the hydration free energies for seven compounds. The atomic multipoles were optimized using the DMA2 approach. Overall Comparisons show that hydration free energies calculated with cc-pVTZ and 6-311G** are similar to each other. The RMSD in the hydration free energy between these two basis sets is 0.23 kcal/mol. Also these two basis sets underestimate the experimental data in most cases, both with RMSEs around 1.0 kcal/mol. On the contrary, 6-311++G(2d,2p) overestimates the hydration free energies in all cases except acetic acid, RMSE 0.77 kcal/mol. Acetic acid is a special case, hydration free energies calculated with all the four methods are underestimated by 1~2 kcal/mol. In general, 6-311++G(2d,2p) gives more favorable hydration free energies for all the small molecules than the calculated results with all the other basis sets. Compared with cc-pVTZ and 6-311G**, hydration free energies calculated from 6-311++G(2d,2p) are closer to that from aug-cc-pVTZ, with a RMSD of 0.58 kcal/mol. The aug-cc-pVTZ set does not show a systematical under- or overestimation.

Overall, the agreement with experimental hydration free energy results improve with the size of the basis set used in the *ab initio* calculations to derive the atomic multipoles. The inclusion of diffuse function the QM basis set has a large effect, as seen in the comparison between the results from aug-cc-pVTZ and 6-311++G(2d,2p) with those of 6-311G** and cc-pVTZ. Obviously, aug-cc-pVTZ gives the best hydration free energy result, and 6-311++G(2d,2p) is also comparable. Plus 6-311++G(2d,2p) is much more affordable since it employs approximately half Gaussian basis functions as the aug-cc-pVTZ basis set, it is recommended as the optimal basis set to MP2 to perform hydration free energy calculations for large molecules. Furthermore, since aug-cc-pVTZ

and 6-311++G(2d,2p) have better performances than the other two basis sets, we conclude that it is important to use basis sets with diffusion functions to capture intermolecular interactions.

Table 2.5 also lists previously reported hydration free energies calculated with some of the fixed charge models GAFF, CHARMm-MSI and OPLS_2005. Although some of the calculated values are close to experimental data, some have erroes as large as 5 kcal/mol. The MUEs and RMSEs of the available hydration free energy valuesare between 1 kcal/mol and 3 kcal/mol. Recent studies over a much wider range of organic molecules using fixed-charge model[147, 148, 162] also reported root mean square error (RMSE) greater than 1 kcal/mol, consistent with the current finding. Our results suggest that the inclusion of polarization effect and/or the use of atomic multipoles moments offer better performance. It is interesting to note that the gas-phase electrostatic potentials of the atomic multipoles derived from the four basis sets are very similar for all the molecules (**Table 2.6**). However, the final hydration free energy values can be as different as more than 1 kcal/mol in some cases. For example, the RMSD of the average magnitude of gas phase potentials is 0.14 kcal/mol between results from aug-cc-pVTZ and 6-311G**; while the RMSD of the hydration free energies between these two basis sets is increased to 0.62 kcal/mol. The hydration free energies with atomic multipoles derived from gas phase show a favorable agreement with the experimental measurements, suggesting that the polarization effect between solute and solvent environment are well captured by the explicit atomic dipole induction model in AMOEBA. Note that the RMS gas phase potentials for isopropanol at 6-311++G(2d,2p) is very different from the other

three basis sets because of the quadrupole scaling on the hydroxyl group, leading to the hydration free energy difference of more than 1 kcal/mol from the other basis sets.

The overall hydration free energy results using the polarizable atomic multipoles are encouraging compared those with fixed atomic charge force fields. It was suggested[147] that the fundamental limit of fixed-charge force fields is roughly 1 kcal/mol (mean unsigned error (MUE)) for hydration free energy. Our work here shows hydration free energy can be calculated within an accuracy of 0.41 kcal/mol (MUE) using gas-phase atomic multipoles from MP2/aug-cc-pVTZ, with the polarization modeled via induced atomic dipoles. Previously we showed that the polarization enhances the solvation of benzamidinium ion in water, and the contribution of polarization to hydration free energy is about 10% of the total electrostatic contributions. Therefore it is expected that even a fixed charge model can lead to a reasonable hydration free energy when the atomic charges are systematically increased to implicitly account for the polarization effect in water environment. The challenge however is that the polarization in the protein-like environment had an opposite effect which actually weakens the interaction between the small molecule and the protein.

2.5 CONCLUSIONS

In this chapter, we computed the hydration free energies for several organic molecules using AMOEBA polarizable force field. We first evaluated the effect of vdW cutoff length and importance of decoupling of the long-range correction by performing a series hydration free energy simulations. We used an alchemical approach in which the electrostatic and then vdW interaction between the solute and solvent molecules were turned off in several steps. We showed that a long vdW cutoff (12 Å) with LRC and 11

perturbation steps in vdW annihilation are necessary for high-precision calculations of hydration free energy. Extra steps in the middle of vdW annihilation are needed as the system energy fluctuation is rather significant when solvent and solute molecules begin to penetrate each other.

With the most appropriate simulation protocol determined, we have investigated various models for intermolecular electrostatic interactions between the organic molecules and water environment via the calculation of hydration free energies. By putting the solute with gas-phase atomic multipoles, the polarization effect between solute and water environment was well captured by the explicit atomic dipole induction model in AMOEBA, as evident by the good agreement between the calculated and experimental hydration free energies for the seven molecules. We tested two methods to derive the gas-phase atomic multipoles from *ab initio* calculations for the solute molecules. One is ESP fitting based on the original DMA method, and the other utilizes the new DMA procedure by Stone. The two methods gave very similar hydration free energies for p-cresol in our test, with the difference comparable to the statistical error of the simulation. We recommend the original-fit procedure to derive atomic multipoles based on our experience. Also this approach allows fitting to the ESP of multi-conformations of flexible molecules simultaneously to derive "conformation-independent" atomic multipoles and allows atoms in symmetry to share the same atom type (e.g. the three H atoms in a methyl group). We subsequently compared the electrostatic parameters (atomic multipoles) derived from MP2/6-311G*, MP2/6-311++G(2d,2p), MP2/cc-pVTZ, and MP2/aug-cc-pVTZ. The hydration free energy results based on the four basis sets were in good agreement with experimental data, with

RMSEs from 0.63 to 1.09 kcal/mol. Among these four *ab initio* methods, the aug-cc-pVTZ basis set gave the best hydration free energy results while the 6-311++G(2d,2p) also performed well and was computationally more affordable. It is encouraging that the overall accuracy increases with the larger basis sets and the inclusion of diffuse functions in the *ab initio* basis set is highly recommended in deriving the atomic multipoles for modeling intermolecular interactions.

Furthermore, we investigated a hybrid PM/MM approach where the solute molecule was modeled with polarizable atomic multipoles and solvent was represented by fixed-charge TIP3P-like water molecules. The saving in computational cost is about 50% when compared to a fully polarizable model. The calculated hydration free energy values were within 1.16 kcal/mol (MUE) of the experimental measurements. While the error is about twice of that from simulations using the polarizable AMOEBA water, there was a systematic overestimation for five out of the six molecules compared. It is likely that a better water model or careful parameterization can further improve the accuracy.

Table 2.1: vdW contributions of ethylbenzene hydration free energies from intermediate decoupling steps. All units are kcal/mol. Results are calculated using the dynamics simulations trajectories up to 500 ps. For schedule (a) and (b), which uses more intermediate steps, the results have been combined for comparison.

Steps[1]	Intermediate Free Energy Changes								Total
	1.0-0.9	0.9-0.8	0.8-0.7	0.7-0.6	0.6-0.5	0.5-0.4	0.4-0.2	0.2-0.0	
(a) 18	5.127	2.281	-0.241	-4.037	-4.848	-0.896	0.039	0.048	-2.53
(b) 11	5.131	2.304	-0.237	-4.059	-4.851	-0.892	0.040	0.048	-2.52
(c) 9	5.135	2.304	-0.169	-3.605	-4.851	-0.898	0.040	0.048	-2.00

1. Three vdW decoupling schedules:
   (a). λ = (1.0, 0.95, 0.9, 0.85, 0.8, 0.775, 0.75, 0.725, 0.7, 0.675, 0.65, 0.625, 0.6, 0.55, 0.5, 0.4, 0.2, 0.0);
   (b). λ = (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0);
   (c). λ = (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.0).

Table 2.2: The effect of vdW cutoff length and LRC on the AMOEBA hydration free energies. Hydration free energies were estimated using both 9 Å and 12 Å with and without LRC decoupling. All the hydration free energies were decomposed into the electrostatic and vdW components. Statistical errors of the total HFE are given in the parenthesis. All values are in kcal/mol.

Cutoff(Å)		Ethylbenze		p-cresol		Isopropanol		imidazole		methylethyl sulfide		acetic acid	
		ele[1]	vdW[2]	ele	vdW	ele	vdW	ele	vdW	ele	vdW	ele	vdW
w/ LRC	12.0	-3.35	2.65	-9.37	2.04	-8.28	2.64	-11.38	1.31	-4.42	2.55	-7.89	2.36
		(0.15)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	
	9.0	-3.37	2.74	-8.75	1.78	-7.99	2.58	-11.43	1.31	-4.38	2.45	-7.93	2.36
		(0.11)		(0.10)		(0.11)		(0.01)		(0.09)		(0.04)	
w/o LRC	12.0	-3.35	2.88	-9.37	2.23	-8.28	2.58	-11.38	1.34	-4.42	2.69	-7.89	2.44
		(0.15)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	
	9.0	-3.37	2.95	-8.75	2.24	-7.99	2.68	-11.43	1.66	-4.38	2.88	-7.93	2.61
		(0.11)		(0.10)		(0.11)		(0.03)		(0.09)		(0.04)	

1. "ele" refers to the free energy contributions of $-\Delta A_{discharging(aq.)}+\Delta A_{discharging(aq.)}$ in equation (1).
2. "vdW" represents the contribtion of $-\Delta A_{decoupling(aq.)}$ in the same equation.

Table 2.3: Comparison of hydration free energies between original-fit and DMA2 methods for p-cresol. All units are kcal/mol. Statistical errors are given in the parenthesis.

MP2 Basis Set	cc-pVTZ	6-311G**	6-311++G(2d,2p)	aug-cc-pVTZ	Experiment[1]
Original-fit	-6.56	-6.47	-7.26	-7.38	-6.10
	(0.10)	(0.10)	(0.11)	(0.10)	-6.60
DMA2	-6.47	-6.89	-7.33	-6.98	
	(0.11)	(0.10)	(0.10)	(0.16)	

1. Experimental values are reported in ref [186].

Table 2.4: Hydration free energies with scaled quadrupole moments of hydroxyl groups in alcohol molecules, using both original-fit and DMA2. Isopropanol and ethanol were selected for comparison. All units are kcal/mol. Statistical errors are given in the parenthesis.

MP2 Basis Set	Isopropanol			Ethanol	
	Original-fit unscaled	Original-fit scaled to 0.6	DMA2 scaled to 0.7	Original-fit scaled to 0.6	DMA2 scaled to 0.7
cc-pVTZ	-5.43	-4.36	-4.04	-3.58	-4.19
	(0.11)	(0.18)	(0.11)	(0.19)	(0.07)
6-311G**	-5.60	-4.16	-4.34	-4.84	-4.51
	(0.10)	(0.15)	(0.10)	(0.20)	(0.09)
6-311++G(2d,2p)	-7.60	-5.59	-5.58	-5.12	-5.67
	(0.11)	(0.18)	(0.12)	(0.19)	(0.09)
aug-cc-pVTZ	-6.98	-4.36	-4.32	-4.53	-5.04
	(0.16)	(0.28)	(0.11)	(0.22)	(0.09)
Experiment[1]	-4.70, -4.80			-4.90	

1. Experimental values were reported in ref [186].

Table 2.5: Comparison of hydration free energies of small molecules with four different basis sets: (i) 6-311G**, (ii) 6-311++G(2d,2p), (iii) cc-pVTZ, and (iv) aug-cc-pVTZ. Hydration free energies calculated with fixed-charge models were also listed for comparison. All units are kcal/mol. Statistical errors are given in the parenthesis.

	Ethylbenzep-cresol	Isopropanol	Imidazole	Methylethyl sulfide	Acetic acid	Ethanol	MUE	RMSE
cc-pvTZ	-0.68	-6.56	-4.04	-9.07	-1.61	-4.57	-4.19	0.63	1.09
	(0.14)	(0.11)	(0.11)	(0.03)	(0.09)	(0.04)	(0.19)		
6-311g**	-0.53	-6.47	-4.34	-8.70	-1.53	-4.45	-4.51	0.61	1.15
	(0.14)	(0.10)	(0.10)	(0.04)	(0.10)	(0.05)	(0.19)		
6-311++g(2d,2p)	-0.70	-7.26	-5.58	-10.11	-1.87	-5.53	-5.67	0.64	0.77
	(0.14)	(0.14)	(0.12)	(0.03)	(0.09)	(0.04)	(0.20)		
aug-cc-pvTZ	-0.44	-7.05	-4.32	-9.72	-1.56	-5.48	-5.04	0.41	0.63
	(0.12)	(0.11)	(0.11)	(0.03)	(0.10)	(0.04)	(0.22)		
GAFF/AM1[1]	0.8	-5.1	-	-	-0.9	-	-3.9	1.09	1.14
GAFF/AM1(SSBP)[2]	-0.56	-	-2.88	-	-	-8.31	-2.88	1.41	1.60
MSI/AM1(SSBP)[2]	0.03	-	-1.42	-	-	-12.52	-1.94	3.21	3.68
OPLS_2005[3]	-0.46	-	-4.15	-	-	-5.44	-1.94	1.26	1.64
Experiment[4]	-0.70	-6.10	-4.70	-9.63	-1.50	-6.70	-4.90		
		-6.60	-4.80						

1. Calculated hydration free energies from ref[147]. Solute molecules used GAFF, along with the semiempirical AM1-BCC charge model. Waters were represented by TIP3P model.
2. Calculated hydration free energies from ref[150]. Solute molecules used GAFF or CHARMm-MSI force fields, along with the semiempirical AM1-BCC charge model. Water molecules close to the solute were represented by TIP3P model. SSBP were incorporated as the remaining bulk.
3. Calculated hydration free energies from ref[163]. Solute molecules used OPLS_2005 force field. Waters were represented by TIP3P model.
4. Experimental values are reported in ref [186], except for imidazole taken from ref [187].

Table 2.6: The root mean square difference of gas phase potentials of the small molecules with multipole parameters generated from four different basis sets. All units are kcal/mol.

	Ethylbenze	p-cresol	Isopropanol	Imidazole	Methylethyl Sulfide	Acetic acid	Ethanol
cc-pVTZ	0.13	0.25	0.00	0.20	0.16	0.32	0.29
6-311g**	0.21	0.34	0.16	0.27	0.07	0.67	0.33
6-311++g(2d,2p)	0.07	0.11	0.41	0.16	0.07	0.12	0.14
aug-cc-pVTZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00

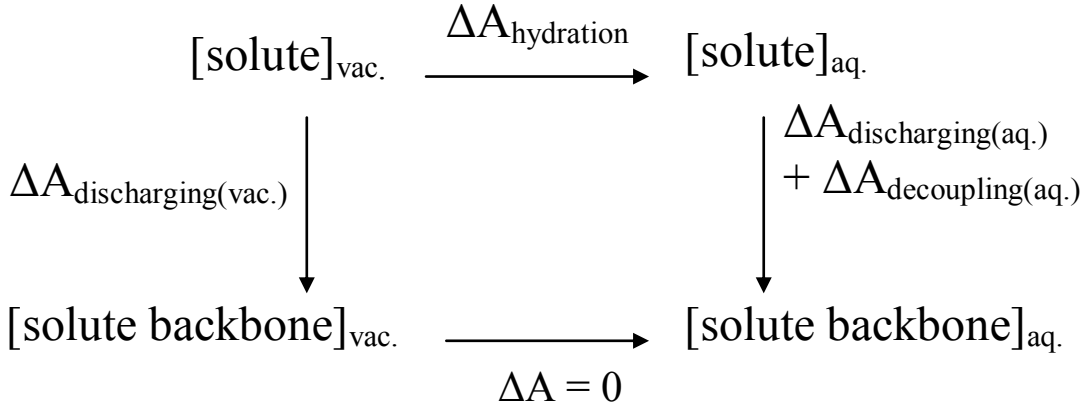

Figure 2.1: Thermodynamic cycle of hydration free energy calculation in explicit water MD simulations. Potential energy of the solute backbone includes valence interactions and vdW interactions within the solute itself.

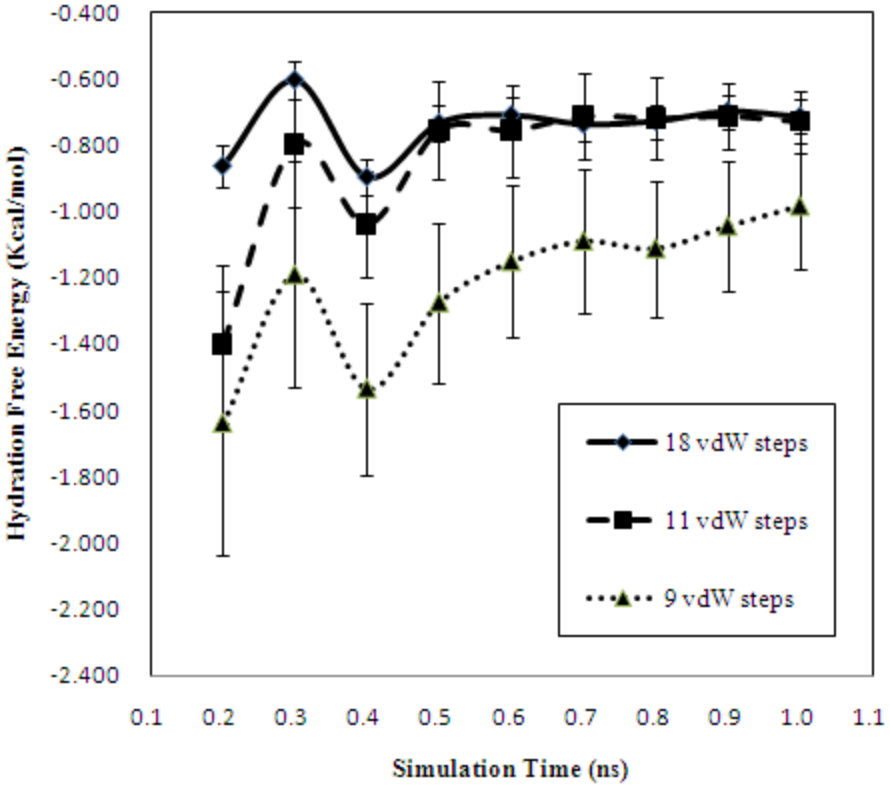

Figure 2.2: Convergence of hydration free energy of ethylbenzene at different simulation time with three vdW decoupling protocols. Hydration free energies with 18 vdW steps (a) are in solid line with diamond, 11 vdW steps (b) in dashed line with square, and 9 vdW steps (c) in dotted line triangle. The bars are the statistical errors computed using Bennett's formula.

3 Development of the Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins

3.1 INTRODUCTION

Proteins are a ubiquitous class of biopolymers whose functionalities depend on the details of their 3D structures, which are encoded by their specific amino acid sequences. Computer modeling and simulations are widely utilized in the study of protein structure, function, dynamics, and interactions with other synthetic or biological molecules. In the so-called molecular mechanics (MM) approaches, interactions among atoms are described by classical empirical potentials that are often referred to as force fields. Unlike the *ab initio* quantum mechanical (QM) methods, the classical MM model treats atoms as rigid particles with electronic degrees of freedom averaged out, thereby lowering the computational cost and allowing simulation of biological events on greater length and time scales. On the other hand, high level *ab initio* theory is becoming more affordable and is now heavily utilized during the development of classic potentials for proteins such as Amber,[9] CFF,[188] CHARMM,[10] MM3[189] and OPLS-AA.[12] This class of force fields typically utilizes fixed atomic charges, point dispersion-repulsion, and empirical functions for valence interactions. The current generation of force field has enjoyed much success in many areas of biological and materials sciences; however, there is still much room for improvement.

Effort to advance molecular mechanics force fields to "next-generation" has been mostly focused on introducing explicit electronic polarization into the electrostatic model. A number of comprehensive reviews on the history and development of polarizable force fields have detailed the significance of polarization effects.[13, 38, 40, 41, 190]. A wide range of studies on water [19, 30, 191], organic molecules [55, 56, 111,

114, 139, 192], peptides[22, 57], protein-ligand binding [59, 60, 62, 130, 193-195], ions[52-54, 122] and ion channels[42] and membrane lipids using polarizable force fields have all demonstrated various benefits of treating the polarization in molecular modeling. In addition to offering more accurate thermodynamic properties, a polarizable force field in principle is more transferable and can be more robustly parameterized by directly utilizing high-level *ab initio* quantum mechanical calculations in gas-phase.

Several different methods for incorporation of many-body effect have been explored. The fluctuating charge approach accounts for polarization by varying the magnitude of atomic charges based on electronegativity equalization.[31, 32, 196-201] Alternatively, the Drude oscillator model, where a point charge moves about the nuclear, has been attempted to model the induced dipole response in water and small molecule systems.[30, 202-206] It has been argued that fluctuating charge model fails for certain situations such as out of plane polarization and bifurcated hydrogen bonding due to the limited charge flow along the bonds.[207] Compared to the classical induced dipole method,[18, 20, 139, 191, 207-209] these schemes involves less complicated numerical algorithm as the point charge framework is sustained. However, the interactive atomic induced dipole model [43, 210] is superior when it comes to reproducing anisotropy and nonadditivity of molecular polarization response across many different types of compounds. Moreover, intramolecular polarization plays important roles when moving from small molecules to peptides that possesses conformational freedom, The conformational dependence of electrostatics, even though significant,[211, 212] has not been given due attention in the development of polarizable or nonpolarizable force fields.

We have shown that interactive induction model of AMOEBA offers an resolution to this issue.[22]

In addition to polarization effect, we also stress that the atomic charge-based representation of permanent electrostatic itself is inadequate. It has been shown that the error in the electrostatic potentials can be reduced by orders of magnitudes by complementing monopoles with dipole and quadrupole moments.[173] The incorporation of higher order atomic multipoles has been shown to greatly improve the quality of crystal structure predictions of organic molecules.[213, 214] One may argue that an alternative to point multipoles, is to use additional off-center charges to enable the same level of description of electron density. In fact, the use of charges at lone-pair site of the water oxygen atom does improve the ability of a model to reproduce water properties such as the density anomaly and dielectric constant.[215] Nevertheless, the determination of both position and magnitude of the charges at the lone pair site is a nontrivial task, requiring fitting to the experimental density-temperature profile. In contrast, the distributed multipole analysis of Stone[49, 50] imparts relatively unambiguous determination of atomic multipoles directly from molecular orbital calculations.

In this chapter, we present the development of a protein potential based on polarizable atomic multipole representation of electrostatics. The intramolecular polarization scheme formulated previously allows us to obtain the permanent atomic multipoles, directly from *ab initio* calculations on blocked dipeptides. A new protocol is applied to combine Distributed Multipole Analysis and electrostatic potential fitting to derive the conformation-independent permanent multipoles. The valence and van der Waals parameters have been derived from liquid simulations of small organic molecules

consisting of similar functional groups. The merging of inter and intramolecular interaction at short separation, including electrostatics, vdW, and torsional contribution, is determined by comparing to gas phase QM data of di- and tetra- peptides. The resulting potential is examined and validated by the simulation of a number of peptides and proteins in solution and compared with experimental data. In the following sections, the potential model, parameterization and initial validations will be presented.

3.2 POTENTIAL ENERGY MODEL

The potential energy model has been explained in details in our previous publications [19, 55]. We will summarize the key features briefly. The total energy of the system is given by

$$U = U_{bond} + U_{angle} + U_{cross} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \quad (3.1)$$

The first five terms are the valence contributions corresponding to the bond, angle, bond-angle coupling, out of plane and torsion energy, respectively. Common functional forms are used for these and the detailed equations have been given previously. [39] The earlier version of AMOEBA protein force field applied a cMAP-style[216] two-dimensional bicubic spline [39] to the backbone torsional energy. In the current version, however, the "traditional" 3-term Fourier expansion function is used for all torsion angles except for the backbone of glycine. The pairwise additive vdW interaction in AMOEBA is described by the buffered 14-7 function.[217] The buffered 14-7 function yields a slightly "softer" repulsive region than the Lennard-Jones 6-12 function but steeper than typical Buckingham exp-6 formulations. For a hydrogen atom connected to a heavy atom X, the vdW site is actually placed along the HX bond such

that the distance between the atom X and vdW site of H is a percentage of the full bond length, referred to as the “reduction factor”.

The permanent electrostatic energy in AMOBEA arises from atomic multipole-multipole interactions. Each atomic center consists of a point monopole (charge), a dipole vector and a quadrupole tensor. Note that there are only five independent quadrupole components due to the symmetry and the traceless requirement. The dipole and quadrupole are defined with respect to the “local” reference frames formed by neighboring atoms. Examples of such local frames are illustrated in Error! Reference ource not found.. As the molecules rotate and diffuse over the course of simulation, the atomic multipoles remain constant with respect to the local frame. The equation for calculating the interaction energy and gradient (forces and torque) between permanent multipoles was presented in ref [19] in traditional Ewald formula.

Electronic polarization accounts for the majority of the many-body effect experienced in biomolecular systems although there are situations where many-body effects for dispersion and repulsion may be important as well.[218] AMOEBA utilizes an interactive atomic dipole induction scheme where the field produced by permanent multipoles and induced dipoles induces a dipole at each polarizable site (atom), which will further polarize other atoms. As a result, the anisotropic molecular polarizability can be described by the isotropic distributed atomic polarizabilities. Based on Thole’s model,[210] the polarization at very short range is damped to avoid the so called polarization catastrophe. The same polarization model is used for both intermolecular and intramolecular polarization. A group-based scheme, where the permanent multipoles will not polarize other atoms within the same group but the induced-induced mutual

polarization occurs among all-atom, allow us to merge the molecular fragments into a chain molecule such as a protein. The extraction of intramolecular polarization to obtain "true" permanent atomic multipoles from *ab initio* quantum mechanical calculations, energy and gradient due to polarization for complex molecules made of more than one polarization groups were explained previously when the organic molecule AMOEBA force fields were reported .

The masking of short range intermolecular interaction is made by means of scaling factors, which were determined by optimizing the transferability of conformational energies from alanine dipeptide to tetrapeptide, as will be discussed in the parameterization section. The final set of scaling factors for interactions between permanent multipoles is 0.4 for the 1-4 interactions (separated by three bonds), 0.8 for 1-5 interactions, and completely neglected for any closer pair. The polarization energy between induced dipoles and permanent multipole moments are computed fully between atoms separated by three (1-4) or more bonds, and are completely neglected for any shorter separation. Consequently, the analytical gradient of the polarization energy is nontrivial because the intramolecular "scaling" for polarization interaction energy differs from the group-based scheme used in induced dipole generation. The derivation of analytical polarization force is given in the Appendix of Reference [55].

A particle-mesh treatment (PME) of polarizable multipole interactions have been developed [219] and implemented in TINKER,[220] Amber[221] and OpenMM[68] for AMOEBA calculations. The latest TINKER has OpenMP shared-memory parallelization of AMOEBA simulations, AMBER (PMEMD) has the MPI parallel capability and OpenMM is specialized for GPU. More than 50% of the computational expense of

AMOEBA is due to iterative calculation of induced dipoles. Recently we have introduced a multi-time step algorithm for molecular dynamics in TINKER and AMBER where the high-frequency valence interactions are computed every 0.25 fs for flexible molecules while the nonbonded interactions including the polarization are updated every 2-3 fs. This simple algorithm leads to a factor of two improvements without constraining any bond stretching. In addition, it has been shown that the iterative solution of induced dipoles can be significantly accelerated by using least-squares fitting based predictor, conjugate gradient iterative method with an efficient pre conditioner [69].

3.3 COMPUTATIONAL DETAILS

Ab initio calculations were performed using Gaussian 09[222] and Q-Chem 4.0.[223] Geometry optimization was carried out at MP2/6-31G* level unless otherwise specified. The atomic multipoles for dipeptide model compounds were derived at MP2/6-311G** level using Stone's original DMA procedure.[50] Diffuse functions in the basis set are avoided for dipeptide at this point as they lead to spurious multipoles especially on the buried atoms. Note that the original DMA procedure can be achieved in the recent gdma program (v2.0 and above)[224] by setting "switch" to 0, H atom radius to 0.65 and S atom radius to 0.80. The resulting atomic multipole values were then optimized against MP2/aug-cc-pVTZ electrostatic potential on a grid around the dipeptide model compounds with the point charges fixed. All single point conformational energies were obtained with complete basis set (CBS) extrapolation from RI-TRIM MP2/cc-pVTZ and cc-pVQZ results.[225] The TINKER v6 and Amber v10 molecular modeling package were used for all the molecular mechanics calculations. Particle-Mesh Ewald summation[182, 183, 219] was applied for treating the electrostatic interactions, with a

real-space cutoff distance of 7.0 Å, grid spacing of 0.8 Å, and a 5^{th} order polynomial. A cutoff with a switching window at 12 Å was applied to the vdW interactions. The induced dipoles, which were also computed with PME, were iterated until the root mean square (RMS) changes were below 0.01 Debye per atom. All the molecular dynamics simulations were performed using an integrator based on Velocity Verlet algorithm.[226] The RESPA algorithm was implemented to enable a 2.5 fs time step. The system temperature was controlled via the Nose-Hoover chain thermostat.[227]

For alanine, glycine, proline and some terminals (Ala-COOH, Gly-COOH, Gly-COO^-), the minimum energy map of the dipeptide was calculated on a uniform 15º grid in the ϕ-ψ space. At each of the 576 points, MP2/6-31G* geometry optimization with constrained ϕ and ψ was performed before a single point energy calculation at the RI-TRIM/MP2 CBS level. In proline, fewer grid points were available for QM calculations due to the limited degree of conformational freedom. For the terminals, the single point energy was also calculated for each optimized structure using the Polarizable Continuum Model (PCM) quantum. The torsion parameters for these model compounds were fit to gas-phase *ab initio* conformational energy first and then adjusted based on the statistical populations sampled from Protein Data Bank (PDB). For the side chain torsions of all other residues, geometry optimization was performed at MP2/6-31G* level with the specific torsion angle constrained at every 30º from 0º to 360º, followed by the single point RI-TRIM MP2/CBS energy calculations.

The potential of mean force (PMF) of a solvated alanine tripeptide, NH_3^+-Ala-Ala-Ala-COO^-, with respect to ϕ and ψ of the middle Ala was computed using the two-dimensional weighted histogram analysis method (2D-WHAM). [228-230] A total of 576

(on the same grid as in the gas-phase map) independent molecular dynamics simulations of alanine dipeptide plus 206 water molecules in a 26.6 Å octohedron box was carried out at 298K. In each simulation, the ϕ and ψ dihedral angles were restrained to one of the grid point on Ramachandran map using weak harmonic potentials (force constant = 0.01 kcal/mol-deg^2). The resulting alanine conformer population, sampled from the 576 × 70 ps trajectory (after 30 ps equilibration), was utilized to construct the PMF or the relative free energy map via the 2D WHAM.

The PDB PMF was calculated from $-ln(P)$ where P is the torsion distribution sampled from PDB.[231] For alanine backbone, the PDB PMF was obtained by averaging the data for alanine with either right or left neighbor residue being alanine (Ala-**Ala**-X or X-**Ala**-Ala, X represents any type of residue, same below). For proline, the data with either right or left neighbor residue being glycine (Gly-**Pro**-X or X-**Pro**-Gly) was averaged. Similarly, the glycine PDB PMF was calculated by averaging the data for Pro-**Gly**-X and X-**Gly**-Gly.

For peptide systems including unblocked and protonated $(Ala)_5$, NH_3^+-Gly-Pro-Gly-Gly-COO^-, and Ac-(Ala-Ala-Glu-Ala-Ala)$_3$–NH_2. Replica exchange molecular dynamics (REMD)[232, 233] simulations were performed with 36 replicas at temperatures between 278 K and 620 K. The $(Ala)_5$ was unblocked and protonated at both N- and C-termini, corresponding to the experimental conditions of pH 2.[99, 234] For each system, the peptide was soaked in an octahedron water box with ~800 water molecules. The MD simulations were performed under the NVT ensemble for at least 30 ns/per replica, using the PMEMD module in Amber 10.[221] The snapshots were saved every 0.5 ps for analysis purpose.

Protein molecular dynamics simulations were performed using Amber10 software package. The proteins simulated are as follows: Crambin (PDB:1EJG, 46 residues),[235] Trp Cage (PDB:1L2Y, 20 residues),[236] Villin Headpiece (PDB:1VII, 35 residues),[237] Ubiquitin (PDB:1UBQ, 76 residues),[238] GB3 Domain (PDB:2OED, 56 residues),[239] RD1 Antifreeze Protein (PDB:1UCS, 64 residues),[240] SUMO-2 Domain (PDB:1WM3, 72 residues),[239] BPTI (PDB:1BPI, 58 residues),[241] FK Binding Protein (PDB:2PPN, 107 residues),[242] and lysozyme (PDB:6LYT, 129 residues).[243] For each system, the protein was soaked in a periodic octahedron box with a buffer distance of 8 Å to the wall. The simulation was performed in the NPT ensemble at a temperature of 298K. Approximately 10 ns of MD trajectories were generated for each system. The first 500 ps portion of the trajectory was discarded corresponding to equilibration and structural information was saved every 0.5 ps. Analyses of the dynamics trajectories and computation of RMSD values and average structures was done using the Ptraj module.[221]

3.4 PARAMETER DERIVATION

3.4.1 Electrostatic Parameters

The permanent atomic multipoles for glycine, alanine and proline residues were derived from a capped (MeN-X-COMe) dipeptide. The goal is to derive the conformational independent permanent electrostatic parameters for the central residue X from QM. It is a critical step to define the intramolecular direct polarization groups since the intramolecular polarization contribution needs to be extracted from the DMA multipoles as described in the Potential Energy Model section. Recall the permanent atomic multipoles (PAMs) only polarize atoms outside its polarization group. **Figure 3.2**

shows the group definition for alanine. For side chains of other residues, the groups are chosen in the same spirit, i.e. no freely rotatable bonds within the group. For example, the -CH_2-, phenyl ring, and hydroxyl group are each classified as one group in the direct polarization. Next, the initial multipole parameters were derived by using DMA at MP2/6-311G** level, which were then optimized against the MP2/aug-cc-pVTZ electrostatic potential computed on a set of grid points around the diepeptide compounds. The grid spacing is set as 0.35 Å with a 1.0 Å grid distance from the vdW surface. The parameters are iterated until the RMS differences from both the target QM potential and gradient are smaller than 0.5 kcal/mol/ Å. Note that the point charges in the optimization process are fixed while the dipole and quadrupole moments are relaxed. The point charges at the boundary between middle residue and caps, or backbone and side chain are manually adjusted to ensure each residue's net charge is zero. The adjustment is usually small and the ESP optimization process also ensure its compensated by dipole and quadrupole changes. The resulting PAMs for the central residue are the "confirmation-independent" PAMs included in the AMOEBA parameter file. When connected with caps or other residues in the peptide and proteins, the intramolecular polarization due to PAMs will take place according to the group definition. For alanine, we chose five local minima (α_L, α', C5, C7a and C7e) to repeat the above procedure using the average as the final multipoles, and one additional conformer (β2) as validation. PAMs for glycine and proline were obtained by following the same procedure.

For all other residues, the capped (MeN-X-COMe) dipeptides are chosen to be the model compounds. For each amino acid, three conformations were used for parameterization, and another three for validation. The conformers have either extended

backbone or compact structures with internal hydrogen bonding as they are gas-phase minima. The ba ckbone atoms have the same PAM as alanine, while the side chain PAMs were derived from the dipeptides using the same procedure mentioned above. When the side chain parameters were merged with alanine backbone parameters, the charge neutralization (in the order of hundredth electron) was made on the C_β. Fitting to MP2/aug-cc-pVTZ electrostatic potential was subsequently done to the three conformers in the parameterizations set by only allowing the side chain atomic dipole and quadrupole moments to vary. The ESP fitting and validation results will be discussed in details below in the Simulation and Validation section. In addition to ACE (-$COCH_3$) and NME (-$NHCH_3$) terminus, other N-terminus (-NH_3^+ and -NH_2) and C-terminus (-COOH and -COO^-) were also parameterized. Note that glycine needs its own -NH_3^+, -COOH due to the different atom type used for glycine Ca. Due to the amount of data involved, most parameters, including the PAMs and direct polarization group definition, are not given in the text, but are included in the publicly available AMOEBA parameter file.

3.4.2 vdW and Valence Parameters

The van der Waals and valence parameters were transferred from the AMOEBA parameters for small organic molecules which was reported in a previous publication.[55] The parameterization essentially followed the same approach used for water[19] and ion[52] where the vdW parameters were optimized to both gas-phase cluster structures and energetics as well as condensed-phase properties. A critical strategy in deriving the vdW parameters, due to its empirical nature, is to ensure chemical consistency among different elements. This was achieved by parameterizing multiple compounds that share the same vdW "classes" simultaneously to improve the transferability. The "atom

classes" (super set of "atom types" used for electrostatics) in AMOEBA are slightly broader than the element types in order to take into account the different hybridizations such as sp^2 vs. sp^3. In Error! Reference source not found., the main vdW parameters are llustrated. The valence parameters, including bond, angle and out-of-plane parameters, were transferred from small organic molecules with minor modifications according to *ab initio* (MP2/6-31G*) geometries of peptides and protein PDB structures.

3.4.3 Torsional Parameters

Once the electrostatic, vdW, and valence parameters were determined, the last step was to derive the backbone torsional parameters by comparing AMOEBA and *ab initio* conformational energy values. Note that the molecular mechanics conformational energy not only depends on the torsional energy term, but also the treatment of nonbonded intramolecular interaction, in particular how the 1-4 interactions are handled. In this work the scaling factors for the intramolecular electrostatic and vdW interactions have been chosen so that they 1) minimize (maximize) the contribution of torsional (nonbonded) terms, and 2) also transfer well from dipeptides to tetrapeptides.

The alanine dipeptide is used to parameterize the backbone torsions for all the amino acids except glycine and proline. The *ab initio* (RI-TRIM MP2/CBS) energy of alanine dipeptide was systematically evaluated at different backbone torsion angles over a 24 × 24 grid (15^o interval in both ϕ and ψ) as described in the computational details. The AMOEBA energy without the ϕ/ψ torsional contribution was computed for the same conformation using torsion constraints. The difference between AMOEBA and RI-TRIM MP2/CBS energy is taken as the fitting target of the torsional parameters using a three-term Fourier expansion. Subsequently 2D WHAM simulations of an $(Ala)_3$ peptide in

explicit water were performed and the Ramachandran potential of mean force (PMF) of the middle residue were obtained. The torsion parameters are further improved by comparing the AMOEBA PMF to the statistical alanine backbone PMF derived from the PDB database.[231] Note that there torsion parameters were not directly fit to the PDB PMF. Instead, the parameter refinement was achieved by assigning relatively higher weight factors to the QM energy of conformers located at the polyproline II (PII), α-helical and β-sheet regions than the other while fitting to the whole QM gas-phase energy map. The torsion parameters are fine-tuned in 3-4 iterations to balance the simulated relative populations in these minimum-energy regions. The gas phase Ramachandran potential energy from RI-TRIM MP2/CBS and AMOEBA with the final torsion parameters are compared in **Figure 3.3**. The simulated solution-phase PMF using AMOEBA and the PDB statistical PMF maps for alanine backbone are shown in **Figure 3.4a**.

The parameterization of proline backbone torsion followed essentially the same procedure as alanine except that fewer grid points are used due to the limited conformational freedom. For glycine, a torsion-torsion spline term is introduced in addition to the Fourier torsional terms for ϕ and ψ. After the Fourier torsions were fit to the gas-phase *ab initio* RI-TRIM MP2/CBS energy, the difference between the *ab initio* and AMOEBA energy were used as the 2-D spline parameters which were fixed in the subsequent optimization of the Fourier torsion parameters. The use of torsion-torsion term improves the fit to both *ab initio* data and solution phase properties. Similar to the parameterization of alanine backbone torsions, the torsion parameters for proline and glycine backbone were refined to match the statistical PMF of proline and glycine from

the PDB database, respectively. REMD simulations were performed with a model tetra-peptide GPGG (NH_3^+-Gly-Pro-Gly-Gly-COO^-) to obtain the simulated torsion angles distribution of proline and glycine backbone. All other residues share the same backbone torsion parameters (together with other valence, vdW and electrostatic parameters) as alanine. The parameterization of the –COOH terminal of alanine (and other non-Gly/Pro residues) and –COOH, -COO^- terminals of glycine were also fit to the RI-TRIM MP2/CBS energies on a 12 × 12 torsional grid Since there are no PDB data available, we have optimized these parameters by matching AMOEBA energy in implicit solvent with QM PCM energies at MP2/6-311G(2d,2p) level.

In addition, conformation energies for a benchmark set of 27 alanine tetrapeptides[244] have been assessed. This comparison was made to validate the transferability and adjusting scaling factors for the short-range intramolecular nonbonded interactions. The dipeptide data itself is obviously not useful for this task as its conformational energy surface has been explicitly fit to. The AMOEBA results are compared with those from MP2, LMP2, DFT and RI MP2 calculations in Error! eference source not found.**3.2**. All the *ab initio* calculations are single point energy evaluation of the same HF/6-31G** geometries. AMOEBA calculations were performed with both full geometry optimization and optimization with ϕ and ψ angles constrained at HF geometry. All comparison is made against RI MP2/CBS results. While the AMOEBA-optimized structures deviate only slightly from those of HF/6-31G** (average srms = 0.47 Å), the rms difference between AMOEBA and RI MP2/CBS energies is 1.15 kcal/mol, similar to those of LMP2/cc-pVTZ(-f) and MP2/6-311+G2d2p. Note that the relative conformational energies of the first two conformers (extended) vs. the third

(compact) given by RI MP2/CBS lie in between the canonical MP2 and LMP2 results, and so are the AMOEBA perditions.

For the side chain torsions for all other residues, the parameters were obtained by fitting to the RI-TRIM MP2/CBS conformational energy of dipeptides. To derive parameters that are independent of the backbone conformation and also to verify the transferability of the backbone torsion parameters from alanine, two or more dipeptide conformation with backbone ϕ/ψ values fixed at α-helix and β-sheet values were used simultaneously to fit the selected side chain χ torsion parameters. For each conformer, the side chain torsion angle of interest was rotated at a step size of 30º. At each point, the dipeptide structure was optimized with ϕ, ψ and χ angles constrained to the same values in both *ab initio* level and AMOEBA calculations. Examples of conformational energy profile of isoleucine and serine side chain are plotted in **Figure 3.5**. The former has an alkane-like side chain, and a single set of torsional parameters performs rather well on both conformers. The order of the local minima is also well captured. The situation for serine on the other hand is a little worse due to intramolecular hydrogen bonding between the side chain and backbone. In both cases, the torsional energy contribution accounts for about 15% of the total conformational energy. In Error! Reference source not found., the veraged RMS error in conformational energy fitting for each residue is given.

3.5 SIMULATION AND VALIDATION

For small molecules, it is possible to systematically compare force field calculations to a range of *ab initio* data and experimental properties, which allows us to securitize the individual components in the potential energy model such as electrostatics vs. vdW. For larger biomolecules, however, one can only compare detailed simulations

with limited experimental data such as statistical population, NMR *J* coupling constants, and atomic structures of native proteins from X-ray or NMR. Such comparison, while absolutely necessary, is mostly useful for inspecting the torsional parameters while providing no feedback to other components (e.g. vdW or electrostatics) within the force field. It would require extensive investigations of a wide range of proteins in various areas, such as protein folding, binding, pKa shift and mutant stability, to fully validate the different components and aspects of a force field. Below we first examined the electrostatic parameters against QM for multiple conformations of each amino acid in gas-phase, and then assessed peptide conformational properties and protein structures in solution using the AMOEBA force field.

3.5.1 Electrostatic Properties in Gas Phase

One of the important advantages of a polarizable force field is its transferability from gas phase to solution. Therefore, it is to our advantage that the gas-phase electrostatic properties can be rigorously compared to QM *ab initio* results. Especially, we would like to ensure the transferability of the alanine backbone electrostatic multipoles to the other residues except for glycine and proline, and the transferability among different conformations for each amino acid. Here, we validated the electrostatic properties of AMOEBA protein force field by computing the dipole moments, including the x, y, and z components, and electrostatic potentials of dipeptide mode compounds for each amino acid. For each amino acid dipeptide, three conformations were chosen as the validation set, which were excluded in the parameterization process. In **Figure 3.6**, we compare the AMOEBA and QM molecular dipole moments of these amino acid dipeptides in the validation set (except ala, gly and pro). The dipole components

including the x, y, and z components of all dipeptides were accurately reproduced, regardless of the conformations and residues. The correlation between AMOEBA and *ab initio* results is 0.998. The comparison between *ab initio* and AMOEBA electrostatic potential for all dipeptide model compounds can be found in the supporting information, for both parameterization and validation sets. The average RMSE is 0.45 kcal/mol per unit charge on a grid surrounding the neutral amino acids, and only slightly higher (0.64) for charged dipeptides, with the absolute values of the latter orders of magnitude higher. Thus, thanks to the intramolecular polarization model in AMOENA, the transferability of backbone and side chain electrostatic multipoles of AMOEBA are very satisfactory. We believe this test performed here, while not commonly done in force field development, is rather important and necessary for validating the electrostatics of the force fields, before other components such as the torsion parameters are empirically adjusted.

3.5.2 Polyalanine Conformational Free Energy in Solution

There has been increasing studies of oligopeptide conformational properties in solution to calibrate the force field torsional parameters.[99, 245-251] Simulated results can be directly compared to the experimental nuclear magnetic resonance (NMR) data for the corresponding peptides. Following the previous work, we have performed simulations on Ala/Gly/Pro based peptides using the current AMOEBA protein force field.

For alanine, we have first examined the solvation of unblocked and protonated $(Ala)_5$ peptide using REMD. The conformational preference is presented as a potential of mean force with respect to ϕ and ψ in **Figure 3.4b**, which is calculated from the averaged ϕ and ψ torsion population distributions of ala-2, ala-3, and ala-4 residues. A distinct global minimum is located around the PII conformation. Two other basins with

energies about half kcal/mol higher are in the β-sheet and α-helix region, respectively. The energy barrier between global and the two local minima is about 1~2 kcal/mol. Overall, the upper left region of the Ramachandran map is distinctively flat compared to the rest of conformational space. The shape and location of this highly populated vicinity agree well with the statistical PMF map from the PDB database (**Figure 3.4c**), [231] suggesting the transferability from $(Ala)_3$ to $(Ala)_5$ as we expected.

The distributions of ϕ/ψ torsion angles of $(Ala)_5$ have been probed experimentally by NMR.[99] The NMR spin-spin coupling (*J*-coupling) constants reflect the ensemble character of the conformational distribution, which were compared with those calculated form REMD simulation trajectories of $(Ala)_5$ via Karplus relations.[87, 250] Totally eight NMR *J*-coupling constants were reported: five for the backbone angle ϕ, $^3J(H_N,H_\alpha)$, $^3J(H_N,C')$, $^3J(H_\alpha,C')$, $^3J(C,C')$, $^3J(H_N,C_\beta)$, two for the backbone angle ψ, $^1J(N,C_\alpha)$, $^2J(N,C_\alpha)$ can be measured, and one for both ϕ and ψ $^3J(H_N,C_\alpha)$.[99] The trajectory of 298 K in $(Ala)_5$ REMD simulation was extracted to calculated the predicted *J*-coupling values. Twenty-seven predicted *J*-coupling values are compared to those measured by NMR experiments in **Table 3.4**. The *J*-coupling constants involved in the N- and the C-termini (COOH) were also included. The predicted *J*-coupling values are in excellent agreement with those probed by the experiments. The chi-square (χ^2) difference between the simulations and experiments, computed using the experimental uncertainties,[99] is about 0.994, and the overall RMS difference is 0.33 (**Table 3.4**). Note that when using the torsion parameters that were directly fit to the gas-phase QM energy of alanine dipeptide, the χ^2 is 3 or 4 times higher. The torsion refinement according to the PDB PMF thus has a significant effect in improving the calculated *J*-

coupling constants. A notable consequence of the adjustment is the location of α-helix population from simulations shifted lower and to the right toward the (φ, ψ) angles in the PDB distribution. In contrast, with torsion terms fit to QM gas-phase energy alone, the simulated "α-helix" population was much broader than that of PDB and centered at much lower (more negative) φ and higher (less negative) ψ. A similar effect has recently been discussed for the CHAMRMM 22/CMAP, CHARMM36-MP2 and CHARMM36 force fields and it was suggested empirical correction to CMAP approach is important.[247, 248] Improving the agreement with PDB distribution both in terms of shape and location, especially for residues not in actual helices, led to thermodynamic properties and cooperativity in helix-coil transition that were more consistent with experiments.

3.5.3 Proline and Glycine Conformational Free Energy in Solution

The ϕ/ψ torsion angles distribution for proline and glycine backbone were validated by via REMD simulations of a tetra-peptide GPGG (NH_3^+-Gly-Pro-Gly-Gly-COO^-). For both proline (pro-2 residue) and glycine (gly-3 residue), the simulated PMF maps with respect to the ϕ/ψ torsion angles show good agreement with the PDB statistical PMF maps (**Figure 3.7**). For Pro-2 residue, the relative free energy of the α-helix and PII regions from the PDB data are well reproduced in our simulation and the local minimum in the α-helix region is about 1 kcal/mole higher than the global minimum in the PII region (**Figure 3.7a**). The torsion distributions of Gly-3 residue from the simulations are also consistent with the PDB data. The global minima are located at the α-helix and the left-handed α-helix regions. Two local minima are located at the PII and the reflection of the PII regions, with about 1 kcal/mole higher than the global (**Figure 3.7c**).

Similar to alanine, the *J*-coupling constants were calculated for Pro-2 and Gly-3 residues. Three *J*-coupling constants, $J(H_\alpha,C')$ for Pro-2, $J(H_\alpha,H_N)$ and $J(H_\alpha,C')$ for Gly-3 were evaluated by using the Karplus coefficients obtained from B972 EPR-III and B3LYP EPR-III calculations.[252] **Table 3.5** compares the *J*-coupling values obtained from the simulations and experiments for the GPGG tetra-peptide. The RMS difference between the calculated and experimental *J*-coupling constants is 0.44 (with B972) and 0.39 (with B3LYP), respectively.

3.5.4 Secondary Structure Distribution in Ac-$(AAQAA)_3$-NH_2 Peptide

It is important to accurately reproduce the distributions and balances of peptide secondary structures in a force field, which is directly related to phenomena such as protein folding, mis-folding, aggregation, and conformational changes. The REMD simulation of short peptide $(Ala)_5$ has already provided the sensitive test of the "intrinsic" secondary structure preferences of the AMOEBA force field. However, $(Ala)_5$ peptide is too short to form a stable α-helix, in which a more extended PII backbone conformation is found more favorable than the right handed α-helix in the Ramachandran free energy map.[99, 234, 250, 253-257] We have further simulated a longer helix-forming peptide, Ac-$(AAQAA)_3$-NH2,[258, 259] to investigate the helix-coil transition. Because the helical populations of Ac-$(AAQAA)_3$-NH2 have been determined from NMR chemical shift data,[260] we can directly compare the helical-fractions calculated from REMD simulations. We use the same definition of helical state as the previous study,[234] where φ is belongs to [-160°, -20°] and ψ is belongs to [-120°, 50°]. **Figure 3.10** shows the fraction of helix $\langle h_i \rangle$ for each residue from our simulation and NMR chemical shifts at 303 K. A reasonable agreement can be seen between the simulated values and those

obtained from NMR experiments at 303K, where both fluctuated between 10 to 30% for most residues. The error was estimated using block average based on six 5-ns simulation windows. In addition, the lower helical propensity trend at the C-terminus was well captured. AMOEBA seems to underestimate the helix fractions of residue 3, 4 and 7.

3.5.5 Molecular Dynamics Simulations of Protein Systems

Ten well-studied proteins were chosen as the validation set to evaluate the new force field. While limited, this set is somewhat representative of different types of protein structures. For example, Trp Cage (1L2Y) and Villin Headpiece (1VII) mainly consist of alpha helices; GB3 domain (2OED) and FK binding protein (2PPN) contain more beta sheets; Crambin (1EJG) and Lysozyme (6LYT) are disulfide bond-rich domains; the others are mixtures of different motifs.

The stability of the protein systems is characterized by the RMSD values relative to the PDB structures after 10 ns MD simulation, as summarized in **Figure 3.8**. The overall average RMSD of the ten simulated protein structures is 1.33 Å, and seven of them are ~1.0 Å. All the systems showed stable RMSD after ~300 ps, indicating the protein structures were well maintained. As a further demonstration, the final snapshot is superposed to the PDB structure as a comparison for all systems. From **Figure 3.9**, we can see that no helical and sheet structures drift away from experiment significantly. For some proteins, the large deviations usually come from the terminals. Take Ubiquitin as an example, inspection of the PDB entry for ubiquitin shows that the final 5 residues have significantly higher temperature factors and the rest of the structure. Flexibility near the C-terminus is also observed in MD results, particularly for the two glycine residues at

positions 75 and 76. Elimination of just these two residues from the data analysis yields a backbone structural RMSD of 1.16 Å against the PDB structure.

3.5.6 Calculation of NMR Order Parameters

The order parameter (S^2) indicates the flexibility of each residue, and lower S^2 values correspond to higher flexibility. We have compared the simulated order parameters of ubiquitin and hen egg white lysozyme to NMR relaxation experiments,[261, 262] which is measured based on the amide N-H bond vibrational motion.[263, 264] The isotropic reorientational eigenmode dynamics (iRED) matrix was extracted from the MD trajectories. The order parameter (S^2) was then computed by solving the Eigen values of the matrix.[265]

Shown in **Figure 3.11** is the comparison of S^2 from 10 ns MD simulations and experimental measurements. The RMSDs from the experimental data are 0.04 and 0.09 for ubiquitin and lysozyme respectively. The S^2 curves in general correlated well with the NMR results. For ubiquitin, the higher flexibility at turn 1 (residue 7-10), turn 3 (residue 37-40) and the terminal was nicely reproduced, while the flexibility at turn 6 (residue 62-65) was underestimated. For lysozyme, the flexibility at turn 1 (residue 46-49), long loop 2 (residue 61-78), loop 3 (residue 85-89), loop 4 (residue 100-107), loop 5 (residue 116-119) and the terminal were reasonably predicted, although the intensities were a little higher than the experimental results. It is satisfying that the trend of the predicted order parameters match well with the experiment, and no remarkable deviation was observed. Overall, the AMOEBA force field is able to reproduce the protein structures and flexibilities

3.5.6 Calculation of Side Chain J-Couplings

The amino acid side chain distributions were evaluated by comparing the simulated side chain *J*-couplings to the experimental NMR values for four protein systems: BPTI, GB3, Ubiquitin and Lysozyme.[262, 266-269]

Table 3.6 lists the side-chain the RMSD between the experimentally derived and simulation-derived scalar couplings. Also listed for comparison are the RMSD values derived using AMBER ff99SB force field, and AMBER ff99SB-ILDN,[270] which were refined against the side chain NMR data of the four proteins. Note that the side chain parameters in the current AMOEBA were directly from fitting to QM energy profile and were not optimized against any experimental data. Overall, the performance of AMOEBA force field is comparable to the ff99SB-ILDN force field. Although a few outliers are found in each protein system, most of the experimental scalar couplings were reasonably reproduced. The correlation between the calculated and experimental *J*-couplings of all the four protein systems is summarized in **Figure 3.12**. The correlation coefficient R^2 is 0.75. The *J*-coupling values of each protein system can be found in the Supporting Information. We can see that within the high scalar-coupling region, AMOEBA failed to distinguish the subtle differences as observed in experiment. For example, most predicted *J*-coupling values by AMOEBA were ~10.8 Hz for experimental data ranging from 8.0 to 15.0 Hz. We further decomposed the RMSD values to each residue. The values ranged from ~0.7 to 3.3 (**Figure 3.12**). For residues with relatively high RMSDs, like Ile and Gln, we did not make further modifications since the sample size of these residues in the four proteins is small. The side chain *J*-coupling data shown above suggests that the AMOEBA force field, derived based on the gas phase QM energy profiles, reasonably captures the side chain conformational

properties in solution while further improvement would require additional reliable experimental data each residues.

3.6 CONCLUSIONS

The development and parameterization of AMOEBA protein force field has been reported in this chapter. A distinct feature of the force field is the atomic multipole-based electrostatics with explicit treatment of dipole polarization. A mutual induction model with damping was applied to describe both inter and intramolecular polarization in a consistent manner. The polarization among permanent multipoles is handled via a group-based scheme while the induced-induced dipole polarization occurs among all atoms (polarizable sites). By extracting the intramolecular polarization as defined by the force field, we are able to derive conformation-independent electrostatic multipole parameters from high-level first principle calculations by using a combination of Distributed Multipole Analysis and electrostatic potential fitting. The rigorously derived electrostatic model will be important for accurate description of protein interactions with other biomolecules as well as the electrostatic forces within the proteins.

With the vdW parameters determined and transferred from liquid simulations of small organic molecules [55], AMOEBA was applied to the simulations of peptides and proteins. In addition to the non-bonded electrostatic and vdW forces, the “torsional” term also contributes to a fraction of the conformational energy. The torsion term is essentially an error function in the classical force field, and yet it plays a crucial role in determining the detailed conformational properties of peptides and proteins. The recent development of Amber [245, 270] and CHARMM force fields[247, 248] have demonstrated that the conformational populations of small peptides are extremely sensitive to subtle changes (a

fraction of kcal/mol) in the torsional parameters. In developing the current force field, we have resorted to both high-level *ab initio* (MP2/CBS) peptide energetics and PDB structural statistics in deriving the backbone torsion parameters. The resulting force field overall performed reasonable well compared with NMR *J* coupling constants and S^2 order parameters of several peptide and protein systems. Nonetheless, these are limited validations focusing on conformational properties and torsional parameters. Extensive investigations on more proteins and a broad range of thermodynamic properties will be necessary to understand the various aspect of the potential energy model and to fully determine the successes and failures of the force field.

As previously noted,[247, 271] the CMAP style spline torsion allows a force field to reproduce the gas-phase *ab initio* conformational energy exactly. This however may also pick up unphysical errors in the force fields (e.g. in the other valence contributions) that are not transferable to the solution phase. While we have strived to derive a balanced and physical force field, further understanding of the limitations of the molecular mechanics force fields is essential for systematic improvement. Better modeling of electrostatics and vdW, as well as their coupling of valence interactions, will be the key to arrive at a transferable protein force field that can perform well in various physical and chemical environments.

Table 3.1: The vdW parameters for atoms in protein backbone. AMOEBA uses atom classes to define vdW and valence parameters while atom types are used for electrostatic parameters. Different atom types may belong to the same atom class.

Atom class	***2r* (Å)**	**ε (kcal/mol)**	**H-reduction factor**
C_α	3.650	0.101	
H_α	2.940	0.026	0.91
C_β	3.820	0.101	
H_β	2.980	0.024	0.92
N (amide)	3.710	0.110	
H (amide)	2.590	0.022	0.90
C (carbonyl)	3.820	0.106	
O (carbonyl)	3.300	0.112	
O^- (in $-COO^-$)	3.700	0.129	
S	4.005	0.355	
S^-	4.200	0.355	

Table 3.2: Comparison of the side chain conformational energy (kcal/mol) calculated by AMOEBA and QM (RI-TRIM MP2/CBS). The conformational energies were calculated by rotating the listed side chain torsion from 0 to 360º at a 30º interval. In all RMSD calculations, data from two backbone conformations (α-helical and β-sheet) were combined.

	Side chain torsion	**RMSD**		**Side chain torsion**	**RMSD**
Cys	χ_1	0.67	Asn	χ_1	1.35
	χ_2	0.10		χ_2	1.06
Met	χ_1 same as Glu and Gln	0.40	Glu	χ_1 from Met	0.79
	χ_2	0.52		χ_2 from Gln	1.86
	χ_3	0.10		χ_3 from Asp χ_2	0.42
Ser	χ_1	0.84	Leu	χ_1	0.32
	χ_2	0.22		χ_2 averaged over Leu/ Ile	0.31
Thr	χ_1 same as Ser and Val	0.96	Val	χ_1	0.58
His	χ_1	1.24	Phe	χ_1	0.65
	χ_2	1.31		χ_2	0.54
Hid	χ_1	0.83	Tyr	χ_1 from Phe	0.55
	χ_2	0.67		χ_2	0.44
Hie	χ_1	1.10	Trp	χ_1	0.54
	χ_2	1.00		χ_2	0.55
Ile	χ_1 from Val	0.53	Asph	χ_1	1.39
	χ_2 averaged over Leu/Ile	0.34		χ_2	0.97
Arg	χ_1 from Met	1.63	Lys	χ_1 from Asn	0.74
	χ_2 same as Lys	1.08		χ_2	0.95
	χ_3	1.76		χ_3	0.96
	χ_4	0.80		χ_4	0.73
Asp	χ_1 from Asph	1.22	Gln	χ_1 from Met	0.79
	χ_2	1.01		χ_2	1.86
				χ_3 from Asn χ_2	0.42

Table 3.3: Comparison of alanine tetrapeptide conformational energy (kcal/mol). This set conformers were used in the previous studies. The RI MP2/CBS and other QM results are taken from taken from Two sets of AMOEBA results were listed. Both are from energy minimization of QM structures, one with backbone ϕ/ψ torsions restrained at the QM values and the other fully relaxed. The RMSD was computed using the RI MP2/CBS energies as references.

RI MP2/ CBS	LMP2/ cc-pVTZ(-f)	MP2/ 6-311+G2d2p	DFT B3LYP/6-31G*	AMOEBA	Struct. RMS (Å)	AMOEBA (/ restrained)
4.13	2.50	4.61	1.62	3.07	0.30	2.54
4.19	2.60	4.21	1.71	3.62	0.42	0.74
0.57	0.00	-0.70	-1.00	0.00	0.21	0.33
5.73	3.87	5.50	3.61	4.07	0.37	3.82
5.26	3.88	5.14	4.25	3.96	0.30	2.27
2.90	2.19	2.10	2.10	2.45	0.53	0.14
6.67	5.73	5.61	6.56	7.64	0.45	0.65
4.64	4.17	3.32	4.99	5.45	0.44	1.06
7.92	6.93	6.98	5.20	10.01	0.25	3.14
7.79	6.99	6.57	7.24	6.34	0.34	0.62
0.00	-0.19	-1.41	0.14	0.75	0.68	0.58
0.29	0.50	-1.07	1.73	0.75	0.91	0.22
3.66	1.77	3.20	1.14	3.56	0.62	0.02
4.68	3.68	4.14	3.89	4.66	0.71	0.00
2.19	2.07	0.65	3.47	2.28	0.59	0.08
3.55	2.83	2.33	3.31	2.93	0.48	0.24
3.42	2.78	2.02	2.00	2.32	0.28	1.09
1.91	0.52	1.15	-0.87	2.19	0.56	0.20
3.82	2.83	2.90	1.13	4.25	0.56	0.19
1.76	0.87	0.88	0.80	3.18	0.47	2.91
2.92	2.11	1.59	1.78	0.00	0.92	8.51
5.82	4.82	4.59	4.84	6.87	0.59	1.60
5.82	4.82	4.57	4.84	6.84	0.33	1.46
3.98	2.98	2.89	3.59	4.11	0.30	0.19
2.50	1.59	1.54	1.92	2.87	0.50	0.35
0.67	0.18	-0.41	1.40	1.60	0.37	1.51
4.02	3.18	3.04	3.53	6.26	0.38	5.57
RMS deviation	1.05	1.06	1.54	*1.15*	*0.47*	*1.22*

Table 3.4: Comparison of *J*-coupling values (Hz) from the AMOEBA simulations and experiments for $(Ala)_5$ peptide. The $(Ala)_5$ was unblocked and protonated at both N- and C-termini, corresponding to the experimental conditions of pH 2. Replica exchange molecular dynamics simulations were performed with 32 replicas at temperatures between 278 K and 620 K (30-ns for each replica). The trajectory at 298 K was extracted for the *J*-coupling calculation.

Rsidue index	*J*-coupling type	*J*-simulation	*J*-expt.[99]
Ala-2	$^1J(N,C_\alpha)$	11.066	11.36
Ala-3	$^1J(N,C_\alpha)$	10.923	11.26
Ala-4	$^1J(N,C_\alpha)$	10.922	11.25
Ala-2	$^2J(N,C_\alpha)$	8.448	9.20
Ala-3	$^2J(N,C_\alpha)$	8.170	8.55
Ala-4	$^2J(N,C_\alpha)$	8.232	8.40
Ala-5	$^2J(N,C_\alpha)$	8.250	8.27
Ala-2	$^3J(C',C')$	0.866	0.19
Ala-2	$^3J(H_\alpha,C')$	1.729	1.85
Ala-3	$^3J(H_\alpha,C')$	1.705	1.86
Ala-4	$^3J(H_\alpha,C')$	1.713	1.89
Ala-5	$^3J(H_\alpha,C')$	1.929	2.19
Ala-2	$^3J(H_N,C')$	1.087	1.13
Ala-4	$^3J(H_N,C')$	1.315	1.15
Ala-5	$^3J(H_N,C')$	1.216	1.16
Ala-2	$^3J(H_N,C_\beta)$	1.819	2.30
Ala-3	$^3J(H_N,C_\beta)$	1.833	2.24
Ala-4	$^3J(H_N,C_\beta)$	1.743	2.14
Ala-5	$^3J(H_N,C_\beta)$	1.584	1.96
Ala-2	$^3J(H_N,H_\alpha)$	6.269	5.59
Ala-3	$^3J(H_N,H_\alpha)$	5.988	5.74
Ala-4	$^3J(H_N,H_\alpha)$	6.079	5.98
Ala-5	$^3J(H_N,H_\alpha)$	6.607	6.54
Ala-2	$^3J(H_N,C_\alpha)$	0.421	0.67
Ala-3	$^3J(H_N,C_\alpha)$	0.614	0.68
Ala-4	$^3J(H_N,C_\alpha)$	0.648	0.69
Ala-5	$^3J(H_N,C_\alpha)$	0.663	0.73
X^2 = 0.994		**RMS=0.33**	

Table 3.5: Comparison of *J*-coupling values (Hz) from AMOEBA simulations and NMR experiments for GPGG tetra-peptide. Replica exchange molecular dynamics simulations were performed with 32 replicas at temperatures between 278 K and 620 K (30-ns for each replica). The trajectory at 298 K was extracted for *J*-coupling calculation.

Residue index	J-coupling type	*J*-simulation (B972 EPR-III)	*J*-simulation (B3LYP EPR-III)	*J*-expt.[252]
Pro-2	$J(H_\alpha,C')$	1.75	1.88	1.30
Gly-3	$J(H_\alpha,H_N)$	4.94	3.67	4.10
Gly-3	$J(H_\alpha,C')$	6.07	6.76	6.30
		RMS=0.44	RMS=0.39	

Table 3.6: Comparison (RMSD) of *J*-coupling values (Hz) from AMOEBA simulations and experiments for BPTI, GB3, Ubiquitin and Lysozyme. Results from AMBER FF99SB force field, and refined AMBER FF99SB-ILDN force field are included for comparison.

	BPTI	GB3	Ubiquitin	Lysozyme
AMOEBA	1.745	1.47	1.55	2.47
AMBER FF99SB[270]	1.782	1.48	2.55	3.59
AMBER FF99SB-ILDN[270]	1.452	0.90	1.63	2.93

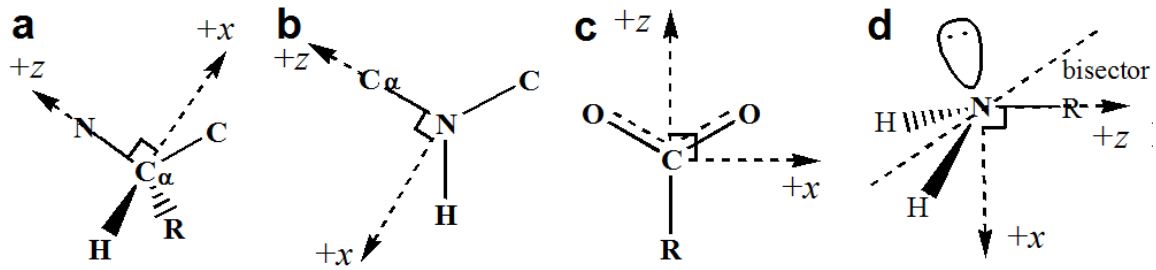

Figure 3.1: Local frame definitions for (a) a protein backbone C_α, (b) backbone amide N, (c) carboxylate carbon, and (d) amine nitrogen. The Cα and amide N use the "Z-then-X" convention, where a first neighboring atom is selected to define the Z-axis, and a second neighbor defines the ZX-plane and the positive x direction. The carboxylate example uses the "Bisector" convention, where the bisector of two neighboring atoms defines Z-axis. This is mainly used in structures with 2-fold symmetry. The amine N is represented by the "Z-Bisector" convention, where the N-R bond defines Z, and the bisector between the two N-H bonds defines X. In all cases, the Y-axis is defined according to the right hand rule.

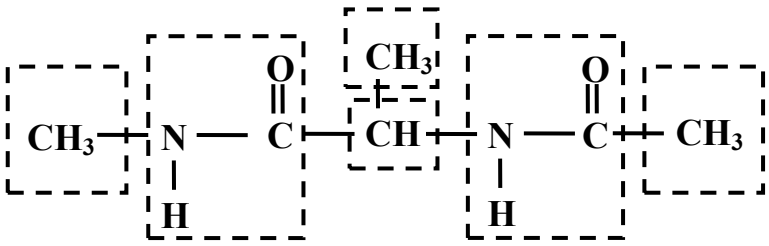

Figure 3.2: Illustration of the intramolecular polarization group definition. Each group consists of a functional group with limited conformational flexibility. The permanent multipole on each atom only polarizes atoms of other groups while induced dipoles on all atoms polarize all other atoms no matter what groups they are in.

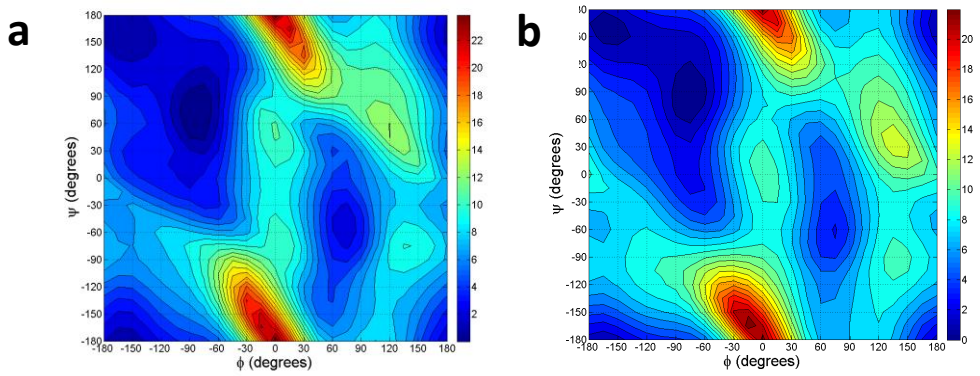

Figure 3.3: Gas-phase energy contours for alanine dipeptide from RI-TRIM MP2/CBS (a) and AMOEBA (b). The energy was computed on a 24 x 24 grid.

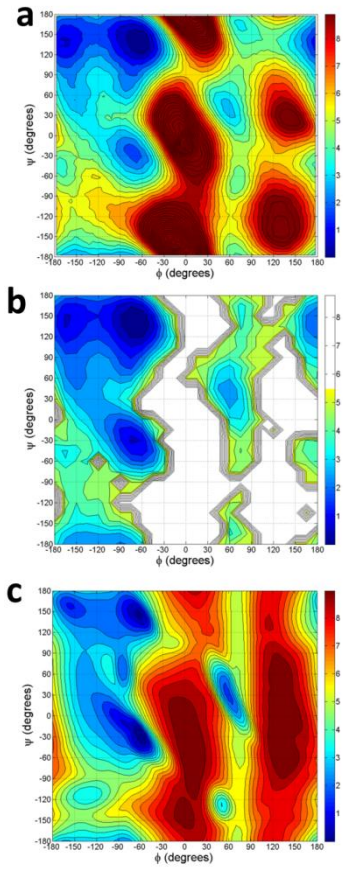

Figure 3.4: Comparison of Ramachandran potential of mean force for alanine. (**a**) Ala-2 residue of $(Ala)_3$ as predicted by 2D-WHAM simulations. (**b**) Average of ala-2, ala-3, and ala-4 residues in replica exchange molecular dynamics simulation of the $(Ala)_5$ peptide. The trajectory at 298 K was used. (**c**) The PDB data.

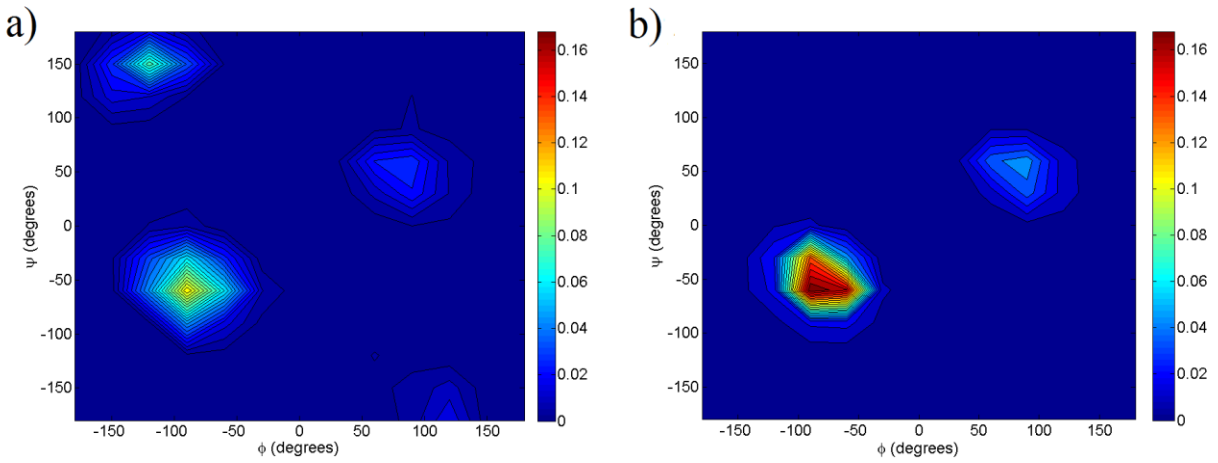

Figure 3.5: Comparison of (a) isoleucine and (b) serine conformational energy about χ^1 angle. The solid lines are RI-TRIM MP2/CBS energy while the dashed lines are AMOEBA values. The AMOEBA curve is shifted to minimize the overall RMS difference between AMOEBA and QM. The top set of curves (with higher energy at 0 degree) corresponds to a backbone conformation of (-60.0, -45.0), and the other corresponds to (-140.0, 135.0).

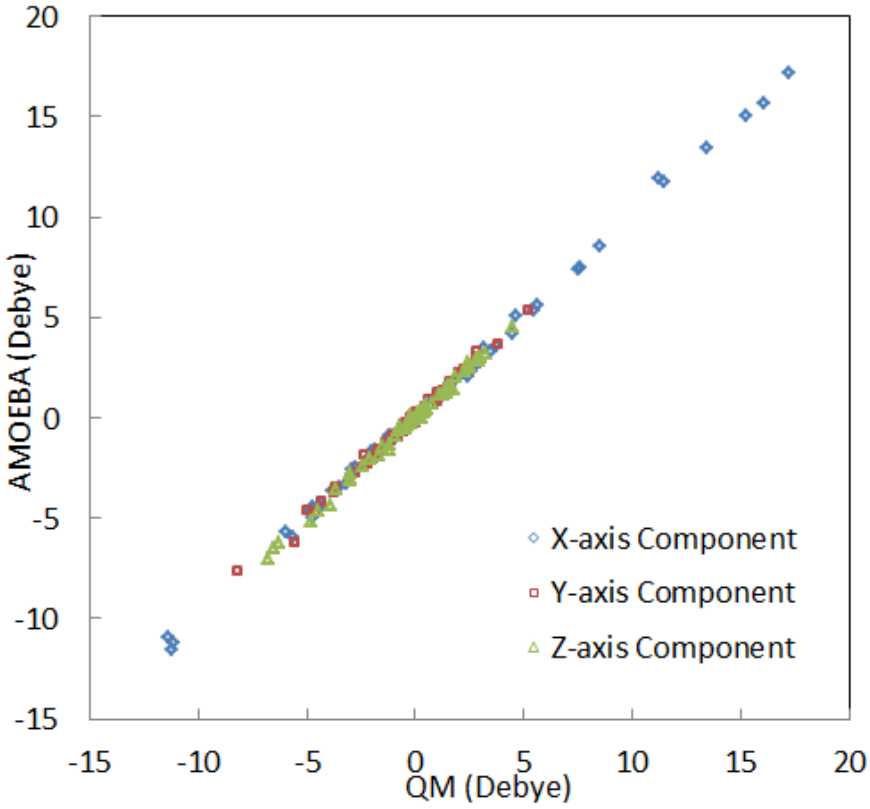

Figure 3.6: Comparison of amino acid molecular dipole moments predicted by AMEOBA and QM (MP2/aug-cc-pVTZ). The AMEOBA permanent atomic multipoles were derived from a set of dipeptides and validated on additional conformations (3 for each amino acid). Only the results for the validation sets are shown. The actual data can be found in the supporting information.

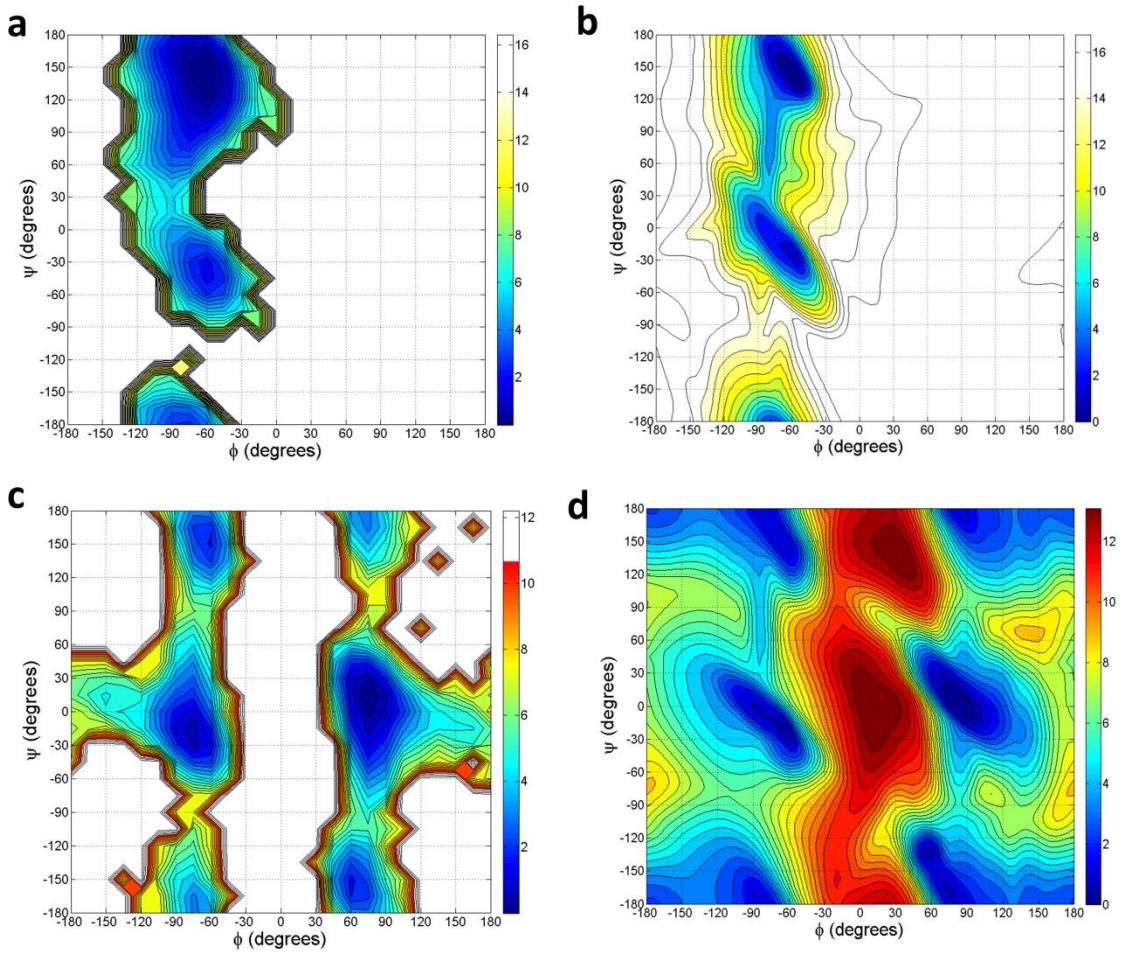

Figure 3.7: Comparison of Ramachandran potential of mean force maps for proline and glycine. (**a**) Pro-2 residue of GPGG from AMOEBA simulations. (**b**) The PDB data for proline. (**c**) Gly-3 residue of GPGG from simulations. (**d**) PDB data for glycine. All the PDB PMF were computed using data from Dunbrack et al.

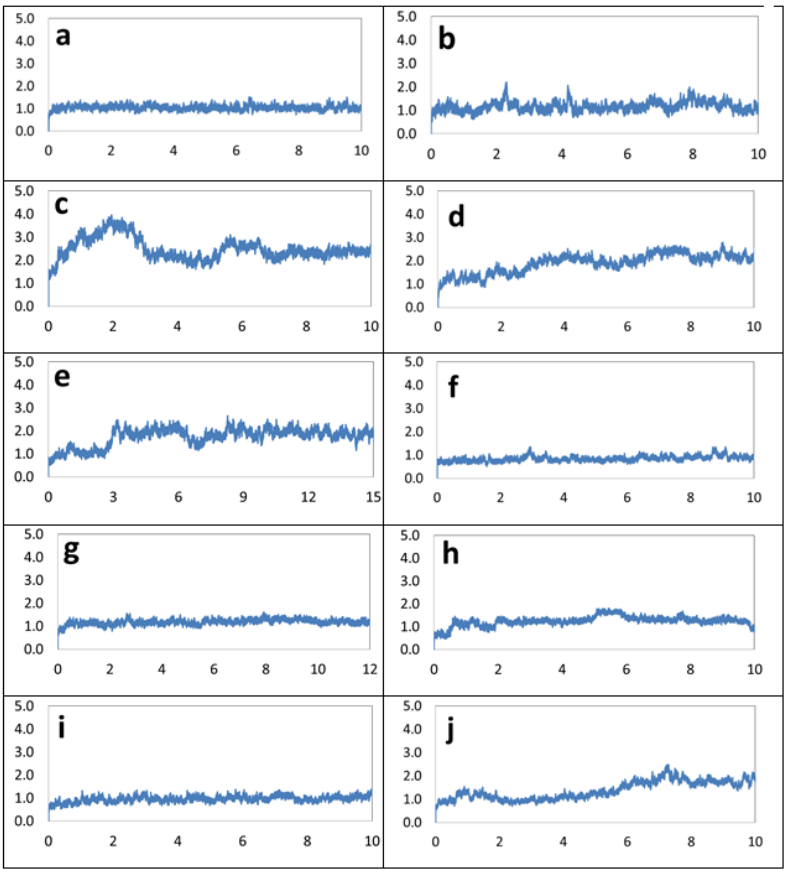

Figure 3.8: The time evolution of backbone RMSDs from the X-ray structures for ten simulated proteins. For each protein, 10 ns simulations were performed with AMOEBA force field in explicit water. The X-axis represents time (ns) and the Y-axis is the RMSD values in Å. (a) Crambin (PDB:1EJG), (b) TRP Cage (PDB:1L2Y), (c) Villin Headpiece (PDB:1VII), (d) ubiquitin (PDB:1UBQ), (e) GB3 Domain (PDB:2OED), (f) RD1 Antifreeze Protein (PDB:1UCS), (g) SUMO-2 Domain (PDB:1WM3), (h) BPTI (PDB:1BPI), (i) FK Binding Protein (PDB:2PPN), (j) and lysozyme (PDB:6LYT).

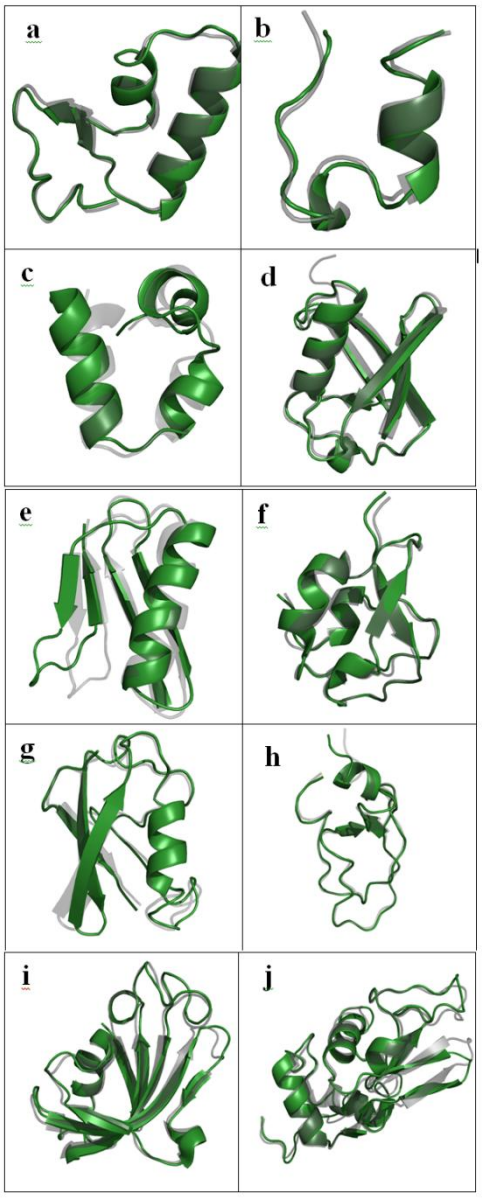

Figure 3.9: Superimposition of the final structures from AMOEBA simulations and the experimental X-ray crystal structures. (**a**) Crambin (PDB:1EJG), (**b**) TRP Cage (PDB:1L2Y), (**c**) Villin Headpiece (PDB:1VII), (**d**) ubiquitin (PDB:1UBQ), (**e**) GB3 Domain (PDB:2OED), (**f**) RD1 Antifreeze Protein (PDB:1UCS), (**g**) SUMO-2 Domain (PDB:1WM3), (**h**) BPTI (PDB:1BPI), (**i**) FK Binding Protein (PDB:2PPN), (**j**) and lysozyme(PDB:6LYT).

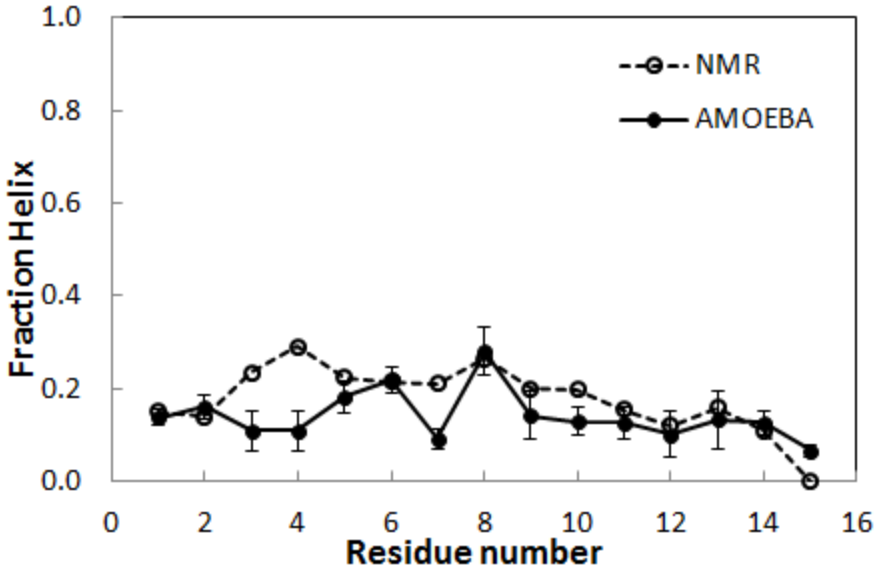

Figure 3.10: Fraction of helix $\langle h_i \rangle$ for each residue in Ac-$(AAQAA)_3$-NH_2 from replica exchange MD simulations and NMR chemical shifts at 303 K.

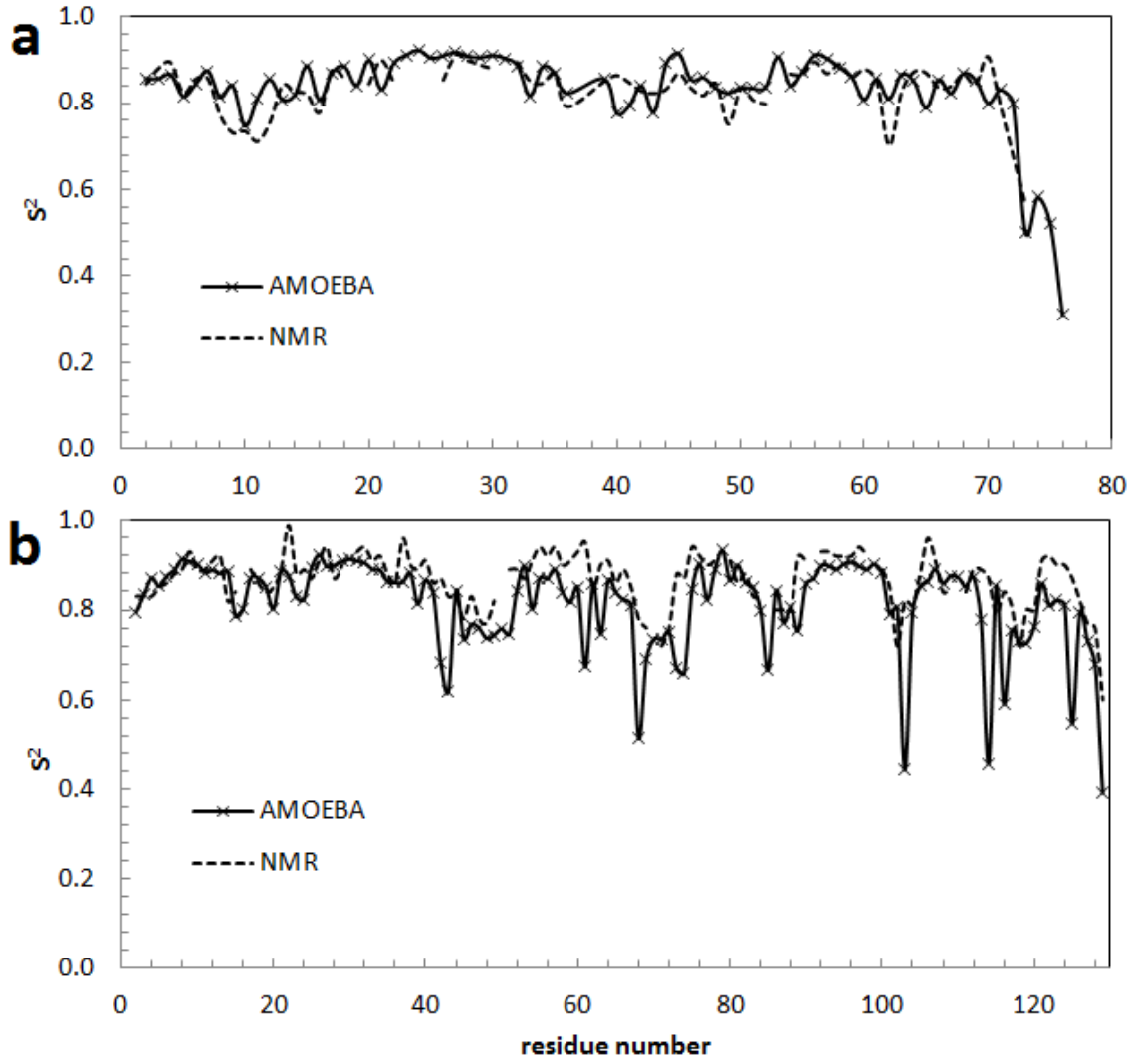

Figure 3.11: Order parameters (S2) derived from experimental NMR[261, 262] (dash lines) and calculated from MD simulations in explicit water using AMOEBA. (a) Ubiquitin, (b) Lysozyme.

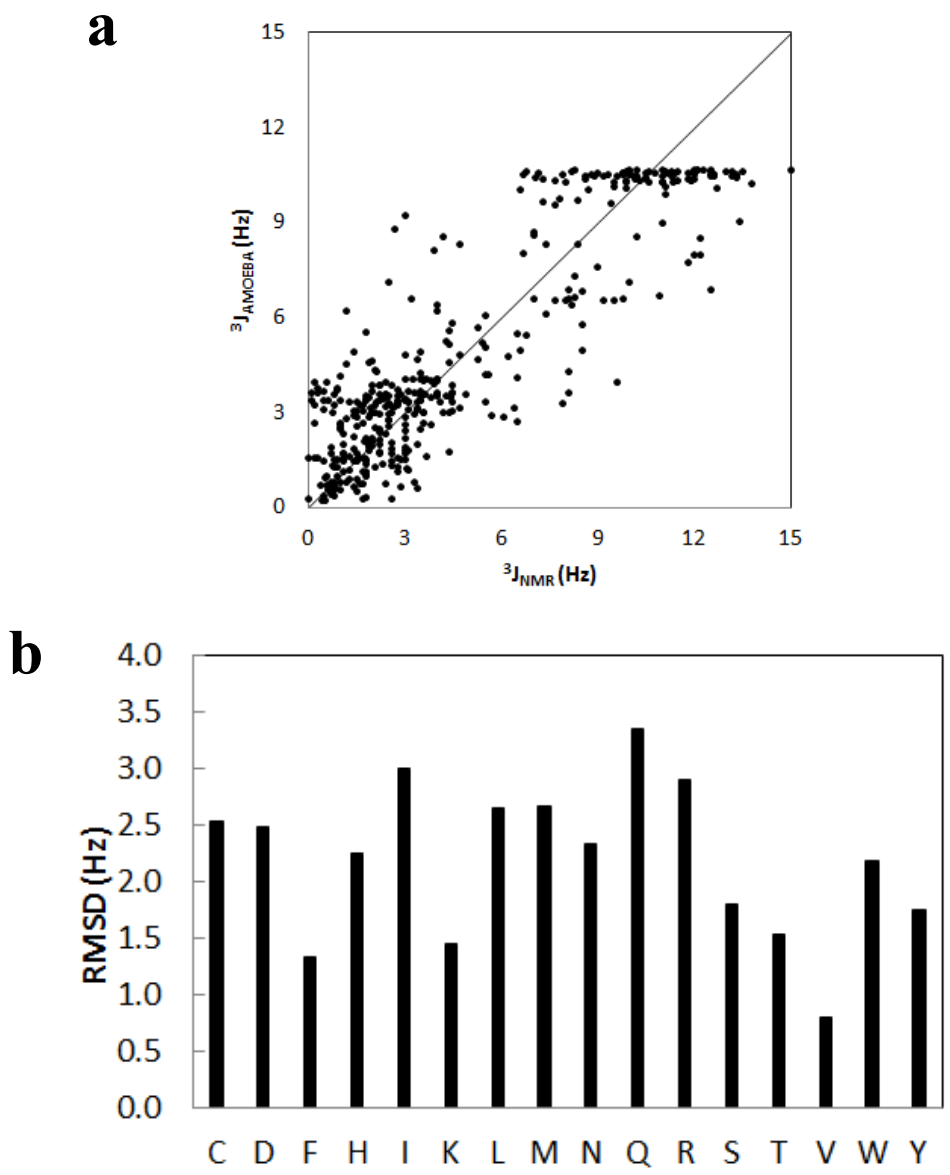

Figure 3.12: (a) Correlation of the experimental NMR *J*-couplings and the calculated *J*-coupling values from the MD simulations of BPTI, GB3, Ubiquitin and Lysozyme. (b) The RMSDs between the experimental and AMOEBA calculated *J*-coupling constants for each residue.

4 Trypsin-Ligand Binding Free Energies Calculation with AMOEBA

4.1 INTRODUCTION

The discovery of a lead molecule that binds to a targeted protein with high affinity is a major preoccupation of early-stage drug design [1, 272]. Accurate calculation of binding free energies is a must in this process. Treatments of protein-ligand binding, ranging from simple empirical scoring functions to thermodynamic free energy simulations with explicit solvent and full atomic details are widely used [273]. In principle, free energy perturbation (FEP) provides formally rigorous means to compute free-energy changes [274]. Although there have been numerous successful applications [3, 161, 273], calculating biomolecule-ligand affinities remains challenging for the highly polarized or charged system. Both the potential energy functions and sampling efficiency need improvement. In a previous work, we reported the absolute and relative binding free energies of charged ligands to trypsin[58]. All the calculated binding free energies are well within the accuracy of experimental measurement. In this chapter, we study the relative binding free energies using a polarizable potential via explicit solvent molecular dynamics simulations. The free energies were decomposed into electrostatic and vdW components to examine the importance of different energy contributions. Also we investigated the relationship between the dipole moment, polarizability and binding free energy of the ligand.

4.2 METHODS

4.2.1 Atomistic Model

The benzamidine-trypsin crystal structure (1BTY)[275] was used to generate new structures for the other ligands. Relative free energy changes of five ligands to

benzamidine were investigated. Ligand B and C replace the phenyl ring of benzamidine with a 1,3-diazine and 1,4-diazine respectively. Ligand D includes an amino group at 4-position of the phenyl ring. Ligand E is the only ligand in this study with an amine group instead of amidinium group. Ligand F is a derivative of ligand D with two carbon atoms in the ring substituted by nitrogen atoms. Ligand B and C were mutated from benzamidine in the trypsin binding pocket with the crystal complex structure whereas ligand D to F were superimposed on the benzamidine in the pocket and new structures were saved. For each ligand, we soaked the protein in an octahedron box with 4515 water molecules and 58 Å on each side.

4.2.2 Force Field and Parameterization

The potential function for the entire system, including trypsin, ligand and water, is given by

$$E = E_{ele} + E_{vdW} + E_{bond} + E_{angle} + E_{torsion} + E_{oop} \tag{4.1}$$

In AMOEBA force field, the electrostatic interaction composes of permanent atomic charges, dipoles, quadrupoles and the polarization effect by atomic induced dipole [19, 22, 276]. The van der Waals interaction is described by a buffered-14-7 function [217]. The electrostatic parameters are derived from quantum mechanical calculation. Each ligand was first optimized with Gaussian03 package at the level of HF/6-31G* [179]. Then the single point calculation was run at MP2/6-311++G(2d,2p) and multipoles of the ligands were calculated with GDMA v2 [51]. The van der Waals (vdW), bond, angle, and atomic polarizability parameters of the ligands were transferred from AMOEBA potential (amoebapro.prm) in TINKER package[277].

4.2.3 Alchemical Transformation

Alchemical Transformation was used to compute the relative binding free energies between the different ligands. One ligand was perturbed from another ligand in both bulk water and the protein complex. The relative binding free energy between these two ligands can be computed as:

$$\Delta\Delta A_{bind}(L1 \to L2) = \Delta A_{pro}(L1 \to L2) - \Delta A_{wat}(L1 \to L2) \quad (4.2)$$

The free energy simulations were performed by changing electrostatic and van der Waals parameters between the ligands in steps. When it comes to the annihilation of atoms, the a soft-core buffered-14-7 vdW function used between the dummy atoms and all other atoms in the system.[278] The free energies between two neighboring steps were calculated using the Bennett Acceptance Ratio estimator [279].

MD simulations were performed in parallel for all steps using PMEMD in AMBER v9. NVT dynamics simulations for 1 ns were run at each step, with a 1 fs time step, and 9 Å vdW cutoff. Particle Mesh Ewald (PME) was used to treat the electrostatic interactions, with a real-space cutoff of 7.0 Å. We used the Bennett acceptance ratio of 10^{-5} D per atom as the convergence criterion.

4.3 RESULTS AND DISCUSSION

4.3.1 Relative Binding Free Energies

Unlike the absolute binding free energy, relative binding free energy is more likely to be predicted accurately due to the small structural change and the systematic error cancellation. Ligands B through E were perturbed from ligand A whose absolute binding affinity was obtained in our previous work and ligand D was then transformed into ligand F. The calculated absolute and relative binding free energies are in excellent

agreement with experimental measurements (**Figure 4.1**). The experimental binding free energies are based on inhibition constants determined by spectrophotometry or isothermal titration calorimetry under various assay conditions [280-284]. The existence of multiple experimental values for single ligand indicates that the experimental uncertainty is almost 1 kcal/mol in energy or one order of magnitude in binding affinity.

4.3.2 Electrostatic Interaction as Driving Force for Binding

Although the separation of electrostatics and vdW contribution to the binding free energy is somewhat artificial because their values may vary in different perturbation path, the decomposition of the free energy change may provide valuable illustrations of the driving force of the binding of the ligands. **Figure 4.2** shows the decomposition of the binding affinity for the 6 ligands we calculated. Deng et al. [285] 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, the benzamdine ligands carry net charges and form a salt bridge with the trypsin. For these systems, the electrostatic contributions range from -4.95 to -7.97 kcal/mol, while the contributions from other interactions are only from -0.50 to 2.60 kcal/mol. Thus the electrostatic interaction is indicated as the driving force of the binding of these highly charged ligands to trypsin.

4.3.3 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. In our previous work, we computed the "polarization free-energy" in both bulk water and trypsin by turning off polarization

between the ligand and trypsin. The results showed that polarization works to diminish the effect of permanent electrostatics in driving the binding of ligand to trypsin. It is not surprising as the benzamidine (+) cancels the polarization effect of Asp 171(-) in the binding pocket while forming strong electrostatic attraction. In **Figure 4.3**, we showed the molecular dipole moments and polarizability of each ligand and their correlation with binding free energy.

Essex et al. [286] and Talhout et al. [283] suggested 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. However, the scattering plot of binding affinities and ligand dipole moments in **Figure 4.3** does not imply any of such correlation, with a poor R square value of 0.026. Ligand B has the smallest dipole moment among the six ligands, yet its solvation free energy is the largest. The significant free energy change in bulk water (-25.51 kcal/mol) is compensated by that in complex (-23.76 kcal/mol) so that the binding affinity is no stronger than some other ligands. At the same time, ligand E, which bears the largest dipole moment (-10.80 Debye) only has a binding free energy of -5.0 kcal/mol. The electrostatic details beyond the molecular dipole moment play the important role.

Interestingly, **Figure 4.3** shows a reasonable correlation between binding free energy and polarizability, indicating that the stronger the polarizability, the weaker the binding affinity will be. Ligand B, which has the smallest dipole moment, takes a polarizability (10.33 Å3) close to ligand C (10.37 Å3) and has a similar value of binding affinity with ligand C (-4.97 kcal/mol versus -4.87 kcal/mol). According to the study by Brian W. Matthews et al. [287], the strength of attraction is directly proportional to both

the polarizability and ionization potential of the interacting molecules. Considering the limited range of values for the ionization potential, the attractive force mainly depends on the polarizability. The outcome of our calculations supports this view point. The R square value of linear fit is 0.85, RMSE 0.52, indicating a good linearity between binding affinity and polarizability.

4.3.4 Structural Analysis

There are a number of hydrogen bonds between ligands and trypsin, including the amidinium group with Asp 171, Gly 196, Ser 172 and water molecules and the amino group with Ser177. In the crystal structure (1TBY), Asp 171 forms double hydrogen bonding with the two nitrogen atoms of benzamidine. However, it is not always the case as shown in the simulations. Take ligand A, C, D and E for example, only one hydrogen bond between Asp 171 and amidinium group was consistently observed in the simulations. This is due to the competitive interaction from a water molecule in the binding pocket. Whenever double hydrogen bonding between the ligand and the protein is missing, there is always a water molecule nearby forming a hydrogen bond with the amidinium. Ligand B forms more stable salt bridge throughout the simulation, with the both N-O distances within 3.5 Å. As for the internal water molecule seen in the crystal structure, our simulation demonstrated the existence of this crystal water was no accident. It interacts with one of the nitrogen atoms of the ligands constantly except for ligand E which has an amine group.

It is worth noting that 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.

4.5 CONCLUSIONS

In this chapter, the binding affinities of five positively charged benzamidine analogs to trypsin were calculated with polarizable AMOEBA force field. The relative binding free energies were computed by mutating each ligand to benzamidine in both water and protein from MD simulations. The calculated binding free energies are well within the experimental uncertainty. Our results also indicate that electrostatic interaction is the dominant force of the binding of all the ligands. Although the correlation between dipole moments and binding free energies as other group has argued were completely invisible, there is a negative correlation between the polarizability and binding free energy. The structures of the binding complexes and hydrogen bonding dynamics were also examined carefully from molecular dynamics simulations. The presence of water seems play an important role.

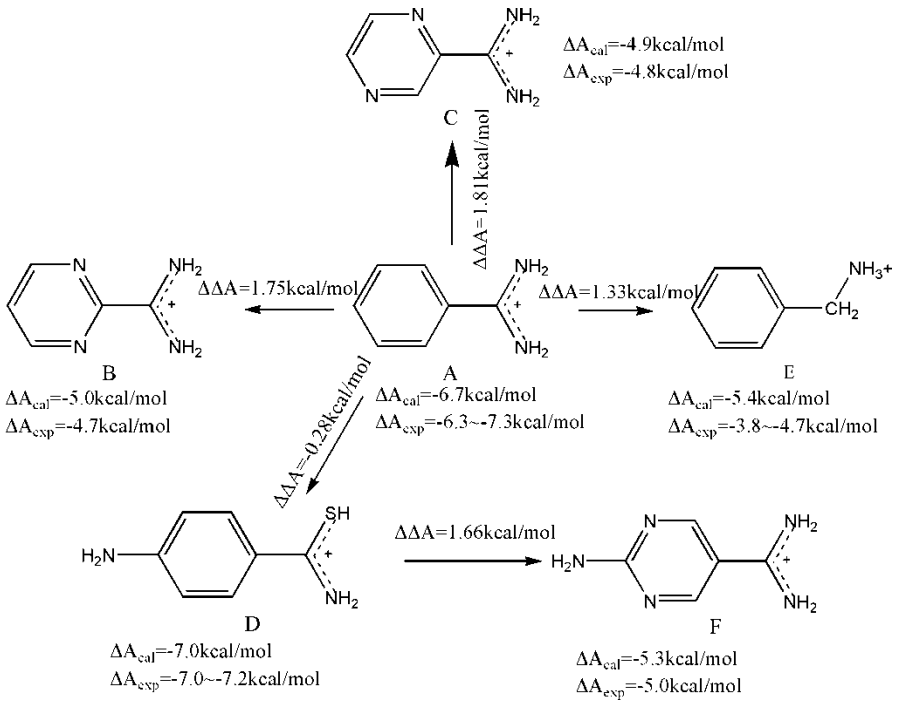

Figure 4.1: Relative binding free energies between ligands.

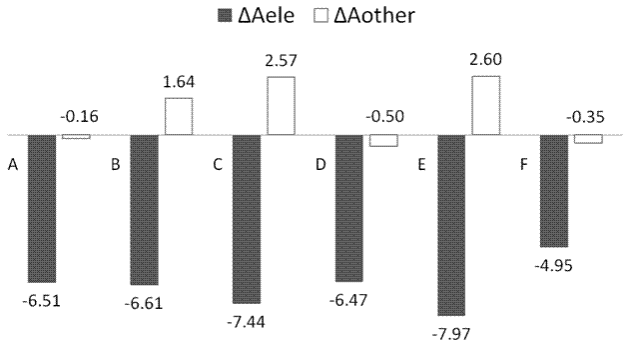

Figure 4.2: Decomposition of binding free energies (kcal/mol). Grey column is the electrostatic free energy and white column is the contribution of other free energy components including vdW and geometry.

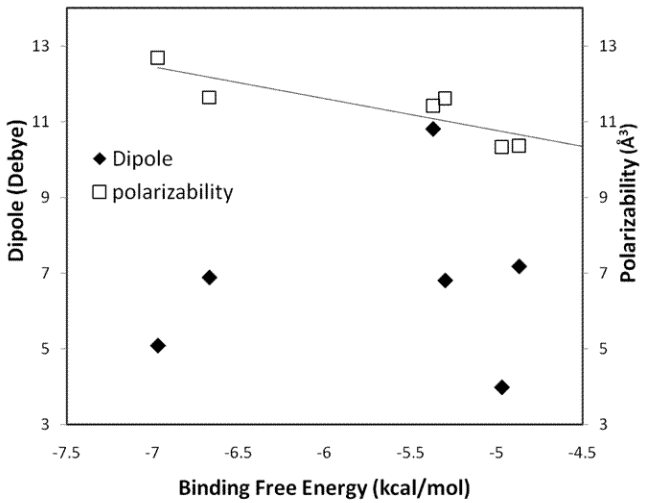

Figure 4.3: Correlation between dipole/polarizability of the ligands and binding free energy. Molecular dipole moments are in black diamond while polarizabilities are in open squares.

5 Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain using Polarizable Force Field simulations

5.1 INTRODUCTION

Understanding the effects of making chemical modifications to ligands upon their relative binding energetics is a critical step in structure-based drug design. Preorganizing a flexible ligand in the conformation it adopts upon binding has long been considered a useful strategy to achieve more a favorable binding entropy and thus an improved binding affinity.[288] Indeed, the program CAVEAT was developed in part to facilitate the design of constrained molecules bearing substituents directed in predefined orientations.[289] Awareness of the potential energetic benefits of ligand preorganization dates back to work by Jencks in the 1970s,[290] and there are numerous reports of the increased affinities that may accompany the introduction of conformational constraints into flexible ligands.[290-296] However, it has recently been reported that the binding entropy of a constrained ligand may actually be *less* favorable than its flexible control.[297-301] Even if ligand preorganization leads to a beneficial entropic contribution to binding, the enhancement to binding affinity may be offset by a compensating enthalpic penalty.[294, 302-310] Entropy-enthalpy compensation has been widely studied, but its origin is not well understood.[292, 311-316] While some consider this effect as an intrinsic physical phenomenon, others argue that the entropy-enthalpy compensation is a statistical artifact arising from obtaining entropy and enthalpy based on temperature dependent data from both experiment and theoretical calculations.[311, 312, 314, 317-320]

In order to understand the detailed energetic effects of ligand preorganization on protein-ligand interactions, it is essential to perform systematic experimental studies to determine the contributions of entropy and enthalpy to binding free energy and to correlate these with structural and dynamic analyses of the protein-ligand complexes using X-ray crystallographic and NMR spectroscopic methods.[321] For example, DeLorbe *et al*.[301] recently examined the binding energetics and structures associated with complexes of a series of constrained and flexible phosphotyrosine-derived peptide analogs with the SH2 domain of growth receptor binding protein 2 (Grb2); Grb2is a 25 KDa cytosolic adapter protein that is involved in activation of the Ras signal transduction pathway.[322] The constrained ligands cpYVN and cpYIN, which were preorganized by incorporating a cyclopropane ring at the pY replacement in the pseudopeptides fpYVN and fpYIN, respectively, (**Figure 5.1**), have the same functional groups and the same number and type of heavy atoms as their flexible controls. The thermodynamic binding parameters of these ligands for the Grb2 SH2 domain were determined by isothermal titration calorimetry (ITC),[323]and the binding entropies for the constrained ligands were found to be *less* favorable than for their flexible analogs. This unexpected finding is contrary to the conventional wisdom that ligand preorganization should be accompanied by a more favorable binding entropy. Their less favorable binding entropies notwithstanding, the constrained ligands bound with higher affinities than their more flexible counterparts because of significantly more favorable binding enthalpies. That the measured heat capacity changes on binding of the flexible/constrained ligand pairs were similar suggests that these thermodynamic differences do not arise from desolvation effects.[324] Although these experimental studies clearly show that ligand

preorganization does not necessarily result in a more favorable binding entropy, the molecular origin of the unanticipated behavior was not examined.

Over the past decades, numerous efforts have been devoted to develop and apply computational approaches to screen and design potent ligands for drug discovery.[272, 273, 325] Computational methods such as docking and molecular dynamics utilizing continuum and explicit solvent models have long been employed toward predicting protein-ligand binding affinities *in silico*.[2, 3, 272, 325] Among the various strategies that have been explored, detailed alchemical pathway simulations using explicit solvent show significant promise for providing energetically accurate predictions of protein-ligand binding affinity.[1, 321, 326]The*absolute* binding free energies calculated from such pathways correlate reasonably well with experimental data, and root mean square (RMS) errors of less than 3 kcal/mol are often reported.[2] On the other hand, the *relative* binding free energies can be calculated more accurately, if there is sufficient sampling of protein-ligand-water configurational space.[327] Decomposition of binding free energy into entropic and enthalpic contributions offers important insights into the driving forces for protein-ligand recognition;[321] however, quantitative estimation of the binding entropy remains a significant challenge. Common approaches to estimating the binding entropy include quasi harmonic analysis,[328, 329] normal mode analysis,[330-332] and knowledge based scoring functions.[333-335] These methods have been applied to several protein-ligand systems,[336-342] but the contributions from solvation are typically neglected or approximated due to the computational expense. More physically rigorous alchemical pathway approaches are also applicable to evaluating entropy, but few investigations have reported using such methods in protein-ligand interactions.[343]

In this chapter, we use molecular modeling to explore how introducing conformational constraints into two closely related phosphotyrosine-derived peptides affects the thermodynamic parameters for their complexation with the Grb2 SH2 domain. Because the phosphate group in these ligands is charged, the AMOEBA polarizable force field,[19, 22, 39, 344] which accounts for multipole electrostatics and polarization effects, was used to model protein, ligand, and water components in each MD simulation. The AMOEBA force field has been successfully applied to accurately model a number of highly polar molecular systems, including water,[19] monovalent and divalent ions,[52, 53, 345, 346] small molecules,[347] peptides,[348] and trypsin-benzamidinium binding.[58, 62, 63] We report herein the calculations of binding free energies, enthalpies and entropies, as well as the results from simulations of the structure and dynamics of the ligands in water and in their complexes with the Grb2 SH2 domain. The results are compared with experimental observations, and a possible molecular origin for the unexpected, unfavorable entropic effect resulting from preorganizing these phosphotyrosine-derived ligands is presented.

5.2 METHOD

5.2.1 Entropy Calculation

The change in entropy (ΔS) is related to the change in free energy by[349]

$$\Delta S = -\left(\frac{\partial \Delta A}{\partial T}\right)_{N,V} \tag{5.1}$$

In this study, we computed the entropy change numerically from the slope of a linear fit to the temperature dependence of free energy change near room temperature. This approach is essentially identical to the finite difference method that has been used

previously.[350-352] While other direct or perturbation methods exist for computing enthalpy of enthalpy, we have found that the numerical approach is the most stable for energetics calculations. In addition, as we are utilizing REMD to simulate free ligands in solution, it is natural to take advantage of the temperature dependence of free energy data.

The relative change in enthalpy is then computed as

$$\Delta H = \Delta A + T\Delta S \tag{5.2}$$

Bennett Acceptance Ratio (BAR) method[180] was utilized to calculate the relative binding free energy (ΔA). In order to avoid the end point singularity during alchemical transformation, the soft-core buffered 14-7 potential was used for van der Waals (vdW) interactions.

5.2.2 Quasiharmonic Analysis

Quasiharmonic analysis was performed to characterize the collective motions of molecules at thermodynamic equilibrium.[328, 329] The quasiharmonic approximation assumes that the spatial fluctuations in the system follow a multivariate Gaussian distribution. In a quasiharmonic frequency analysis on an n-atom system, the eigenvalues λ_i (i=1,2,…,3n-6) of the mass weighted covariance matrix of atomic fluctuations are calculated to determine the quasiharmonic frequencies, $\omega = \sqrt{(RT/\lambda_i)}$, which are then used to estimate the conformational entropy, given by

$$S = R\sum_{i}^{3n-6} \frac{\hbar\omega_i / RT}{e^{\hbar\omega_i/RT} - 1} - \ln(1 - e^{-\hbar\omega_i/RT}) \tag{5.6}$$

where $\hbar$ is the planck constant, and R is the gas constant. The quasiharmonic method has been successfully applied to simple molecular systems with few energy minima.

However, this method will overestimate the configurational entropy for more complex systems having multiple occupied energy wells, particularly when using Cartesian coordinates rather than internal coordinates.

5.2.3 Force Field and Parameterization

AMOEBA polarizable force field [19, 39] is applied to model the protein, water molecules, and most part of the peptide-analog ligands fpYVN, fpYIN, cpYVN and cpYIN (in **Figure 5.1**). Additional parameterization was necessary for the phosphotyrosine (pY) residue in the ligands. Missing valence parameters for the constrained and unconstrained pY segments were derived from Quantum Mechanics (QM) calculations by using the "Valence" module in TINKER software package.[220] "Valence" sets the equilibrium bond lengths and angles based on the HF/6-31G* optimized structures of the pY segments; the force constants and vdW parameters were transferred from existing parameters of the same atom types in AMOEBA force field. Trimethyl and dimethyl phosphates (TMP and DMP) were used to derive the vdW parameters of the phosphate group, by fitting to both QM structure and energy of TMP/DMP-water dimers. The vdW parameters were fine-tuned to match experimental liquid density and heat of vaporization of TMP as well. The electrostatic parameters for the pY side chain were obtained from QM at the MP2/aug-cc-pVTZ level by the "original-fit" approach.[347, 353] In this approach, the atomic multipoles for the pY segment were initially derived from MP2/6-311g** density matrix using the original distributed multipole analysis (DMA).[169, 354] The dipole and quadrupole moments of the pY residue were then optimized to the electrostatic potential (ESP) around the whole ligand computed at the MP2/6-311++G(2d,2p) level (electrostatic parameters of other

residues were fixed). The ESP root mean square differences between QM and final atomic multipoles, evaluated over roughly 35000 grid points around the peptide, are 1.55 kcal/mol per unit charge and 0.92 kcal/mol per unit charge for cpYVN and fpYVN, respectively.

The model compounds of both unconstrained and constrained pY segments and the dihedral angles a~h for which parameters were derived are illustrated in **Figure 5.2**. Dihedral angle parameters in the pY subunit that were missing were obtained by comparing the QM conformational energy profile to the ones computed from corresponding MM (Molecular Mechanics) using all energy terms except the dihedral angle term. The difference in energy was then fit to a 3-term Fourier series torsional function. The "torsional energy" in MM works as an "error" function. In order to achieve better transferability, its contribution to overall conformational energy should be minimized. Typically, the torsional energy parameters (V1, V2, and V3) are less than 1-3 kcal/mol for rotation about a single bond (compared to 15-20 kcal/mol for double bonds), whereas the overall conformational energy barriers are on the order of tens of kcal/mol.

5.2.4 Computational Details

In order to evaluate the thermodynamic driving forces for ligand binding to the SH2 domain, we calculated the relative binding free energy, enthalpy and entropy contributions for constrained and unconstrained ligand pairs, according to the alchemical pathway shown in **Figure 5.1**. The relative binding free energies for each pair of constrained and unconstrained ligands were computed from the free energy differences between the ligands in water and in the protein binding pocket. The fpYVN ligand was gradually transformed into cpYVN by performing 26 steps of simulations; 10 steps were

performed in the alchemical transformations of Val to Ile in the constrained and unconstrained ligands. For the ligand-water systems, replica exchange molecular dynamics (REMD) [233, 355, 356] simulations were performed with 48 replicas at temperatures between 260 and 620 K (detailed schedule can be found in the Supporting Information).[232] For perturbations from fpYVN to cpYVN, a bond was grown to form the cyclopropane by gradually increasing the force constant (from 0 to 550 kcal/mol/ $Å^2$); at the same time, two hydrogen atoms were turned into "dummy atoms" by turning off their vdW and electrostatic interactions with other atoms. Note that the valence contributions due to the dummy atoms are canceled between perturbations of ligand in water and in solvated complex. NVT simulations of 2.5 ns were performed at each step. Replicas were exchanged every 2 ps, and the exchange success rates for all replicas were greater than 20%. All the REMD simulations were performed using the parallel SANDER module in AMBER10. We modified the REMD implemented in AMBER10[221] for use with the AMOEBA force field. For the protein-ligand complexes in explicit water, relative free energies for constrained and unconstrained ligands having Val and Ile at the pY+1 position were calculated at 288 K, 298 K and 308 K. For each set of alchemical calculations at each temperature, a total of 65 ns NVT simulations were performed over 26 steps, using the PMEMD module in AMBER10. For all the simulations, the vdW cutoff was set to 12 Å, and the long-range electrostatics were treated using Particle Mesh Ewald (PME) summation[182-184] with a grid space of 0.8 Å and a real space cutoff of 7 Å. The induced dipoles, also computed with PME, were iterated until the root mean square change was below 0.01 D/atom. A tighter induced dipole convergence of 10^{-5} D/atom was used in the energy calculation for the post free

energy analysis with Bennett Acceptance Ratio (BAR). By using a bootstrap procedure, the statistical uncertainty was estimated as the standard deviation of the average free energy values that are computed using100 partial simulation trajectory blocks (1.0 ns).Given that the uncertainty in the entropy calculation was dominated by the free energies at the lowest and highest temperatures (288K and 308K), the statistical error for $-T\Delta S$ was estimated from the upper and lower bound of free energy changes at 288K and 308K; thus the statistical error for$-T\Delta S$ is twice the statistical error of free energy.

The entropy was extracted from the temperature dependency of free energies via linear regression (Eq. 5.1). For ligands in water, the relative free energy ΔA_{wat} was obtained from REMD at 18 temperatures between 260 K and 360 K. A linear fit was used to interpolate ΔA_{wat} at 288K, 298 K and 308 K. These values were then subtracted from the free energy changes of ligand in complex (ΔA_{comp}) at the same temperatures to obtain relative binding free energy $\Delta\Delta A_{bind}$. The entropic contribution was computed from the slope of the fitted linear temperature dependence of relative binding free energy. The enthalpy was evaluated via $\Delta\Delta A+T\Delta\Delta S$. The total simulation time for the combined MD simulations of free ligands and complexes at all temperature is up to 6 μs.

5.3 Result And Discussions

5.3.1 Ligand Conformational Property and Sampling

An accurate description of the conformational distribution of the peptide-like ligands is essential in this study. The ligands of interest possess a total of more than a dozen of rotatable bonds, with four being in the constrained pY residue (cpY) and six in the unconstrained pY residue (fpY). The large number of degrees of freedom suggests the ligand can be highly flexible, thus presenting a significant challenge for molecular

simulations. On the other hand, the strong intramolecular interactions between the charged phosphate moiety and the other polar groups in the gas phase result in certain stable conformations with low potential energy. The conformational energy profile for the pY residue indicates the energy barrier for escaping these stable conformations can be tens of kcal/mol (**Figure 5.2**). In solution, it is likely that the conformational population will be significantly different due to the competition by water molecules. As a matter of fact, it is well known that alanine dipeptide has distinctly different local minimum energy structures in the gas and solution phases.[344] As AMOEBA is a polarizable model that responds to electrostatic environment via changes in the induced dipoles, we believe that it is important to reproduce the gas-phase conformational properties as well as to compare the computational and experimental data in the liquid phase as discussed below. We have examined each of the main conformational degrees of freedom for the pY residue that we have parameterized. **Figure 5.2** shows that the AMOEBA conformational energy profiles are in good agreement with high-level *ab initio* QM results. The average root-mean-square derivation (RMSD) between QM and MM minimized structures is about 0.25 Å per atom.

Adequate sampling of configurational space of the molecular system, including the ligand, is critical in order to obtain reliable thermodynamic information from the simulations. When the peptide-like ligands are bound to SH2 domain, they are restricted within the protein binding pocket, and their structures are relatively well-defined. They are similar to the X-ray crystallographic structures.[301] The ensemble of structures of the free ligands in solvent is unknown, so we performed MD simulations of the unconstrained fpYVN ligand in a dielectric medium (ε=80) at room temperature and

above in order to explore the conformational space of the peptide analogs in water and to guide the more elaborated explicit-solvent simulations. The potential energy and torsional RMSD distributions obtained from multiple independent simulation trajectories were compared to examine the convergence of the MD sampling. We found that MD simulations at room temperature do not produce converged distributions after 30 ns. On the other hand, when these simulations are performed at 600K or above, the distributions quickly converge after a few nanoseconds. Detailed information about these distributions can be found in the supporting information. Based on the information obtained from the simple continuum simulations, replica-exchange molecular dynamics (REMD) simulations in explicit solvent at temperatures ranging from 260 K to 620 K were performed to compute the relative free energy/entropy between different ligands in the solvent environment. Using these simulations, we examined the distribution of the main torsions of fpYVN in solution at 298K. Overall, we observe a broad sampling in torsional space and a number of torsional transitions. The autocorrelation functions of these torsions exhibit fast decay of only a few ps to reach $1/e$, perhaps because of the artificially fast kinetics of REMD; some examples are included in the Supporting Information.

5.3.2 Calculated Binding Thermodynamics Consistent with Experimental Measurements

We evaluated the relative binding free energy and entropy for complex formation of the constrained ligands cpYVN and cpYIN and their corresponding unconstrained analogs fpYVN and fpYIN with the Grb2 SH2 domain. The free energy and entropy were computed from molecular dynamics simulations using the AMOEBA polarizable force field, and results are summarized and compared with the values that were determined

experimentally by ITC in **Table 5.1**.[301]In order to facilitate the comparison of the calculated and experimental values, we used the fpYVN ligand as the reference and set the calculated values for it to those determined experimentally. The calculated thermodynamic parameters for the other ligands were then obtained by difference according to the relative binding free energies, entropies and enthalpies derived from the simulations.

The order of the calculated binding free energies correspond tothose determined experimentally with fpYVN≈fpYIN<cpYIN<cpYVN. The calculated binding free energy difference between fpYVN and cpYVN was overestimated by about 1.2 kcal/mol compared to experiment, whereas the difference in free energies of the Val and Ile variants match experimental results reasonably well. The transformation of fpYVN into cpYVN requires that a covalent bond be created. A small change in the bond length corresponds to large energy fluctuations due to the stiff bond stretching and angle bending energy term.[357, 358] During the transformation of fpYVN into cpYVN in water or in their complexes with the Grb2 SH2 domain, 26 intermediate steps are necessary to obtain sufficient overlap between the configurational spaces associated with neighboring steps. As illustrated in **Figure 5.3**, the relative free energies converge after 1.5 ns simulations for each intermediate step at different temperatures, with a deviation of less than 0.1 kcal/mol. REMD simulations for the ligands in water at 48 different temperatures were performed to enhance the sampling near room temperature. Accordingly, the relative binding free energies for each pair of ligands are effectively the result of 3 μs of simulation time.

Additional analyses were performed in order to scrutinize the convergence and effectiveness of the free energy calculations. To inspect the thermodynamic cycle closure, we reevaluated the relative binding free energy associated with the alchemical transformation fpYVN to cpYVN at 298K after systematically skipping intermediate perturbation steps (skipping every 4^{th}, 5^{th} ... 13^{th}steps). The results indicate that the relative binding free energy converges within 0.3 kcal/mol after 22 to 26 perturbation steps. For neighboring steps during the perturbation, the histograms of potential energy differences from forward and backward perturbation have been obtained and verified to overlap with each other. In addition, statistical inefficiency calculations[359] show that ~22 ps and ~45 ps are needed for ligand sampling in water and in the complex to lose "memory" of their previous configurations, respectively. Further details of the thermodynamic cycle, energy overlap, and statistical inefficiency tests are found in the Supporting Information.

The relative binding free energies were then decomposed into enthalpy and entropy contributions. Perturbation methods, including finite difference, single state perturbation, β -perturbation, modified β -perturbation, and the perturbation and correction method, provide a physically rigorous evaluations of changes in entropy and enthalpy.[352, 360] However, compared with the free energy calculations, the entropy evaluated from computational methods has much greater error. With the exception of the finite difference approach, all of the above methods suffer from numerical underflow problems for systems with large energy fluctuations such as solvated protein-ligand complexes. Recently, Wyczalkowski *et al.*[361] calculated solvation entropy changes based on the analytical temperature derivative of the Bennett Acceptance Ratio and

Multistate Bennett Acceptance Ratio (mBAR) estimators. The BAR/mBAR approach to computing entropy is sensitive to energy fluctuations and thus requires long simulation times to achieve convergence. Using the restraint release (RR) approach, Warshel*et al.*[343] reported the binding entropy of three protein-ligand systems with encouraging agreement to experimental measurements. In the RR approach, multiple structures of the ligands must be selected before free energy perturbation and quasiharmonic calculations. Considering the potentially high flexibility of our peptide ligands, a large number of representative configurations are needed to account for the global minima in each perturbation. In the current study, the REMD simulations that were performed on the free ligands in water already provided free energy data at different temperatures. Therefore, we took advantage of the simulations at different temperatures and adopted the finite difference method that has been utilized to calculate entropy change for both small and relatively large systems.[342, 349, 362-364] We used a linear regression to fit the temperature dependence of the relative binding free energies at 288 K, 298 K and 308 K (see Computation Details). For all the systems, the R^2 values for the linear fit range from 0.81 to 0.97.

The calculated relative binding entropies and enthalpies for the two ligand pairs show the same trends as experiment. Both calculation and experiment indicate that the constrained ligands bind to the Grb2 SH2 domain with less favorable binding entropy and more favorable binding enthalpy than their more flexible controls (**Table 5.1**). Relative to the unconstrained ligands fpYVN and fpYIN, the increased binding affinity, or lower binding free energy, that is observed for the two constrained ligands cpYVN and cpYIN is attributed to an enthalpic advantage rather than an entropic one. Based upon the

simulations, preorganization results in an unfavorable binding entropy change(higher $-T\Delta S$)of 8 kcal/mol, which is offset by an enthalpic gain of about –10 kcal/mol. Hence, although constraining the ligands fpYVN and fpYIN increases their binding affinities, *both computational and experimental results show that this affinity enhancement does not arise from more favorable entropic factors as would normally be expected.* The calculated absolute binding entropy $-T\Delta S$ of the constrained cpYV(I)N ligands is unfavorable (positive), whereas slightly favorable (negative) values were obtained experimentally. This difference is a reflection of our overestimation of the magnitude of the relative binding entropy between constrained and unconstrained ligand pairs, even though the sign of the relative change was predicted correctly. The mutation of Val to Ile in both constrained and unconstrained ligands had an insignificant effect upon both the calculated and the experimental values for binding enthalpies and entropies. Moreover, both the simulated and experimental data seem to suggest that enthalpy/entropy compensation limits the enhancement to the binding affinity, as reflected by the linear relationship between entropy and enthalpy in **Figure 5.4**.

Binding is a process that involves ligand desolvation followed by formation of the protein-ligand complex, so the behavior of unbound ligands in water plays an important role in the overall process. In the current study, the relative binding free energy and entropy was evaluated as the difference between the free energy and entropy changes of the protein and the ligands in solvent and complex environments. The entropy decomposition given in **Table 5.2** suggests that the majority of the unfavorable binding entropy observed for constraining fpYVN and fpYIN arises from differences in the entropies of the free ligands in solvent rather than in their complexes.

5.3.3 Effect of Constraint on Organization of Unbound Ligands

The concept underlying ligand preorganization is that constraining the unbound ligand in the three-dimensional shape that corresponds to that of the bound ligand will lead to a more favorable entropic term for binding. Provided the constrained ligand interacts in the same way with the protein and the solvent as the more flexible parent (*i.e.,* $\Delta\Delta H° \sim 0$ kcal/mol), this entropic advantage would lead to an increased binding affinity. We thus analyzed the simulated structure and dynamics of constrained and unconstrained ligands to examine the effect of constraining fpYVN.

The structures of the ligands extracted from molecular dynamics trajectories were hierarchically clustered based on the all-atom RMSD of ligands using average-linkage algorithm[365] over 2 ns at 298 K. In an average-linkage algorithm, the distance between one cluster and another cluster is computed as the average distance from any member of one cluster to any member of the other cluster. Recall that REMD simulations with 48 replicas at various temperatures were performed to facilitate the sampling of ligand configurations at 298 K. This is equivalent to a total simulation time of 100 ns. By using 1.0 Å as the limit for the average distance to centroid, 22 and 20 clusters were obtained for unconstrained and constrained ligand trajectories, respectively. Interestingly, the most dominant configuration (34.7%) of fpYVN is a rather compact, macrocyclic-like structure, as shown in **Figure 5.5**, with a prominent intramolecular contact between the phosphate group and the amide groups of the pY+2 residue. On the other hand, the cyclopropane ring in the constrained ligands cpYVN and cpYIN prevents these functional groups from interacting with each other, so these residues interact with water and are oriented in opposite directions. Superimposition of the dominant structure of cpYVN with the structure extracted from the complex[301] reveals that the cyclopropane

ring does indeed preorganize the flexible pY replacement in its biologically active conformation when bound to the domain. Thus, even though the cyclopropane ring *locally* preorganized the pY replacement, the simulations suggest that the constrained ligands possess significant flexibility at the pY+1 and pY+2 positions.

The entropy of both constrained and unconstrained ligands was estimated from quasiharmonic analysis (**Table 5.3**). The absolute entropy of solvated cpYVN (557 cal mol^{-1} K^{-1}) is indeed about 25% greater than that of fpYVN (448 cal mol^{-1} K^{-1}), which corresponds to a difference of 30 kcal/mol in $T\Delta S$ at room temperature. The same trend is also observed for fpYIN and cpYIN ligands, which have essentially the same entropies as that of fpYVN and cpYVN, respectively. Vibration contributes about 80% of the total entropy, while translation and rotation account for the remainder. The entropy difference between the constrained and unconstrained ligands thus arises primarily from the vibrational components. We further decomposed the vibrational entropy contribution to each atom and found that the modes contributing the most to the entropy difference arise from the pY+2 residue. This finding is consistent with the structural analysis above. Since the rigid cyclopropane ring separates the pY+2 residue and phosphate group of the pY replacement in the constrained ligands, the pY+2 residue is well exposed in solvent and tends to be mobile. On the other hand, the unconstrained ligands lack this ring, so motion of the pY+2 residue is restrained by strong hydrogen bonding interactions with the phosphate moiety.

Water molecules also contribute to the entropy of the system and thus may affect binding thermodynamics. The average intermolecular hydrogen-bonds between water and the ligands were computed from the MD simulations at 298.79 K, and the results shown

in **Table 5.4** reveal noticeable differences between the constrained and unconstrained ligands. The constrained ligands are solvated by about two more water molecules than their unconstrained counterparts on average. The higher number of contacts with water molecules for constrained ligands is consistent with the observation that constrained ligands are more extended and better exposed to solvent (**Table 5.4**). The unconstrained ligands possess more intramolecular hydrogen bonds than their constrained counterparts (2.6 vs. 1.4), so fewer water molecules bind to the unconstrained ligands. Although the interactions with more water molecules could translate into a higher enthalpic penalty for binding and more favorable binding entropy arising from ligand desolvation, the total relative binding enthalpy or entropy comprises the sum of contributions from both ligand and water molecules. The simulations and quasiharmonic analysis have thus far suggested that the unconstrained ligands actually have lower entropy than the constrained analogs owing to the intramolecular hydrogen-bonding interactions between the phosphate group and the amide groups of the terminal pY+2 residue. We conclude that variation in the structures of the ligands in solution is the predominant source of differences in observed differences in binding entropies and enthalpies. Although the cyclopropane ring in cpYVN and cpYIN does orient those atoms that interact with the domain in a manner closely similar to that of fpYVN and fpYIN, this ring also obviates the opportunity of forming intramolecular interactions between the phosphate group and the pY+2 residue. This may account for the fact that the constrained ligands bind with the less favorable entropy, seen by both our alchemical free energy calculations and experiment.

5.3.4 Flexibility of Solvated Complexes with Unconstrained and Constrained Ligands

There is the possibility that introducing a conformational constraint into a flexible molecule might affect protein dynamics in the resultant complexes in a manner that could have entropic consequences. In order to probe this question, structures of the four solvated complexes were analyzed using the MD trajectories obtained from the free energy calculations. A comparison of the structures of the solvated complexes of the Grb2 SH2 domain with each of fpYVN, cpYVN, fpYIN and cpYIN derived from simulations and X-ray crystallography yielded all-atom RMSDs of 1.6 Å, 1.6 Å, 1.1 Å and 1.5 Å, respectively, suggesting that the structures obtained computationally are generally in good agreement with those determined by experiment. The RMSDs between the complexes of cpYVN and cpYIN and the complexes of fpYVN and fpYIN are both 1.4 Å. Representative structures of the complexes of fpYVN and cpYVN containing complexes from MD simulations are compared with crystal structures in **Figure 5.7**.

The calculated B-factors about the mean coordinates of the simulations over 2 ns for all Cα atoms are shown in **Figure 5.6**. The four complexes share very similar fluctuation modes, and the coordinates of most of the residues vary within only 1 Å, although residues in some of the loops fluctuate by as much as 2.1 Å. Structural variations in the BC loop (GluBC1-GluBC4) were observed upon comparing the crystal structures of complexes of the Grb2 SH2 domain with constrained and unconstrained ligands, although comparable variations in the BC loops were also found in coexisting complexes in the asymmetric unit.[298, 301, 366] In our B-factor calculations, the fluctuations (1.5 to 2.0 Å) in the BC loops of the four complexes are the highest among all other loops. This observation suggests that the variations of the BC loop are more

likely to result from the intrinsic flexibility of the loop rather than from differences in the binding modes of the constrained and unconstrained ligands. Based on the simulated B-factors in **Figure 5.6**, the B-factors for the protein backbone in the complexes of the constrained ligands are slightly greater than in the complexes of their unconstrained counterparts. Although the trends in the B-factors obtained computationally and experimentally are the same, the differences are more prominent in the experimental values.[301]

The entropies calculated from quasiharmonic analysis (**Table 5.3**) suggest that the Cα-atoms in the Grb2 SH2 domain complexed with each of the four different ligands have similar entropies of about 820 cal/mol/K. The entropies of the domain in the complexes of the constrained ligands cpYVN and cpYIN are slightly higher than in the complexes with their unconstrained controls (about 1%). The higher entropies of the domain in the complexes of the constrained ligands suggest a more favorable contribution to binding free energy, however, the opposite trend is seen by both experiment and calculation. Based on the analyses above, we are unable to find evidence that a change in dynamics of the SH2 domain is responsible for differences in the experimentally observed binding entropies for the various complexes.

The structures of fpYVN, cpYVN, fpYIN, and cpYIN in their respective complexes with the SH2 domain were also examined. The structures of the complexes were clustered based on RMSD of all atoms in the ligands from all MD snapshots using the same method described above for free ligands. For complexes containing fpYVN and cpYVN, there is one dominant conformation of the ligand that forms strong polar interactions with the protein, whereas there are two dominant conformers for fpYIN and

cpYIN because of the flexibility of the pY+1 (Ile) residue. Other portions of the pseudopeptides fpYIN and cpYIN are highly similar to each other as well as to fpYVN and cpYVN. The most representative structures of fpYVN and cpYVN are depicted in **Figure 5.7**, while the corresponding structures for fpYIN and cpYIN not shown due to their high similarity. The entropies calculated from quasiharmonic analysis of the four ligands in the bound state range from 345.6 to 358.0 cal/mol/K, with the largest entropy difference being only 12 cal/mol/K, which is within statistical error, between cpYVN and fpYIN. Therefore, quasiharmonic analysis of the protein and the ligands in the respective complexes suggests that the entropic advantage for binding of the unconstrained ligands fpYVN and fpYIN does not arise from any differences in the entropy of either the protein or the ligands in bound state.

5.4 CONCLUSION

In order to probe the origin of the unexpected entropic consequences observed for binding of a series of constrained and flexible phosphotyrosine-derived peptide analogs to the Grb2 SH2 domain,[14] we performed a series of calculations involving alchemical transformations at different temperatures on two sets of these analogs. Consistent with experimental results, these computations predicted that the binding affinity of the unconstrained peptides fpYV(I)N for the Grb2 SH2 domain is lower than that of the corresponding constrained peptides cpYV(I)N and that the mutation of V to I is not accompanied by significant changes in binding free energy. The experimental observation that unconstrained peptide analogs bind with more favorable entropies but significantly less favorable enthalpies than their constrained counterparts is also well reproduced by our computations, but the differences in the relative binding entropy and

enthalpy components are overestimated. Even though these simulations nicely reproduce those trends observed experimentally, further refinements to the method are needed in order to improve the accuracy of predicting differences in the relative binding thermodynamic parameters. Sampling of the conformations of the pseudo peptide ligands in solution is the most demanding calculation in this study. Some 26 perturbation steps were applied to introduce a bond between carbon atoms that are separated by two bonds, and REMD simulations with 48 replicas were performed at each perturbation step. Hence, in order to evaluate the relative solvation free energy/entropy of the two pairs of peptide ligands, about 6 μs MD simulations were performed; however, the statistical uncertainties remain significant. More advanced sampling techniques beyond the first-order scheme adopted in this study are needed in order to compute the solvation and binding free energy more efficiently and precisely.

We analyzed the structures and dynamics of ligands in solution and in their complexes with the SH2 domain in order to probe the molecular origin of the effects of ligand preorganization on binding thermodynamics in this system. Conformational clustering and quasiharmonic analysis of the free ligands in solution suggest that the unconstrained ligands possess significantly lower entropy than their constrained counterparts. This unexpected finding is the consequence of intramolecular hydrogen bonding interactions between the phosphate group of the flexible pY replacement and the C-terminalamide moieties of the pY+2 residue that lead to a more compact and rigid, macrocyclic-like structure. The presence of the cyclopropane ring in the constrained pY replacement prevents this interaction, thereby resulting in more extended conformations in solution. MD simulations of the protein-ligand complexes show that the unconstrained

and constrained ligands share similar binding modes; the distribution of conformations of the bound forms of the unconstrained and constrained ligands are comparable as are their nonbonded interactions with the domain. These findings are consistent with the structures determined experimentally by X-ray crystallography.

These simulations reveal an important caveat that has not been previously acknowledged regarding the use of ligand preorganization, which is widely presumed to have a favorable effect upon binding entropy as a general design strategy. Namely, this work demonstrates that *introducing a conformational constraint into a flexible ligand does not necessarily lower its entropy in solution*, because the flexibility of a ligand in solution is determined by a subtle balance between any intramolecular interactions and the intermolecular interactions between the ligand and its aqueous environment. Comparing the dominant structures of the constrained ligands in their bound and unbound states shows that the cyclopropane ring in the constrained ligands, cpYVN and cpYIN, locally constrains and orients functionality on the flexible phosphotyrosine replacement in fpYVN and fpYIN in the bound conformation as predicted from modeling studies. However, the macrocyclic-like structures of fpYVN and fpYIN in solution, *which do not correspond to their bound conformations,* reduce the global flexibility of these ligands to an even greater degree than the cyclopropane ring. Because the binding entropies for fpYVN and fpYIN are more favorable than for their constrained derivatives cpYVN and cpYIN, it is now apparent that one cannot think simply in terms of introducing constraints to stabilize the biologically active conformation of a small molecule as a strategy for enhancing ligand binding affinities. Rather, *lowering the entropy of a ligand in any way that allows it to adopt its bound conformation can lead to more favorable*

binding entropies. These studies also reveal that knowing the structures of small molecules in their unbound states is a critical prerequisite to correlating changes in their structures with protein binding entropies and enthalpies.

Table 5.1: Calculated and experimental thermodynamics (kcal/mol) for phosphotyrosine (pY)-containing peptide analogs and their constrained counterparts binding to the SH2 domain of Grb2. fpYV(I)N is the unconstrained tri-peptide analog consisting of pY, V (or I) and N residues; cpYV(I)N are the constrained counterparts (see Figure 5.1). The ΔG, ΔH and ΔS are the absolute binding free energy, enthalpy and entropy, respectively. With the calculated values of fpYVN set to experimental values, thermodynamics for the remaining ligands have been computed from the relative binding free energy and enthalpy obtained from MD simulations. Statistical errors of the calculated binding free energy are given in the parenthesis.

	Calculation			Experiment[301]		
	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
fpYVN	-7.7	-5.4	-2.3	-7.7	-5.4	-2.3
cpYVN	-10.0(0.6)	-15.7	5.7(1.2)	-8.8	-7.9	-0.8
fpYIN	-7.7(0.1)	-3.2	-4.5(0.2)	-7.7	-5.5	-2.2
cpYIN	-9.8(0.4)	-14.0	4.3(0.8)	-8.6	-8.3	-0.3

Table 5.2: Comparison of relative binding energetics (kcal/mol) between the unconstrained and constrained ligands (fpYVN/cpYVN and fpYIN/cpYIN), and decomposition of the relative binding free energy and entropy into the unbound (in solvent) and complex contributions.

	$\Delta\Delta G$ kcal/mol	$\Delta G_{solvent}$* kcal/mol	$\Delta G_{complex}$* kcal/mol	$-T\Delta\Delta S$ kcal/mol	$-T\Delta S_{solvent}$ kcal/mol	$-T\Delta S_{complex}$ kcal/mol
fpYVN->cpYVN	-2.27	26.68	24.41	8.05	-6.88	1.16
fpYIN->cpYIN	-2.07	26.80	24.73	7.45	-7.42	0.03

*The free energy contributions from solvent and complex (ΔGsolvent and ΔGcomplex) both include the relevant valence contributions from fpYV(I)N to cpYV(I)N. The valence contributions from water and protein were canceled in the calculation of relative binding free energy.

Table 5.3: Estimation of absolute configurational entropy by quasiharmonic analysis. Spro(complex), Slig(complex), and Slig(solvent) represent the entropy contributions of α-carbons of the Grb2 SH2 in solvated complex, ligands in solvated complex and unbound ligands in solvent, respectively. All the entropy contributions are in cal mol^{-1} K^{-1}.

	$S_{pro(complex)}$	$S_{lig(complex)}$	$S_{lig(solvent)}$
fpYVN	822.37	353.80	448.07
fpYIN	814.78	345.58	444.43
cpYVN	831.80	358.08	557.30
cpYIN	818.00	350.00	563.84

Table 5.4: The average numbers of intermolecular water-ligand hydrogen-bonds around the four solvated ligands, and the average numbers of intramolecular hydrogen bonds within the ligands in solution at 298.79 K.

	fpYVN	fpYIN	cpYVN	cpYIN
Intermolecular H-bond	33.5	33.2	35.7	35.2
Intramolecular H-bond	2.7	2.4	1.4	1.4

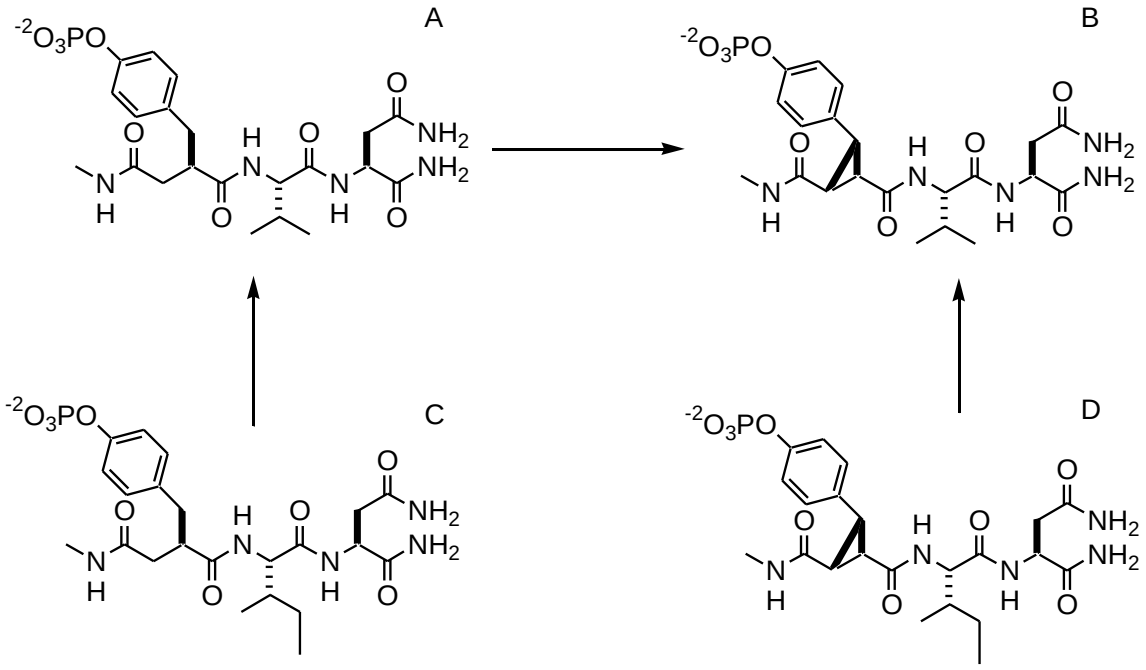

Figure 5.1: Chemical structures of the ligands studied and the perturbation scheme. A. fpYVN; B. cpYVN; C. fpYIN; D. cpYIN.

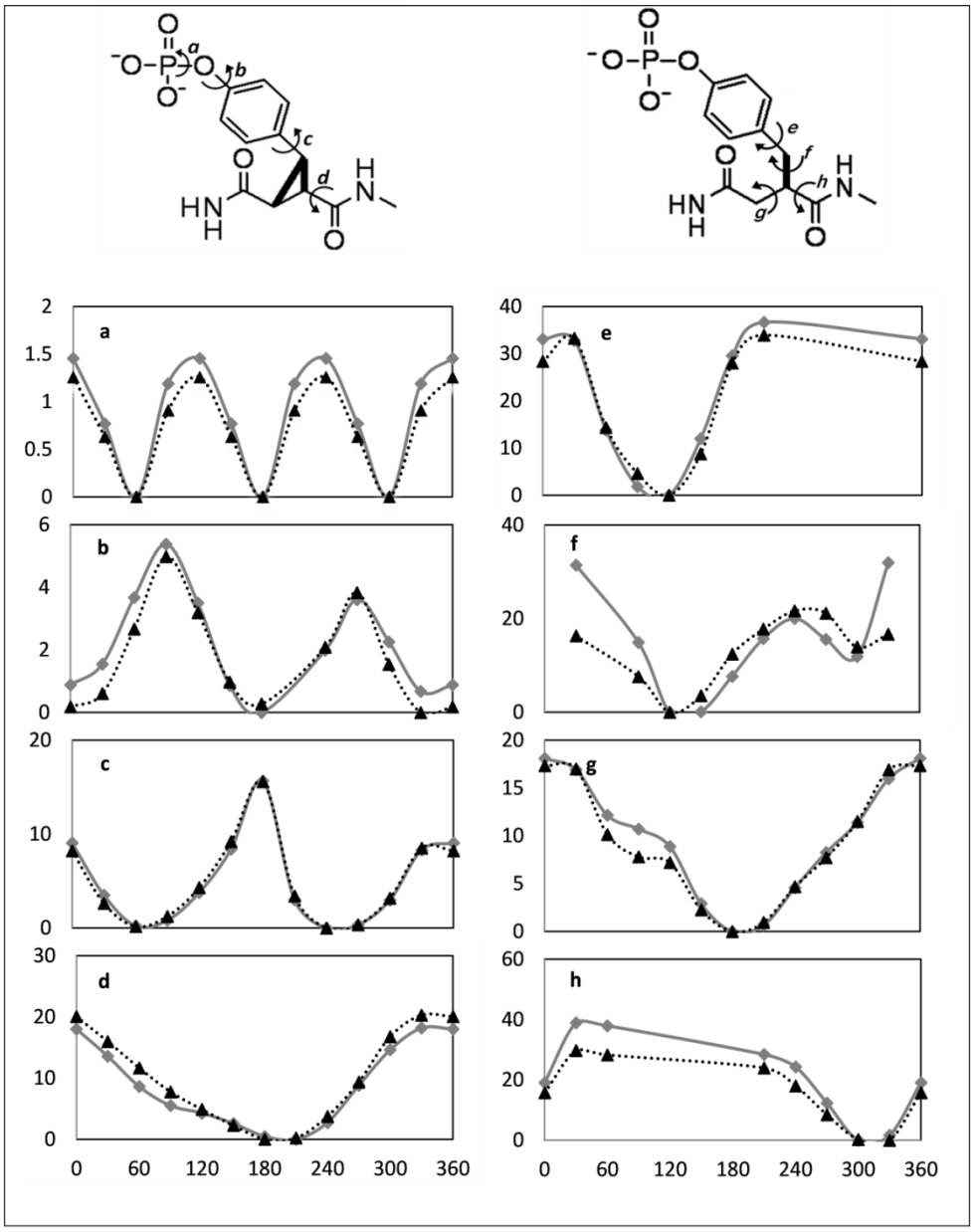

Figure 5.2: Conformational energy profiles for constrained and unconstrained pY segments. The grey lines with squares are QM relative energy, and the back dotted lines with triangles are the MM relative energy. Y-axis is the relative energy in kcal/mol. X-axis is the dihedral angle in degree.

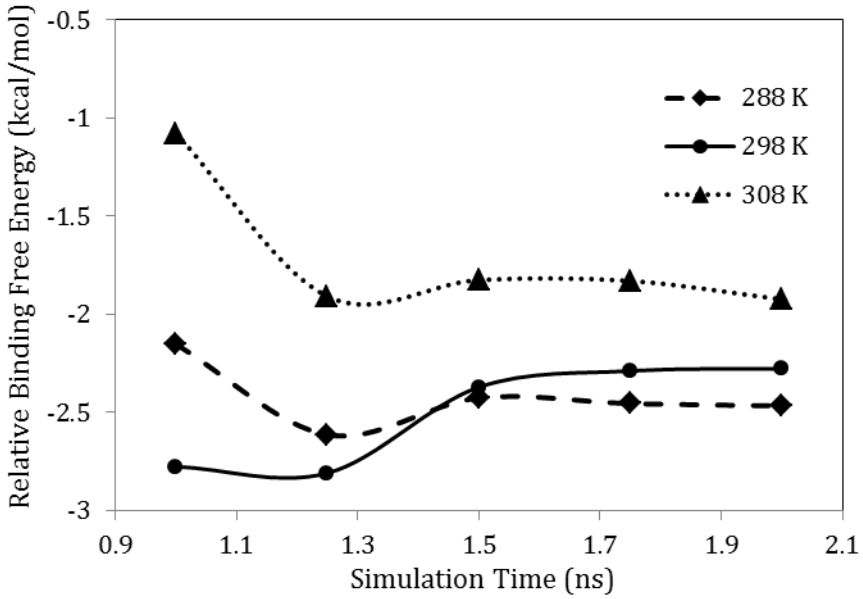

Figure 5.3: Convergence of relative binding free energy between fpYVN and cpYVN over simulation time at selected temperatures.

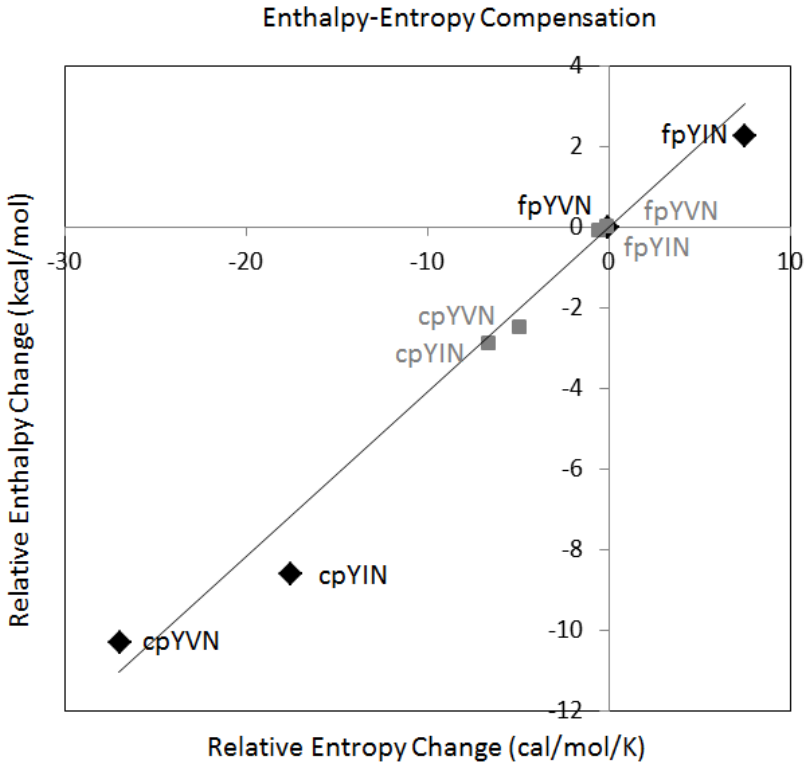

Figure 5.4: Correlation between binding enthalpy and binding entropy of fpYVN, fpYIN, cpYVN and cpYIN. Blue diamonds are calculated values; Red squares are experimental data. Both calculated and experimental binding enthalpy and entropy of fpYVN are shifted to zero for comparison purpose.

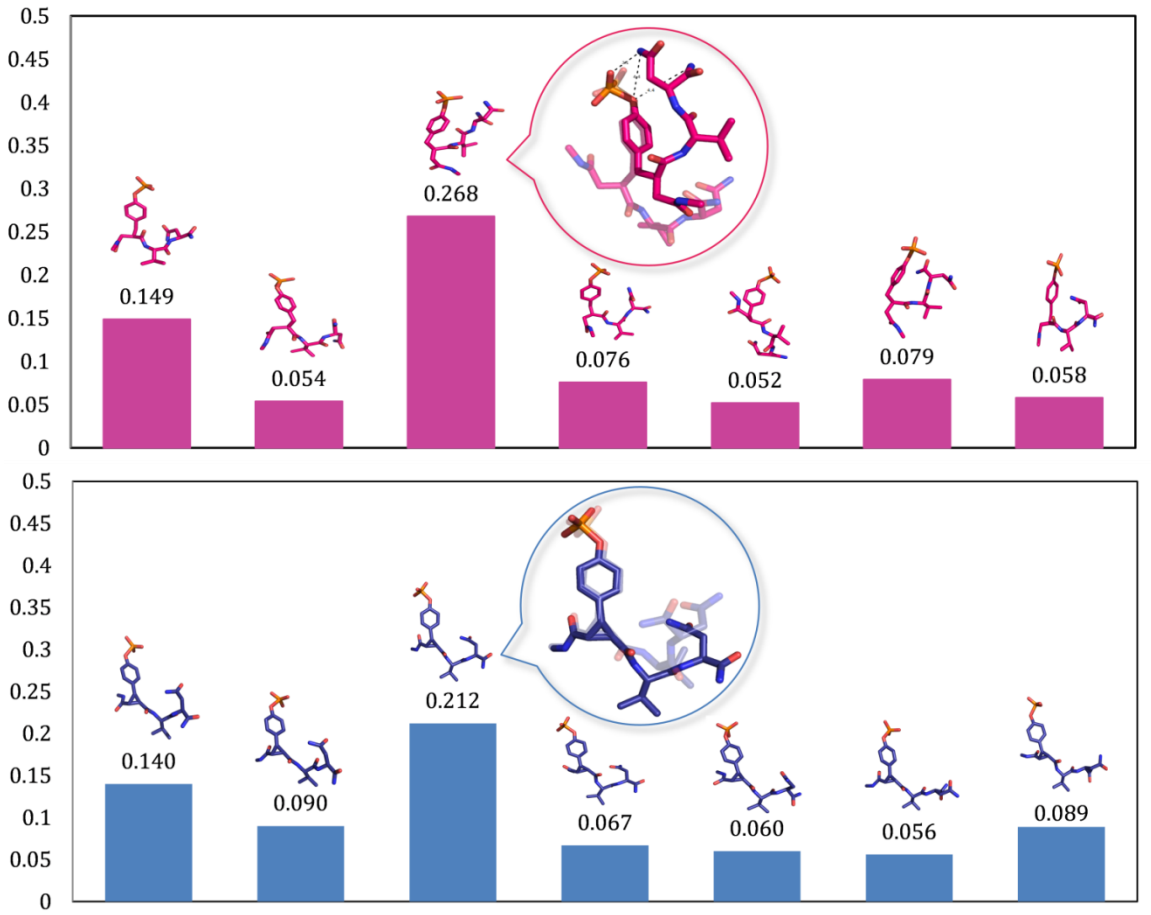

Figure 5.5: Clustering of the solvated ligand structures for fpYVN (pink) and cpYVN (blue). The most representative structures are plotted for clusters higher than 10%.

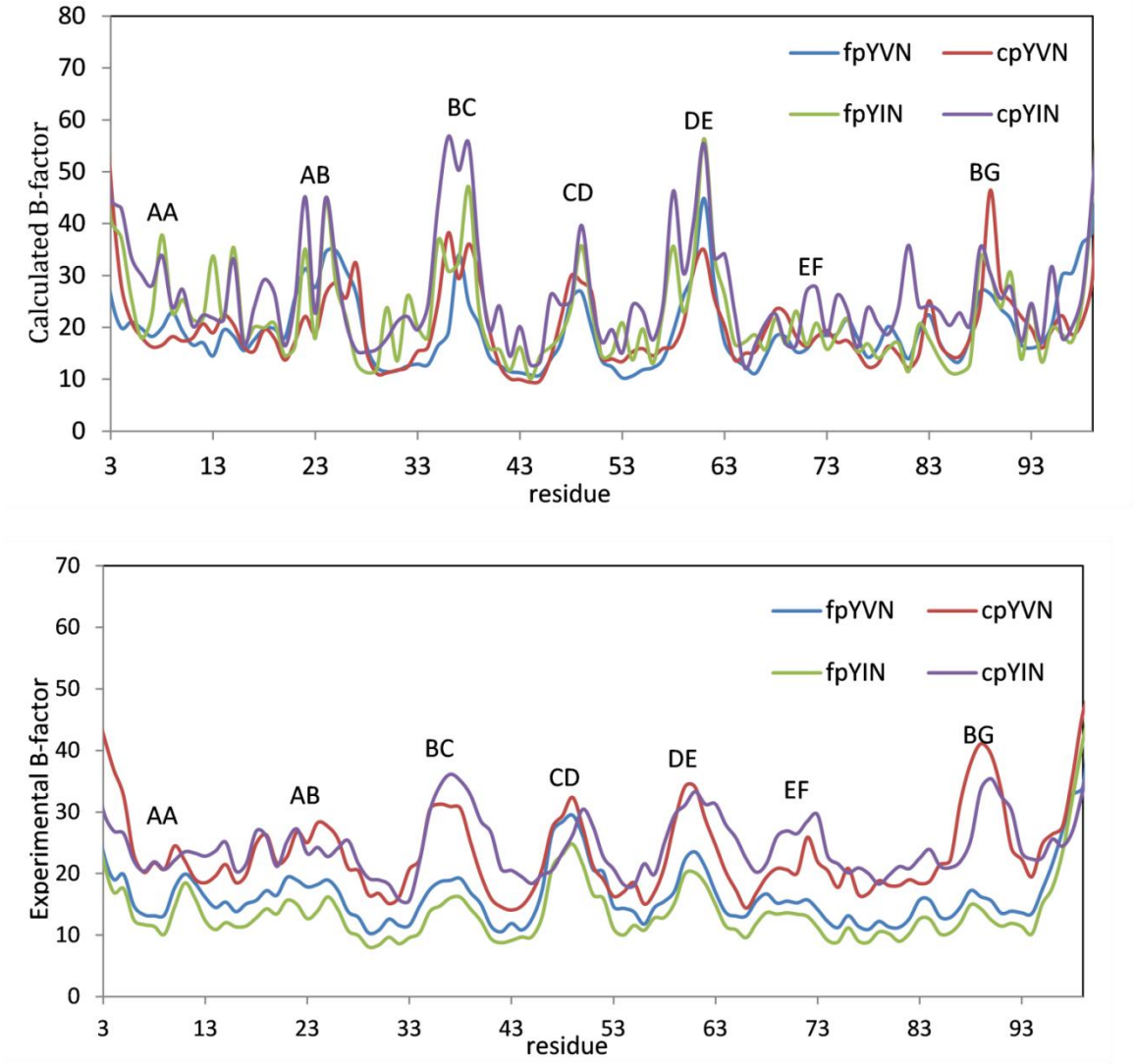

Figure 5.6: B-factor of α-carbons calculated from MD trajectories of fpYVN (blue), fpYIN (green), cpYVN (red) and cpYIN (purple) binding to Grb2 SH2 domain (top); B-factor of α-carbons of the four ligands binding to Grb2 SH2 domain from experiment (bottom).

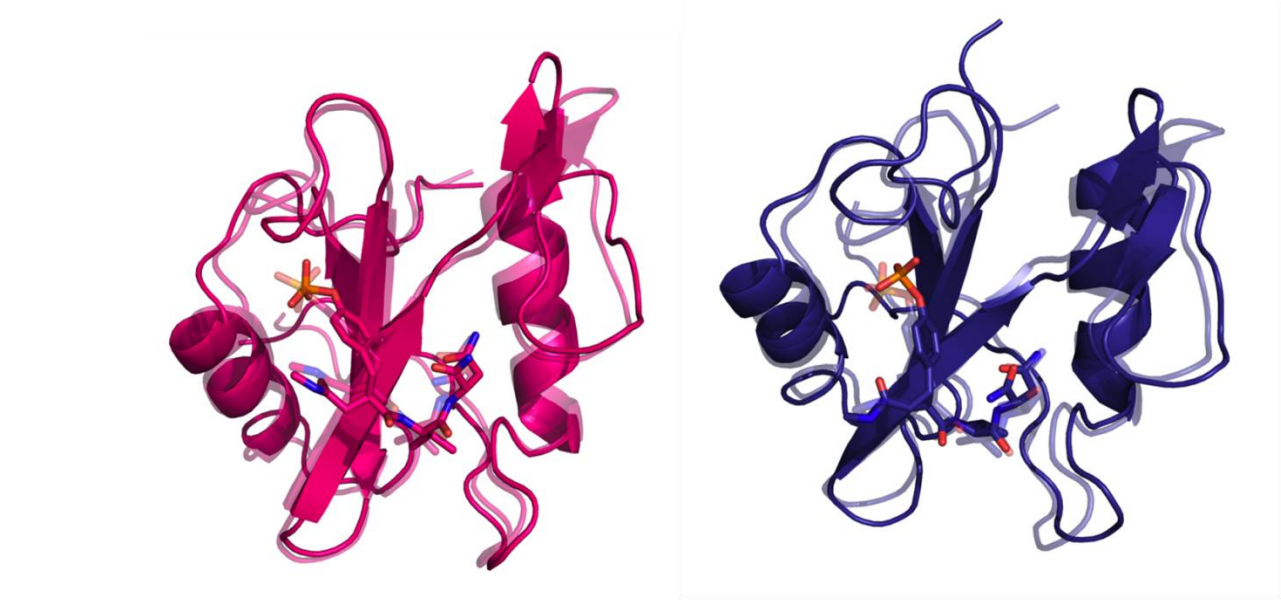

Figure 5.7: Representative structures from MD simulations of fpYVN (pink) and cpYVN (deep blue) binding to Grb2 SH2 domain. The structures are generated by pyMOL.

6 Conclusion

Modeling and prediction of binding affinity of a ligand to protein receptor will not only help advance our understanding of the underlying recognition mechanism but also facilitate the experimental drug discovery.[1, 272, 360] Although computational techniques have been extensively used in virtual screening and de novo drug design, predicting binding affinity within chemical accuracy remains challenging.[1, 2] Representation of electrostatic interactions with fixed atomic challenges imposes serious limitations on the accuracy of molecular modeling.[2, 3] Therefore, AMOEBA polarizable force field was developed, providing a better representation of the electrostatic interactions.

Accurately predicting HFE is recognized as one fundamental capability of molecular mechanics force field. We present a systematic investigation on HFE calculations with AMOEBA polarizable force field at various parameterization and simulation conditions. The HFEs of seven small organic molecules have been obtained alchemically using the Bennett Acceptance Ratio (BAR) method. We have compared two approaches to derive the atomic multipoles from quantum mechanical (QM) calculations: one directly from the new distributed multipole analysis (DMA) and the other involving fitting to the electrostatic potential around the molecules. Wave functions solved at the MP2 level with four basis sets (6-311G*, 6-311++G(2d,2p), cc-pVTZ, and aug-cc-pVTZ) are used to derive the atomic multipoles. HFEs from all four basis sets show a reasonable agreement with experimental data (root mean square error 0.63 kcal/mol for aug-cc-pVTZ). We conclude that aug-cc-pVTZ gives the best performance when used with AMOEBA, and 6-311++G(2d,2p) is comparable but more efficient for larger systems.

The results suggest that the inclusion of diffuse basis functions is important for capturing intermolecular interactions. The effect of long-range correction to van der Waals interaction on the hydration free energies is about 0.1 kcal/mol when the cutoff is 12Å, and increases linearly with the number of atoms in the solute/ligand.

Next, the AMOEBA force field for proteins is developed. The current version (AMOEBA'12) continues to utilize permanent electrostatic multipole moments through the quadrupole at each atom, and treats explicitly the polarization effects in various chemical and physical environments. The atomic electrostatic multipoles for each amino acid type were derived from high-level gas phase quantum mechanical calculations via a consistent and extendable protocol. Both the inter- and intramolecular polarization is treated via the mutual-interactive Thole-style model with short-range damping. The intramolecular polarization ensures the transferability of the electrostatic parameters among different conformations as demonstrated by good agreement between QM and AMOEBA electrostatic potentials and dipole moments of the dipeptides. The backbone and side chain torsional parameters were determined by comparing to both gas-phase QM (RI-TRIM MP2/CBS) conformational energy of dipeptides and statistical distributions from the Protein Data Bank. Molecular dynamics simulations have been performed on short peptides in explicit water to examine the conformational properties in solution. The calculated conformational free energy and *J*-coupling constants are consistent with the PDB statistics and experimental NMR results, respectively. In addition, the experimental crystal structures of a number of proteins were well maintained during the MD simulations. While extensive calculations will be necessary to fully validate the force field, the initial results suggest that this polarizable multipole-based force field is able to

describe the structure and energetics peptides and proteins in gas-phase and solution environments.

AMOEBA was then applied to the calculation of protein-ligand binding free energies of several benzamidine-like inhibitors to trypsin. All the computed binding free energies are in good agreement with the experimental data. From free energy decomposition, electrostatic interaction was found to be the driving force for the binding. Structural analysis shows that the ligands form hydrogen bonds with residues and water molecules nearby in a competitive fashion. The dependence of binding free energy on molecular dipole moment and polarizability was also studied. While the binding free energy is independent on the dipole moment, it shows a negative correlation with the polarizability.

Finally, AMOEBA was used to investigate the thermodynamic effect of constraining and hydrophobicity on binding energetics. Preorganizing a ligand in the conformation it adopts upon binding to a protein has long been considered to be an effective way to improve affinity by making the binding entropy more favorable. However, recent thermodynamic studies of a series of complexes of the Grb2 SH2 domain with peptide analogs having constrained and flexible replacements for a phosphotyrosine residue revealed that less favorable binding entropies may result from constraining ligands in their biologically active conformations. Toward probing the origin of this unexpected finding, we examined the complexes of four phosphotyrosine-derived analogs with the Grb2 SH2 domain using molecular dynamics simulations with a polarizable force field. Significantly, the computed values for the relative binding free energies, entropies and enthalpies of two pairs of constrained and unconstrained ligands

reproduced the trends that were determined experimentally, although the relative differences were overestimated. These calculations also revealed that a large fraction of the ligands lacking the constraining element exist in solution as compact, macro cyclic-like structures that are stabilized by interactions between the phosphate groups and the amide moieties of the *C*-terminal pY+2 residues. In contrast, the three-membered ring in the constrained ligands prevents the formation of such macro cyclic structures, leading instead to globally extended, less ordered conformations. Quasiharmonic analysis of these conformational ensembles suggests that the unconstrained ligands possess significantly lower entropies in solution, a finding that is consistent with the experimental observation that the binding entropies for the unconstrained ligands are more favorable than for their constrained counterparts. This study suggests that introducing local constraints in flexible molecules may have unexpected consequences, and a detailed understanding of the conformational preferences of ligands in their unbound states is a critical prerequisite to correlating changes in their chemical structure with protein binding entropies and enthalpies.

References

1. Gilson, M.K. and H.X. Zhou, Calculation of protein-ligand binding affinities. Annual Review of Biophysics and Biomolecular Structure, 2007. **36**: p. 21-42.
2. Mobley, D.L. and K.A. Dill, Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". Structure, 2009. **17**(4): p. 489-498.
3. Lamb, M.L. and W.L. Jorgensen, Computational approaches to molecular recognition. Current Opinion in Chemical Biology, 1997. **1**(4): p. 449-457.
4. Wang, W. and R.D. Skeel, Fast evaluation of polarizable forces. The Journal of Chemical Physics, 2005. **123**(16): p. 164107.
5. van der Avoird, A., et al., Vibration-rotation-tunneling states of the benzene dimer: an ab initio study. Physical chemistry chemical physics : PCCP, 2010. **12**(29): p. 8219-40.
6. Mccammon, J.A., B.R. Gelin, and M. Karplus, Dynamics of Folded Proteins. Nature, 1977. **267**(5612): p. 585-590.
7. Spackman, M.A., The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. Chemical Physics Letters, 2006. **418**(1-3): p. 158-162.
8. Nachtigallova, D., P. Hobza, and V. Spirko, Assigning the NH stretches of the Guanine tautomers using adiabatic separation: CCSD(T) benchmark calculations. Journal of Physical Chemistry A, 2008. **112**(9): p. 1854-1856.
9. Cornell, W.D., et al., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
10. MacKerell, A.D., et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
11. Valdes, H., et al., Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. Physical Chemistry Chemical Physics, 2008. **10**(19): p. 2747-2757.
12. Jorgensen, W.L., D.S. Maxwell, and J. TiradoRives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of

organic liquids. Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.

13. Ponder, J.W. and D.A. Case, Force fields for protein simulations. Advances in Protein Chemistry, 2003. **66**: p. 27-85.

14. Rezac, J., et al., Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems (Www.Begdb.Com): A Users Manual and Examples. Collection of Czechoslovak Chemical Communications, 2008. **73**(10): p. 1261-1270.

15. Rezac, J., K.E. Riley, and P. Hobza, S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. Journal of Chemical Theory and Computation, 2011. **7**(8): p. 2427-2438.

16. Berka, K., et al., Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. Journal of Chemical Theory and Computation, 2009. **5**(4): p. 982-992.

17. Barker, J.A., Statistical Mechanics of Interacting Dipoles. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1953. **219**(1138): p. 367-372.

18. Caldwell, J.W. and P.A. Kollman, Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. Journal of Physical Chemistry, 1995. **99**: p. 6208-6219.

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

20. Kaminski, G.A., et al., Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. Journal of Computational Chemistry, 2002. **23**(16): p. 1515-1531.

21. Friesner, R.A., Modeling polarization in proteins and protein-ligand complexes: Methods and preliminary results. Peptide Solvation and H-Bonds, 2006. **72**: p. 79-+.

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

23. Molnar, L.F., et al., Further analysis and comparative study of intermolecular interactions using dimers from the S22 database. The Journal of Chemical Physics, 2009. **131**(6): p. 065102.

24. Cieplak, P., J. Caldwell, and P. Kollman, Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases. Journal of Computational Chemistry, 2001. **22**(10): p. 1048-1057.

25. Wang, Z.X., et al., Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides (vol 27, pg 781, 2006). Journal of Computational Chemistry, 2006. **27**(8): p. 994-994.

26. Holt, A. and G. Karlstrom, Inclusion of the quadrupole moment when describing polarization. The effect of the dipole-quadrupole polarizability. Journal of Computational Chemistry, 2008. **29**(12): p. 2033-2038.

27. Moghaddam, S., et al., New ultrahigh affinity host-guest complexes of cucurbit[7]uril with bicyclo[2.2.2]octane and adamantane guests: thermodynamic analysis and evaluation of M2 affinity calculations. Journal of the American Chemical Society, 2011. **133**(10): p. 3570-81.

28. Geerke, D.P., W.F. van Gunsteren, and Sk, 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, 2007. **111**(23): p. 6425-6436.

29. Lamoureux, G., et al., A polarizable model of water for molecular dynamics simulations of biomolecules. Chemical Physics Letters, 2006. **418**(1-3): p. 245-249.

30. Lamoureux, G., A.D. MacKerell, and B. Roux, A simple polarizable model of water based on classical Drude oscillators. Journal of Chemical Physics, 2003. **119**(10): p. 5185-5197.

31. Banks, J.L., et al., Parametrizing a Polarizable Force Field from ab Initio Data. I. The Fluctuating Point Charge Model. Journal of Chemical Physics, 1999. **110**(2): p. 741-754.

32. Patel, S. and C.L. Brooks, CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. Journal of Computational Chemistry, 2004. **25**(1): p. 1-15.

33. Rappe, A.K. and W.A. Goddard, Charge Equilibration for Molecular-Dynamics Simulations. Journal of Physical Chemistry, 1991. **95**(8): p. 3358-3363.

34. Song, L., et al., Explicit polarization (X-Pol) potential using ab initio molecular orbital theory and density functional theory. The journal of physical chemistry. A, 2009. **113**(43): p. 11656-64.

35. Mills, M.J.L. and P.L.A. Popelier, Polarisable multipolar electrostatics from the machine learning method Kriging: an application to alanine. Theoretical Chemistry Accounts, 2012. **131**(3).

36. Cieplak, P., et al., Polarization effects in molecular mechanical force fields. Journal of Physics-Condensed Matter, 2009. **21**(33): p. 333102 1-21

37. Halgren, T.A. and W. Damm, Polarizable force fields. Current Opinion in Structural Biology, 2001. **11**(2): p. 236-242.

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

39. Ponder, J.W., et al., Current Status of the AMOEBA Polarizable Force Field. Journal of Physical Chemistry B, 2010. **114**(8): p. 2549-2564 PMCID: PMC2918242

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

41. Rick, S.W. and S.J. Stuart, Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. Reviews in Computational Chemistry, 2002. **18**: p. 89-146.

42. Patel, S., J.E. Davis, and B.A. Bauer, Exploring ion permeation energetics in gramicidin A using polarizable charge equilibration force fields. Journal of the American Chemical Society, 2009. **131**(39): p. 13890-1.

43. Applequist, J., J.R. Carl, and K.-K. Fung, An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. Journal of the American Chemical Society, 1972. **94**(9): p. 2952-2960.

44. Lesueur, C.R. and A.J. Stone, Practical Schemes for Distributed Polarizabilities. Molecular Physics, 1993. **78**(5): p. 1267-1291.

45. Garmer, D.R. and W.J. Stevens, Transferability of Molecular Distributed Polarizabilities from a Simple Localized Orbital Based Method. Journal of Physical Chemistry, 1989. **93**(25): p. 8263-8270.

46. Thole, B.T., Molecular Polarizabilities Calculated With A Modified Dipole Interaction. Chemical Physics, 1981. **59**(3): p. 341-350.

47. Truchon, J.F., et al., Accurate molecular polarizabilities based on continuum electrostatics. Journal of Chemical Theory and Computation, 2008. **4**(9): p. 1480-1493.

48. Vanbelle, D., et al., Molecular-Dynamics Simulation of Polarizable Water by an Extended Lagrangian Method. Molecular Physics, 1992. **77**(2): p. 239-255.

49. Stone, A.J., Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. Chemical Physics Letters, 1981. **83**(2): p. 233-239.

50. Stone, A.J., Distributed Multipole Analysis: Methods and Applications. Molecular Physics, 1985. **56**(5): p. 1047-1064.

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

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

53. Jiao, D., et al., Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. Journal of Physical Chemistry B, 2006. **110**(37): p. 18553-18559.

54. Wu, J.C., et al., Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. Journal of Chemical Theory and Computation, 2010. **6**(7): p. 2059-2070 PMCID: PMC2992432.

55. Ren, P., C. Wu, and J.W. Ponder, Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. Journal of Chemical Theory and Computation, 2011. **7**(10): p. 3143-3161 PMCID: PMC3196664.

56. Shi, Y., et al., Multipole electrostatics in hydration free energy calculations. Journal of Computational Chemistry, 2011. **32**(5): p. 967-977 PMCID: PMC 3073856.

57. Jiang, J.L., et al., Assessing the Performance of Popular Quantum Mechanics and Molecular Mechanics Methods and Revealing the Sequence-Dependent Energetic Features Using 100 Tetrapeptide Models. Journal of Chemical Theory and Computation, 2010. **6**(4): p. 1199-1209.

58. Jiao, D., et al., 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, 2008. **105**(17): p. 6290-6295. PMCID: PMC2359813.

59. Zhang, J., et al., Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. Journal of Chemical Theory and Computation, 2012. **8**: p. 1314-1324 PMCID: PMC3383645

60. Shi, Y., et al., Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations.

Journal of Physical Chemistry B, 2012. **116**(5): p. 1716-27 PMCID: PMC3277292.

61. Jiang, W. and B. Roux, Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. Journal of Chemical Theory and Computation, 2010. **6**(9): p. 2559-2565.

62. Jiao, D., et al., Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. Journal of Computational Chemistry, 2009. **30**(11): p. 1701-1711 PMCID: PMC2752704.

63. Shi, Y., et al., Trypsin-ligand binding free energy calculation with AMOEBA. Engineering in Medicine and Biology Society (EMBC). EMBC Annual International Conference of the IEEE, 2009: p. 2328-2331 PMCID: PMC2819397.

64. Schnieders, M.J., T.D. Fenn, and V.S. Pande, Polarizable Atomic Multipole X-Ray Refinement: Particle Mesh Ewald Electrostatics for Macromolecular Crystals. Journal of Chemical Theory and Computation, 2011. **7**(4): p. 1141-1156.

65. Schnieders, M.J., et al., Polarizable atomic multipole X-ray refinement: application to peptide crystals. Acta Crystallographica Section D-Biological Crystallography, 2009. **65**: p. 952-965.

66. Schnieders, M.J., et al., The Structure, Thermodynamics and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. Journal of Chemical Theory and Computation, 2012. **8**(5): p. 1721-1736 PMCID: PMC3348590.

67. Ponder, J.W., TINKER: Software Tools for Molecular Design Washington University http://dasher.wustl.edu/tinker/, 2012. **v6.0**.

68. Friedrichs, M.S., et al., Accelerating molecular dynamic simulation on graphics processing units. Journal of Computational Chemistry, 2009. **30**(6): p. 864-72.

69. Case, D.A., et al., The Amber biomolecular simulation programs. Journal of Computational Chemistry, 2005. **26**(16): p. 1668-1688.

70. Schnieders, M.J., et al., Force Field X open source, platform independent modules for molecular biophsyics simulations. http://ffx.kenai.com, 2011.

71. Ponder, J.W., TINKER: Software Tools for Molecular Design, 2009: Saint Louis, MO. p. TINKER: Software Tools for Molecular Design.

72. Schnieders, M.J., et al., Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. Journal of Chemical Physics, 2007. **126**(12).

73. Baker, N.A., et al., Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-41.

74. Schnieders, M.J. and J.W. Ponder, Polarizable atomic multipole solutes in a generalized Kirkwood continuum. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2083-2097.

75. Brooks, B.R., et al., CHARMM: the biomolecular simulation program. Journal of Computational Chemistry, 2009. **30**(10): p. 1545-614.

76. Woodcock, H.L., et al., MSCALE: A General Utility for Multiscale Modeling. Journal of Chemical Theory and Computation, 2011. **7**(4): p. 1208-1219.

77. Zheng, L., M. Chen, and W. Yang, Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proceedings of the National Academy of Sciences, 2008. **105**(51): p. 20227-20232.

78. Fenn, T.D. and M.J. Schnieders, Polarizable atomic multipole X-ray refinement: weighting schemes for macromolecular diffraction. Acta Crystallographica Section D, 2011. **67**(11): p. 957-65.

79. Eastman, P. and V.S. Pande, Efficient Nonbonded Interactions for Molecular Dynamics on a Graphics Processing Unit. Journal of Computational Chemistry, 2010. **31**(6): p. 1268-1272.

80. Gresh, N., P. Claverie, and A. Pullman, Theoretical-Studies of Molecular-Conformation - Derivation of an Additive Procedure for the Computation of Intramolecular Interaction Energies - Comparison with Abinitio Scf Computations. Theoretica Chimica Acta, 1984. **66**(1): p. 1-20.

81. Gresh, N., Energetics of Zn2+ Binding to a Series of Biologically Relevant Ligands - a Molecular Mechanics Investigation Grounded on Ab-Initio Self-Consistent-Field Supermolecular Computations. Journal of Computational Chemistry, 1995. **16**(7): p. 856-882.

82. Piquemal, J.P., et al., Inclusion of the ligand field contribution in a polarizable molecular mechanics: SIBFA-LF. Journal of Computational Chemistry, 2003. **24**(16): p. 1963-1970.

83. Piquemal, J.P., H. Chevreau, and N. Gresh, Toward a separate reproduction of the contributions to the Hartree-Fock and DFT intermolecular interaction energies by polarizable molecular mechanics with the SIBFA potential. Journal of Chemical Theory and Computation, 2007. **3**(3): p. 824-837.

84. Gresh, N., et al., Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-

up strategy. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 1960-1986.

85. Piquemal, J.P., N. Gresh, and C. Giessner-Prettre, Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. Journal of Physical Chemistry A, 2003. **107**(48): p. 10353-10359.

86. Gresh, N., P. Claverie, and A. Pullman, Intermolecular Interactions - Elaboration on an Additive Procedure Including an Explicit Charge-Transfer Contribution. International Journal of Quantum Chemistry, 1986. **29**(1): p. 101-118.

87. Karplus, M., Contact Electron-Spin Coupling of Nuclear Magnetic Moments. Journal of Chemical Physics, 1959. **30**(1): p. 11-15.

88. Antony, J., J.P. Piquemal, and N. Gresh, Complexes of thiomandelate and captopril mercaptocarboxylate inhibitors to metallo-beta-lactamase by polarizable molecular mechanics. Validation on model binding sites by quantum chemistry. Journal of Computational Chemistry, 2005. **26**(11): p. 1131-1147.

89. Roux, C., et al., Binding of 5-phospho-D-arabinonohydroxamate and 5-phospho-D-arabinonate inhibitors to zinc phosphomannose isomerase from Candida albicans studied by polarizable molecular mechanics and quantum mechanics. Journal of Computational Chemistry, 2007. **28**(5): p. 938-957.

90. Jenkins, L.M.M., et al., Specificity of acyl transfer from 2-mercaptobenzamide thioesters to the HIV-1 nucleocapsid protein. Journal of the American Chemical Society, 2007. **129**(36): p. 11067-11078.

91. Foret, J., et al., Synthesis and evaluation of non-hydrolyzable D-mannose 6-phosphate surrogates reveal 6-deoxy-6-dicarboxymethyl-D-mannose as a new strong inhibitor of phosphomannose isomerases. Bioorganic & Medicinal Chemistry, 2009. **17**(20): p. 7100-7107.

92. Gresh, N., et al., Analysis of the Interactions Taking Place in the Recognition Site of a Bimetallic Mg(II)-Zn(II) Enzyme, Isopentenyl Diphosphate Isomerase. A Parallel Quantum-Chemical and Polarizable Molecular Mechanics Study. Journal of Physical Chemistry B, 2010. **114**(14): p. 4884-4895.

93. Roux, C., et al., The reaction mechanism of type I phosphomannose isomerases: new information from inhibition and polarizable molecular mechanics studies. Proteins-Structure Function and Bioinformatics, 2011. **79**(1): p. 203-20.

94. Ledecq, M., et al., Modeling of copper(II) complexes with the SIBFA polarizable molecular mechanics procedure. Application to a new class of HIV-1 protease inhibitors. Journal of Physical Chemistry B, 2003. **107**(38): p. 10640-10652.

95. Marjolin, A., et al., Toward accurate solvation dynamics of lanthanides and actinides in water using polarizable force fields: from gas-phase energetics to hydration free energies. Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta), 2012. **131**(4): p. 1-14.

96. Rogalewicz, F., G. Ohanessian, and N. Gresh, Interaction of neutral and zwitterionic glycine with Zn2+ in gas phase: Ab initio and SIBFA molecular mechanics calculations. Journal of Computational Chemistry, 2000. **21**(11): p. 963-973.

97. Gresh, N., Inter- and intramolecular interactions. Inception and refinements of the SIBFA, molecular mechanics (SMM) procedure, a separable, polarizable methodology grounded on ab initio SCF/MP2 computations. Examples of applications to molecular recognition problems. Journal De Chimie Physique Et De Physico-Chimie Biologique, 1997. **94**(7-8): p. 1365-1416.

98. Gresh, N., G. Tiraboschi, and D.R. Salahub, Conformational properties of a model alanyl dipeptide and of alanine-derived oligopeptides: Effects of solvation in water and in organic solvents - A combined SIBFA/Continuum reaction field, ab initio Self-Consistent Field, and Density Functional Theory investigation. Biopolymers, 1998. **45**(6): p. 405-425.

99. Graf, J., et al., Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study. Journal of the American Chemical Society, 2007. **129**(5): p. 1179-1189.

100. Cisneros, G.A., et al., Design of next generation force fields from ab initio computations: beyond point charges electrostatics, in *Multi-scale Quantum Models for Biocatalysis*, D.M. York and T.S. Lee, Editors. 2009, Springer Science.

101. Piquemal, J.P., et al., Towards a force field based on density fitting. Journal of Chemical Physics, 2006. **124**(10).

102. Piquemal, J.-P., et al., Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. The Journal of Chemical Physics, 2006. **125**(5): p. 054511-7.

103. Brdarski, S. and G. Karlstrom, Modeling of the exchange repulsion energy. Journal of Physical Chemistry A, 1998. **102**(42): p. 8182-8192.

104. Carignano, M.A., G. Karlstrom, and P. Linse, Polarizable ions in polarizable water: A molecular dynamics study. Journal of Physical Chemistry B, 1997. **101**(7): p. 1142-1147.

105. Holt, A. and G. Karlstrom, Improvement of the NEMO Potential by Inclusion of Intramolecular Polarization. International Journal of Quantum Chemistry, 2009. **109**(6): p. 1255-1266.

106. Gagliardi, L., R. Lindh, and G. Karlstrom, Local properties of quantum chemical systems: The LoProp approach. Journal of Chemical Physics, 2004. **121**(10): p. 4494-4500.

107. Hermida-Ramon, J.M., et al., Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. Journal of Computational Chemistry, 2003. **24**(2): p. 161-176.

108. Hagberg, D., et al., The coordination of uranyl in water: A combined quantum chemical and molecular simulation study. Journal of the American Chemical Society, 2005. **127**(41): p. 14250-14256.

109. Kollman, P.A., et al., Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Accounts of Chemical Research, 2000. **33**(12): p. 889-897.

110. Drude, P., C.R. Mann, and R.A. Millikan, The theory of optics1902, New York etc.: Longmans, Green, and Co. 1 p. l., xxi, 546 p.

111. Harder, E., et al., Understanding the dielectric properties of liquid amides from a polarizable force field. Journal of Physical Chemistry B, 2008. **112**(11): p. 3509-3521.

112. Dror, R.O., et al., Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annual Review of Biophysics, 2012. **41**(1): p. 429-452.

113. Myshakin, E.M., H. Jiang, and K.D. Jordan, PHYS 549-Molecular dynamics simulations of methane hydrate decomposition using a polarizable force field. Abstracts of Papers of the American Chemical Society, 2007. **234**.

114. Lopes, P.E.M., et al., Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. Journal of Physical Chemistry B, 2007. **111**(11): p. 2873-2885.

115. Davis, J.E. and S. Patel, Charge Equilibration Force Fields for Lipid Environments: Applications to Fully Hydrated DPPC Bilayers and DMPC-Embedded Gramicidin A. Journal of Physical Chemistry B, 2009. **113**(27): p. 9183-9196.

116. Xiong, R., et al., Some insights into the binding mechanism of Aurora B kinase gained by molecular dynamics simulation. Journal of molecular modeling, 2012.

117. Kaoud, T.S., et al., Manipulating JNK Signaling with (-)-Zuonin A. ACS chemical biology, 2012.

118. Warren, G.L. and S. Patel, Hydration free energies of monovalent ions in transferable intermolecular potential four point fluctuating charge water: An assessment of simulation methodology and force field performance and transferability. Journal of Chemical Physics, 2007. **127**(6): p. -.

119. Baker, C.M., et al., Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. Journal of Chemical Theory and Computation, 2010. **6**(4): p. 1181-1198.

120. Wang, J.M., et al., Development and testing of a general amber force field. Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.

121. Whitfield, T.W., et al., Theoretical study of aqueous solvation of K(+) comparing ab initio, polarizable, and fixed-charge models. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2068-2082.

122. Yu, H.B., et al., Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. Journal of Chemical Theory and Computation, 2010. **6**(3): p. 774-786.

123. Lu, Z.Y. and Y.K. Zhang, Interfacing ab initio quantum mechanical method with classical Drude osillator polarizable model for molecular dynamics simulation of chemical reactions. Journal of Chemical Theory and Computation, 2008. **4**(8): p. 1237-1248.

124. Brooks, B.R., et al., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

125. Phillips, J.C., et al., Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 2005. **26**(16): p. 1781-802.

126. Jiang, W., et al., High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. Journal of Physical Chemistry Letters, 2011. **2**(2): p. 87-92.

127. Kaoud, T.S., et al., From in Silico Discovery to Intracellular Activity: Targeting JNK‚ÄìProtein Interactions with Small Molecules. Acs Medicinal Chemistry Letters, 2012. **3**(9): p. 721-725.

128. Chodera, J.D., et al., Dynamical reweighting: improved estimates of dynamical properties from simulations at multiple temperatures. The Journal of Chemical Physics, 2011. **134**(24): p. 244107.

129. Shirts, M.R., et al., Accurate and efficient corrections for missing dispersion interactions in molecular simulations. The journal of physical chemistry. B, 2007. **111**(45): p. 13052-63.

130. Bauer, B.A., et al., Water Permeation Through DMPC Lipid Bilayers using Polarizable Charge Equilibration Force Fields. Chemical Physics Letters, 2011. **508**(4-6): p. 289-294.

131. Bauer, B.A., G.L. Warren, and S. Patel, Incorporating Phase-Dependent Polarizability in Nonadditive Electrostatic Models for Molecular Dynamics Simulations of the Aqueous Liquid-Vapor Interface. Journal of Chemical Theory and Computation, 2009. **5**(2): p. 359-373.

132. Xie, W. and J. Gao, The Design of a Next Generation Force Field: The X-POL Potential. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 1890-1900.

133. Gao, J.L., A molecular-orbital derived polarization potential for liquid water. Journal of Chemical Physics, 1998. **109**(6): p. 2346-2354.

134. Wierzchowski, S.J., D.A. Kofke, and J.L. Gao, Hydrogen fluoride phase behavior and molecular structure: A QM/MM potential model approach. Journal of Chemical Physics, 2003. **119**(14): p. 7365-7371.

135. Xie, W.S., et al., Incorporation of a QM/MM Buffer Zone in the Variational Double Self-Consistent Field Method. Journal of Physical Chemistry B, 2008. **112**(45): p. 14124-14131.

136. Xie, W.S., et al., The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. Journal of Chemical Physics, 2008. **128**(23).

137. Wang, Y.J., et al., Multilevel X-Pol: A Fragment-Based Method with Mixed Quantum Mechanical Representations of Different Fragments. Journal of Physical Chemistry B, 2012. **116**(23): p. 6781-6788.

138. Kaminski, G.A., R.A. Friesner, and R.H. Zhou, A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. Journal of Computational Chemistry, 2003. **24**(3): p. 267-276.

139. Kaminski, G.A., et al., Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. Journal of Physical Chemistry A, 2004. **108**(4): p. 621-627.

140. Kim, B.C., et al., Structure and dynamics of the solvation of bovine pancreatic trypsin inhibitor in explicit water: A comparative study of the effects of solvent

and protein polarizability. Journal of Physical Chemistry B, 2005. **109**(34): p. 16529-16538.

141. Kaminski, G.A., S.Y. Ponomarev, and A.B. Liu, Polarizable Simulations with Second order Interaction Model - force field and software for fast polarizable calculations: Parameters for small model systems and free energy calculations. Journal of Chemical Theory and Computation, 2009. **5**(11): p. 2935-2943.

142. Ponomarev, S.Y. and G.A. Kaminski, Polarizable Simulations with Second order Interaction Model (POSSIM) force field: Developing parameters for alanine peptides and protein backbone. Journal of Chemical Theory and Computation, 2011. **7**(5): p. 1415-1427.

143. Ponomarev, S.Y. and G. Kaminski, Polarizable force field for protein simulations POSSIM: Alanine dipeptide and tetrapeptide parameters, and stability of the alanine 13 alpha-helix in water. Abstracts of Papers of the American Chemical Society, 2011. **241**.

144. Shoichet, B.K., A.R. Leach, and I.D. Kuntz, Ligand solvation in molecular docking. Proteins: Structure, Function, and Genetics, 1999. **34**(1): p. 4-16.

145. Jiang, H., K.D. Jordan, and C.E. Taylor, Molecular dynamics simulations of methane hydrate using polarizable force fields. Journal of Physical Chemistry B, 2007. **111**(23): p. 6486-6492.

146. Kaminski, G., et al., Free-Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the Opls All-Atom Model. Journal of Physical Chemistry, 1994. **98**(49): p. 13077-13082.

147. Mobley, D.L., et al., Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. Journal of Physical Chemistry B, 2007. **111**(9): p. 2242-2254.

148. Mobley, D.L., et al., Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. Journal of Chemical Theory and Computation, 2009. **5**(2): p. 350-358.

149. Rizzo, R.C., et al., Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial, charge models and optimization of nonpolar contributions. Journal of Chemical Theory and Computation, 2006. **2**(1): p. 128-139.

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

151. Gallicchio, E. and R.M. Levy, AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. Journal of Computational Chemistry, 2004. **25**(4): p. 479-499.

152. Gallicchio, E., K. Paris, and R.M. Levy, The AGBNP2 Implicit Solvation Model. Journal of Chemical Theory and Computation, 2009. **5**(9): p. 2544-2564.

153. Tan, C., Y.H. Tan, and R. Luo, Implicit nonpolar solvent models. Journal of Physical Chemistry B, 2007. **111**(42): p. 12263-12274.

154. Wagoner, J.A. and N.A. Baker, Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(22): p. 8331-8336.

155. Onufriev, A., Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview Annual Reports in Computational Chemistry, 2008. **4**: p. 125-137.

156. Mobley, D.L., et al., Charge asymmetries in hydration of polar solutes. Journal of Physical Chemistry B, 2008. **112**(8): p. 2405-2414.

157. Purisima, E.O. and T. Sulea, Restoring Charge Asymmetry in Continuum Electrostatics Calculations of Hydration Free Energies. Journal of Physical Chemistry B, 2009. **113**(4): p. 8206–8209.

158. Mongan, J., et al., Generalized Born model with a simple, robust molecular volume correction. Journal of Chemical Theory and Computation, 2007. **3**(1): p. 156-169.

159. Straatsma, T.P. and H.J.C. Berendsen, Free-Energy of Ionic Hydration - Analysis of a Thermodynamic Integration Technique to Evaluate Free-Energy Differences by Molecular-Dynamics Simulations. Journal of Chemical Physics, 1988. **89**(9): p. 5876-5886.

160. Frenkel, D. and A.J.C. Ladd, New Monte-Carlo Method to Compute the Free-Energy of Arbitrary Solids - Application to the Fcc and Hcp Phases of Hard-Spheres. Journal of Chemical Physics, 1984. **81**(7): p. 3188-3193.

161. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. Chemical Reviews, 1993. **93**(7): p. 2395-2417.

162. Jorgensen, W.L., et al., Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

163. Shivakumar, D., et al., Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. Journal of Chemical Theory and Computation, 2010. **6**(5): p. 1509-1519.

164. Shirts, M.R., et al., Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. Journal of Chemical Physics, 2003. **119**(11): p. 5740-5761.

165. Chipot, C., Rational determination of charge distributions for free energy calculations. Journal of Computational Chemistry, 2003. **24**(4): p. 409-415.

166. Hess, B. and N.F.A. van der Vegt, Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. Journal of Physical Chemistry B, 2006. **110**(35): p. 17616-17626.

167. Baker, C.M., et al., Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. Journal of Chemical Theory and Computation, 2010. **6**(4): p. 1181-1198.

168. Stone, A.J., Distributed Multipole Analysis, or How to Describe a Molecular Charge-Distribution. Chemical Physics Letters, 1981. **83**(2): p. 233-239.

169. Stone, A.J. and M. Alderton, Distributed multipole analysis - Methods and applications (Reprinted from Molecular Physics, vol 56, pg 1047-1064, 1985). Molecular Physics, 2002. **100**(1): p. 221-233.

170. Bader, R.F.W., Atoms in Molecules—A Quantum Theory. International Series of Monographs on Chemistry1990, Oxford: Clarendon Press.

171. Sokalski, W.A., et al., Point-Charge Representation of Multicenter Multipole Moments in Calculation of Electrostatic Properties. Theoretica Chimica Acta, 1993. **85**(1-3): p. 209-216.

172. Sokalski, W.A. and R.A. Poirier, Cumulative Atomic Multipole Representation of the Molecular Charge-Distribution and Its Basis Set Dependence. Chemical Physics Letters, 1983. **98**(1): p. 86-92.

173. Williams, D.E., Representation of the Molecular Electrostatic Potential by Atomic Multipole and Bond Dipole Models. Journal of Computational Chemistry, 1988. **9**(7): p. 745-763.

174. Shi, Y., et al., Trypsin-ligand binding free energy calculation with AMOEBA. IEEE Engineering in Medicine and Biology Society, 2009(3-6): p. 2328-2331.

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

176. Halgren, T.A., Representation of Vanderwaals (Vdw) Interactions in Molecular Mechanics Force-Fields - Potential Form, Combination Rules, and Vdw Parameters. Journal of the American Chemical Society, 1992. **114**(20): p. 7827-7843.

177. Ponder, J.W., TINKER: Software Tools for Molecular Design Washington University Medical School, 2009. **Version 5.1 Available at:http://dasher.wustl.edu/tinker/.**

178. Stone, A.J., GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System. University of Cambridge, 2005. **version 2.2. 02**.

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

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

181. Berendsen, H.J.C., et al., Molecular-Dynamics with Coupling to an External Bath. Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.

182. Essmann, U., et al., A Smooth Particle Mesh Ewald Method. Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.

183. Darden, T., D. York, and L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.

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

185. Shirts, M.R., et al., Accurate and efficient corrections for missing dispersion interactions in molecular Simulations. Journal of Physical Chemistry B, 2007. **111**(45): p. 13052-13063.

186. Cabani, S., et al., Group Contribution to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution. Journal of Solution Chemistry, 1981. **10**: p. 563-595.

187. Wolfenden, R., et al., Cooperativity and Anticooperativity in Solvation by Water: Imidazoles, Quinones, Nitrophenols, Nitrophenolate, and Nitrothiophenolate Ions. Journal of the American Chemical Society, 1987. **109**: p. 463–466.

188. Maple, J.R., et al., Derivation of Class-Ii Force-Fields .1. Methodology and Quantum Force-Field for the Alkyl Functional-Group and Alkane Molecules. Journal of Computational Chemistry, 1994. **15**(2): p. 162-182.

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

190. Cieplak, P., et al., Polarization effects in molecular mechanical force fields. J. Phys. Condens.Matter., 2009. **21**(33): p. 333102 1-21

191. Stern, H.A., et al., Combined Fluctuating Charge and Polarizable Dipole Models: Application to a Five-Site Water Potential Function. Journal of Chemical Physics, 2001. **115**(5): p. 2237-2251.

192. Anisimov, V.M., et al., Polarizable empirical force field for the primary and secondary alcohol series based on the classical drude model. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 1927-1946.

193. Jiao, D., et al., Calculation of protein-ligand binding free energy by using a polarizable potential. Proc. Nat. Acad. Sci. USA, 2008. **105**(17): p. 6290-6295. PMCID: PMC2359813.

194. Zhang, J., Y. Shi, and P. Ren, Polarizable Force Fields for Scoring Protein–Ligand Interactions, in *Protein-Ligand Interactions*2012, Wiley-VCH Verlag GmbH & Co. KGaA. p. 99-120.

195. Harder, E., A.D. MacKerell, and B. Roux, Many-Body Polarization Effects and the Membrane Dipole Potential. Journal of the American Chemical Society, 2009. **131**(8): p. 2760-2761.

196. Rappé, A.K. and W.A. Goddard III, Charge Equilibration for Molecular Dynamics Simulations. Journal of Physical Chemistry, 1991. **95**: p. 3358-3363.

197. Rick, S.W., S.J. Stuart, and B.J. Berne, Dynamical Fluctuating Charge Force-Fields - Application to Liquid Water. Journal of Chemical Physics, 1994. **101**(7): p. 6141-6156.

198. Ando, K., A stable fluctuating-charge polarizable model for molecular dynamics simulations: Application to aqueous electron transfers. Journal of Chemical Physics, 2001. **115**(11): p. 5228-5237.

199. Yoshii, N., et al., A molecular-dynamics study of the equation of state of water using a fluctuating-charge model. Chemical Physics Letters, 2000. **317**(3-5): p. 414-420.

200. Patel, S. and C.L. Brooks, Fluctuating charge force fields: Recent developments and applications from small molecules to macromolecular biological systems. Molecular Simulation, 2006. **32**(3-4): p. 231-249.

201. Bauer, B.A. and S. Patel, Recent applications and developments of charge equilibration force fields for modeling dynamical charges in classical molecular dynamics simulations. Theoretical Chemistry Accounts, 2012. **131**(3).

202. Sprik, M. and M.L. Klein, A polarizable model for water using distributed charge sites. Journal of Chemical Physics, 1988. **89**(12): p. 7556-7560.

203. van Maaren, P.J. and D. van der Spoel, Molecular dynamics simulations of water with novel shell-model potentials. Journal of Physical Chemistry B, 2001. **105**(13): p. 2618-2626.

204. Yu, H.B., T. Hansson, and W.F. van Gunsteren, Development of a simple, self-consistent polarizable model for liquid water. Journal of Chemical Physics, 2003. **118**(1): p. 221-234.

205. Anisimov, V.M., et al., Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. Journal of Chemical Theory and Computation, 2005. **1**(1): p. 153-168.

206. Zhu, X., P.E. Lopes, and A.D. Mackerell, Jr., Recent Developments and Applications of the CHARMM force fields. Wiley Interdiscip Rev Comput Mol Sci, 2012. **2**(1): p. 167-185.

207. Stern, H.A., et al., Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from ab Initio Quantum Chemistry. Journal of Physical Chemistry B, 1999. **103**: p. 4730-4737.

208. Dang, L.X. and T.M. Chang, Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. Journal of Chemical Physics, 1997. **106**(19): p. 8149-8159.

209. Brdarski, S., P.-O. Åstrand, and G. Karlström, The Inclusion of Electron Correlation in Intermolecular Potentials: Applications to the Formamide Dimer and Liquid Formamide. Theoretical Chemistry Accounts, 2000. **105**(1): p. 7-14.

210. Thole, B.T., Molecular Polarizabilities Calculated with a Modified Dipole Interaction. Chemical Physics, 1981. **59**: p. 341-350.

211. Faerman, C.H. and S.L. Price, A Transferable Distributed Multipole Model for the Electrostatic Interactions of Peptides and Amides. Journal of the American Chemical Society, 1990. **112**: p. 4915-4926.

212. Price, S.L., C.H. Faerman, and C.W. Murray, Toward Accurate Transferable Electrostatic Models for Polypeptides: A Distributed Multipole Study of Blocked Amino Acid Residue Charge Distributions. Journal of Computational Chemistry, 1991. **12**(10): p. 1187-1197.

213. Mooij, W.T.M. and F.J.J. Leusen, Multipoles versus charges in the 1999 crystal structure prediction test. Physical Chemistry Chemical Physics, 2001. **3**(22): p. 5063-5066.

214. Lommerse, J.P.M., et al., A test of crystal structure prediction of small organic molecules. Acta Crystallogr. Sect. B Struct. Sci., 2000. **56**: p. 697-714.

215. Mahoney, M.W. and W.L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. Journal of Chemical Physics, 2000. **112**(20): p. 8910-8922.

216. MacKerell, A.D., M. Feig, and C.L. Brooks, Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society, 2004. **126**(3): p. 698-699.

217. Halgren, T.A., Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. Journal of the American Chemical Society, 1992. **114**: p. 7827-7843.

218. Stone, A.J., The Theory of Intermolecular Forces. International Series of Monographs on Chemistry1996, Oxford: Oxford University Press.

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

220. Ponder, J.W., TINKER: Software Tools for Molecular Design v5.0. Washington University Medical School http://dasher.wustl.edu/tinker/, 2010.

221. Pearlman, D.A., et al., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. Computer Physics Communications, 1995. **91**(1-3): p. 1-41.

222. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09. Gaussian, Inc., Wallingford CT, 2009.

223. Kong, J., et al., Q-chem 2.0: A high-performance ab initio electronic structure program package. Journal of Computational Chemistry, 2000. **21**(16): p. 1532-1548.

224. Stone, A.J., GDMA, 1998, Cambridge University Technical Services: Cambridge, England.

225. DiStasio, R.A., Y.S. Jung, and M. Head-Gordon, A resolution-of-the-identity implementation of the local triatomics-in-molecules model for second-order Moller-Plesset perturbation theory with application to alanine tetrapeptide conformational energies. Journal of Chemical Theory and Computation, 2005. **1**(5): p. 862-876.

226. Verlet, L., Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. Physical Review, 1967. **159**(1): p. 98-&.

227. Martyna, G.J., et al., Explicit Reversible Integrators for Extended Systems Dynamics. Molecular Physics, 1996. **87**: p. 1117-1157.

228. Roux, B., The Calculation of the Potential of Mean Force Using Computer-Simulations. Computer Physics Communications, 1995. **91**(1-3): p. 275-282.

229. Kumar, S., et al., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. Journal of Computational Chemistry, 1992. **13**(8): p. 1011-1021.

230. Kumar, S., et al., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. Journal of Computational Chemistry, 1995. **16**(11): p. 1339-1350.

231. Ting, D., et al., Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. PLoS Comput. Biol., 2010. **6**(4).

232. Patriksson, A. and D. van der Spoel, A temperature predictor for parallel tempering simulations. Physical Chemistry Chemical Physics, 2008. **10**(15): p. 2073-2077.

233. Sugita, Y. and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.

234. Best, R.B. and G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. Journal of Physical Chemistry B, 2009. **113**(26): p. 9004-15.

235. Jelsch, C., et al., Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. Proc. Nat. Acad. Sci. USA, 2000. **97**(7): p. 3171-6.

236. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, Designing a 20-residue protein. Natural Structural Biology, 2002. **9**(6): p. 425-30.

237. McKnight, C.J., P.T. Matsudaira, and P.S. Kim, NMR structure of the 35-residue villin headpiece subdomain. Natural Structural Biology, 1997. **4**(3): p. 180-4.

238. Vijaykumar, S., C.E. Bugg, and W.J. Cook, Structure of Ubiquitin Refined at 1.8 a Resolution. Journal of Molecular Biology, 1987. **194**(3): p. 531-544.

239. Ulmer, T.S., et al., Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. Journal of the American Chemical Society, 2003. **125**(30): p. 9179-9191.

240. Ko, T.P., et al., The refined crystal structure of an eel pout type III antifreeze protein RD1 at 0.62-A resolution reveals structural microheterogeneity of protein and solvation. Biophysical Journal, 2003. **84**(2): p. 1228-1237.

241. Parkin, S., B. Rupp, and H. Hope, Structure of bovine pancreatic trypsin inhibitor at 125 K: Definition of carboxyl-terminal residues Gly57 and Ala58. Acta Crystallogr. D Biol. Crystallogr., 1996. **52**: p. 18-29.

242. Szep, S., et al., Structural coupling between FKBP12 and buried water. Proteins., 2009. **74**(3): p. 603-611.

243. Young, A.C.M., et al., Comparison of Radiation-Induced Decay and Structure Refinement from X-Ray Data Collected from Lysozyme Crystals at Low and Ambient-Temperatures. Journal of Applied Crystallography, 1993. **26**: p. 309-319.

244. Beachy, M.D., et al., Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. Journal of the American Chemical Society, 1997. **119**(25): p. 5908-5920.

245. Hornak, V., et al., Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins, 2006. **65**(3): p. 712-725.

246. Hegefeld, W.A., et al., Helix Formation in a Pentapeptide Experiment and Force-field Dependent Dynamics. Journal of Physical Chemistry A, 2010. **114**(47): p. 12391-12402.

247. Best, R.B., et al., Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of alpha-Helix and beta-Hairpin Formation. Biophysical Journal, 2012. **103**(5): p. 1045-1051.

248. Best, R.B., et al., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. Journal of Chemical Theory and Computation, 2012. **8**(9): p. 3257-3273.

249. Best, R.B. and J. Mittal, Balance between a and beta Structures in Ab Initio Protein Folding. Journal of Physical Chemistry B, 2010. **114**(26): p. 8790-8798.

250. Best, R.B., N.V. Buchete, and G. Hummer, Are current molecular dynamics force fields too helical? Biophysical Journal, 2008. **95**(1): p. L7-L9.

251. Lindorff-Larsen, K., et al., Systematic Validation of Protein Force Fields against Experimental Data. Plos One, 2012. **7**(2).

252. Aliev, A.E. and D. Courtier-Murias, Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. Journal of Physical Chemistry B, 2010. **114**(38): p. 12358-12375.

253. Shi, Z.S., et al., Polyproline II structure in a sequence of seven alanine residues. Proc. Nat. Acad. Sci. USA, 2002. **99**(14): p. 9190-9195.

254. Chen, K., et al., Spin relaxation enhancement confirms dominance of extended conformations in short alanine peptides. Angewandte Chemie International Edition, 2007. **46**(47): p. 9036-9039.

255. Schweitzer-Stenner, R. and T.J. Measey, The alanine-rich XAO peptide adopts a heterogeneous population, including turn-like and polyproline II conformations. Proc. Nat. Acad. Sci. USA, 2007. **104**(16): p. 6649-6654.

256. Schweitzer-Stenner, R., et al., Conformational manifold of alpha-aminoisobutyric acid (Aib) containing alanine-based tripeptides in aqueous solution explored by vibrational spectroscopy, electronic circular dichroism spectroscopy, and molecular dynamics simulations. Journal of the American Chemical Society, 2007. **129**(43): p. 13095-13109.

257. Mukhopadhyay, P., G. Zuber, and D.N. Beratan, Characterizing Aqueous Solution Conformations of a Peptide Backbone Using Raman Optical Activity Computations. Biophysical Journal, 2008. **95**(12): p. 5574-5586.

258. Marqusee, S., V.H. Robbins, and R.L. Baldwin, Unusually Stable Helix Formation in Short Alanine-Based Peptides. Proc. Nat. Acad. Sci. USA, 1989. **86**(14): p. 5286-5290.

259. Spek, E.J., et al., Alanine is an intrinsic alpha-helix stabilizing amino acid. Journal of the American Chemical Society, 1999. **121**(23): p. 5571-5572.

260. Shalongo, W., L. Dugad, and E. Stellwagen, Distribution of Helicity within the Model Peptide Acetyl(Aaqaa)(3)Amide. Journal of the American Chemical Society, 1994. **116**(18): p. 8288-8293.

261. Buck, M., et al., Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. Biochemistry, 1995. **34**(12): p. 4041-55.

262. Tjandra, N., et al., Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. Journal of the American Chemical Society, 1995. **117**(50): p. 12562-12566.

263. Lipari, G. and A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. Journal of the American Chemical Society, 1982. **104**(17): p. 4559-4570.

264. Lipari, G. and A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.

265. Prompers, J.J. and R. Bruschweiler, General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. Journal of the American Chemical Society, 2002. **124**(16): p. 4522-4534.

266. Smith, L.J., et al., Analysis of phi and chi 1 torsion angles for hen lysozyme in solution from 1H NMR spin-spin coupling constants. Biochemistry, 1991. **30**(4): p. 986-96.

267. Grimshaw, S.B., Novel approaches to characterizing native and denatured proteins by NMR. Doctoral Thesis, University of Oxford, 1999.

268. Miclet, E., J. Boisbouvier, and A. Bax, Measurement of eight scalar and dipolar couplings for methine-methylene pairs in proteins and nucleic acids. Journal of Biomolecular NMR, 2005. **31**(3): p. 201-16.

269. Berndt, K.D., et al., Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. Journal of Molecular Biology, 1992. **227**(3): p. 757-75.

270. Lindorff-Larsen, K., et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins., 2010. **78**(8): p. 1950-8.

271. Mackerell, A.D., M. Feig, and C.L. Brooks, Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry, 2004. **25**(11): p. 1400-1415.

272. Jorgensen, W.L., The many roles of computation in drug discovery. Science, 2004. **303**(5665): p. 1813-1818.

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

274. Jorgensen, W.L., Interactions between amides in solution and the thermodynamics of weak binding. Journal of American Chemical Society, 1989. **111**: p. 2.

275. Katz, B.A., et al., Episelection - Novel K-I-Similar-to Nanomolar Inhibitors of Serine Proteases Selected by Binding or Chemistry on an Enzyme Surface. Biochemistry, 1995. **34**(26): p. 8264-8280.

276. Ponder, J.W. and D.A. Case, Force fields for protein simulations. Protein Simulations, 2003. **66**: p. 27-+.

277. Ren, P. and J.W. Ponder, Tinker polarizable atomic multipole force field for proteins. Abstracts of Papers of the American Chemical Society, 2002. **224**: p. U473-U473.

278. Beutler, T.C., et al., Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. Chemical Physics Letters, 1994. **222**(6): p. 529-539.

279. Bennett, C., Efficient Estimation of Free Energy Differences From Monte Carlo Data. Journal of Computational Physics, 1976. **22**: p. 245-268.

280. Schwarzl, S.M., et al., Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? Journal of Computational Chemistry, 2002. **23**(12): p. 1143-1149.

281. Grater, F., et al., Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. Journal of Physical Chemistry B, 2005. **109**(20): p. 10474-10483.

282. Ota, N., et al., Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: Relative free energy of binding of trypsin to benzamidine and benzylamine. Proteins-Structure Function and Genetics, 1999. **37**(4): p. 641-653.

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

284. Katz, B.A., et al., A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. Journal of Molecular Biology, 2001. **307**(5): p. 1451-1486.

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

286. Essex, J.W., et al., Monte Carlo simulations for proteins: Binding affinities for trypsin-benzamidine complexes via free-energy perturbations. Journal of Physical Chemistry B, 1997. **101**(46): p. 9663-9669.

287. Quillin, M.L., et al., Size versus polarizability in protein-ligand interactions: Binding of noble gases within engineered cavities in phage T4 lysozyme. Journal of Molecular Biology, 2000. **302**(4): p. 955-977.

288. Mann, A., In The Practice of Medicinal Chemistry, Second Edition, ed. C.G. Wermuth2003, London: Academic Press. 233.

289. Lauri, G. and P.A. Bartlett, Caveat - a Program to Facilitate the Design of Organic-Molecules. Journal of Computer-Aided Molecular Design, 1994. **8**(1): p. 51-66.

290. Page, M.I. and W.P. Jencks, Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and Chelate Effect. Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(8): p. 1678-1683.

291. Khan, A.R., et al., Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. Biochemistry, 1998. **37**(48): p. 16839-16845.

292. Widlanski, T., S.L. Bender, and J.R. Knowles, Dehydroquinate Synthase - a Sheep in Wolfs Clothing. Journal of the American Chemical Society, 1989. **111**(6): p. 2299-2300.

293. Bartlett, P.A. and W.W. Smith, Macrocyclic inhibitors of penicillopepsin. 3. Design, synthesis, and evaluation of an inhibitor bridged between P2 and P1 '. Journal of the American Chemical Society, 1998. **120**(19): p. 4622-4628.

294. Davidson, J.P., et al., Calorimetric and structural studies of 1,2,3-trisubstituted cyclopropanes as conformationally constrained peptide inhibitors of Src SH2 domain binding. Journal of the American Chemical Society, 2002. **124**(2): p. 205-215.

295. Davidson, J.P. and S.F. Martin, Use of 1,2,3-trisubstituted cyclopropanes as conformationally constrained peptide mimics in SH2 antagonists. Tetrahedron Letters, 2000. **41**(49): p. 9459-9464.

296. Verdine, G.L., et al., High-affinity muopioid receptor ligands discovered by the screening of an exhaustively stereodiversified library of 1,5-enediols. Journal of the American Chemical Society, 2002. **124**(45): p. 13352-13353.

297. Udugamasooriya, D.G. and M.R. Spaller, Conformational constraint in protein ligand design and the inconsistency of binding entropy. Biopolymers, 2008. **89**(8): p. 653-667.

298. DeLorbe, J.E., et al., Thermodynamic and Structural Effects of Macrocyclic Constraints in Protein-Ligand Interactions. Acs Medicinal Chemistry Letters, 2010. **1**(8): p. 448-452.

299. Martin, S.F., Preorganization in biological systems: Are conformational constraints worth the energy? Pure and Applied Chemistry, 2007. **79**(2): p. 193-200.

300. Martin, S.F., et al., Ligand preorganization may be accompanied by entropic penalties in protein-ligand interactions. Angewandte Chemie-International Edition, 2006. **45**(41): p. 6830-6835.

301. DeLorbe, J.E., et al., Thermodynamic and structural effects of conformational constraints in protein-ligand interactions. Entropic paradoxy associated with ligand preorganization. Journal of the American Chemical Society, 2009. **131**(46): p. 16758-70.

302. Liu, L. and Q.X. Guo, Isokinetic relationship, isoequilibrium relationship, and enthalpy-entropy compensation. Chemical Reviews, 2001. **101**(3): p. 673-695.

303. Gilli, P., et al., Enthalpy-Entropy Compensation in Drug-Receptor Binding. Journal of Physical Chemistry, 1994. **98**(5): p. 1515-1518.

304. Dunitz, J.D., Win Some, Lose Some - Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. Chemistry & Biology, 1995. **2**(11): p. 709-712.

305. Cooper, A., et al., Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. Biophysical Chemistry, 2001. **93**(2-3): p. 215-230.

306. Breslauer, K.J., et al., Enthalpy Entropy Compensations in Drug DNA-Binding Studies. Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(24): p. 8922-8926.

307. Grunwald, E. and C. Steel, Solvent Reorganization and Thermodynamic Enthalpy-Entropy Compensation. Journal of the American Chemical Society, 1995. **117**(21): p. 5687-5692.

308. Rekharsky, M. and Y. Inoue, Chiral recognition thermodynamics of beta-cyclodextrin: The thermodynamic origin of enantioselectivity and the enthalpy-entropy compensation effect. Journal of the American Chemical Society, 2000. **122**(18): p. 4418-4435.

309. Chen, W., C.E. Chang, and M.K. Gilson, Calculation of cyclodextrin binding affinities: Energy, entropy, and implications for drug design. Biophysical Journal, 2004. **87**(5): p. 3035-3049.

310. Al Omari, M.M., et al., Thermodynamic enthalpy-entropy compensation effects observed in the complexation of basic drug substrates with beta-cyclodextrin. Journal of Inclusion Phenomena and Macrocyclic Chemistry, 2007. **57**(1-4): p. 379-384.

311. Cornish-Bowden, A., Enthalpy-entropy compensation: a phantom phenomenon. Journal of Biosciences, 2002. **27**(2): p. 121-126.

312. Krug, R.R., W.G. Hunter, and R.A. Grieger, Enthalpy-Entropy Compensation .1. Some Fundamental Statistical Problems Associated with Analysis of Vant Hoff and Arrhenius Data. Journal of Physical Chemistry, 1976. **80**(21): p. 2335-2341.

313. Leung, D.H., R.G. Bergman, and K.N. Raymond, Enthalpy-entropy compensation reveals solvent reorganization as a driving force for supramolecular encapsulation in water. Journal of the American Chemical Society, 2008. **130**(9): p. 2798-2805.

314. Sharp, K., Entropy-enthalpy compensation: Fact or artifact? Protein Science, 2001. **10**(3): p. 661-667.

315. Reynolds, C.H. and M.K. Holloway, Thermodynamics of Ligand Binding and Efficiency. Medicinal Chemistry Letters, 2011. **2**: p. 433.

316. Williams, D.H., et al., Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. Angewandte Chemie-International Edition, 2004. **43**(48): p. 6596-6616.

317. Exner, O., Statistics of Enthalpy-Entropy Relationship .5. Enthalpy-Entropy Relationship in Organic Reactions. Collection of Czechoslovak Chemical Communications, 1975. **40**(9): p. 2762-2780.

318. Exner, O. and V. Beranek, Statistic of Enthalpy-Entropy Relationship .2. General Case. Collection of Czechoslovak Chemical Communications, 1973. **38**(3): p. 781-798.

319. Schowen, R.L., Isergonic Relations and Their Significance for Catalysis. Journal of Pharmaceutical Sciences, 1967. **56**(8): p. 931-&.

320. Linert, W. and R.F. Jameson, The Isokinetic Relationship. Chemical Society Reviews, 1989. **18**(4): p. 477-505.

321. Zhou, H.X. and M.K. Gilson, Theory of Free Energy and Entropy in Noncovalent Binding. Chemical Reviews, 2009. **109**(9): p. 4092-4107.

322. Bradshaw, J.M. and G. Waksman, Molecular recognition by SH2 domains. Protein Simulations, 2002. **61**: p. 161-210.

323. Ladbury, J.E., Isothermal titration calorimetry: application to structure-based drug design. Thermochimica Acta, 2001. **380**(2): p. 209-215.

324. Sturtevant, J.M., Heat capacity and entropy changes in processes involving proteins. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(6): p. 2236-40.

325. Brandsdal, B.O., et al., Free energy calculations and ligand binding. Protein Simulations, 2003. **66**: p. 123-+.

326. Pande, V.S., et al., Alchemical free energy methods for drug discovery: progress and challenges. Current Opinion in Structural Biology, 2011. **21**(2): p. 150-160.

327. Shirts, M.R., D.L. Mobley, and J.D. Chodera, Alchemical free energy calculations: ready for prime time? Annual Reports in Computational Chemistry, 2007. **3**: p. 41-57.

328. Andricioaei, I. and M. Karplus, On the calculation of entropy from covariance matrices of the atomic fluctuations. Journal of Chemical Physics, 2001. **115**(14): p. 6289-6292.

329. Brooks, B.R., D. Janezic, and M. Karplus, Harmonic-Analysis of Large Systems .1. Methodology. Journal of Computational Chemistry, 1995. **16**(12): p. 1522-1542.

330. Gohlke, H. and D.A. Case, Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. Journal of Computational Chemistry, 2004. **25**(2): p. 238-250.

331. Tidor, B. and M. Karplus, The Contribution of Vibrational Entropy to Molecular Association - the Dimerization of Insulin. Journal of Molecular Biology, 1994. **238**(3): p. 405-414.

332. Case, D.A., Normal-Mode Analysis of Protein Dynamics. Current Opinion in Structural Biology, 1994. **4**(2): p. 285-290.

333. Bohm, H.J., Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. Journal of Computer-Aided Molecular Design, 1998. **12**(4): p. 309-323.

334. Jackson, R.N., M. Kulharia, and R.S. Goody, Information Theory-Based Scoring Function for the Structure-Based Prediction of Protein-Ligand Binding Affinity. Journal of Chemical Information and Modeling, 2008. **48**(10): p. 1990-1998.

335. Muegge, I. and Y.C. Martin, A general and fast scoring function for protein-ligand interactions: A simplified potential approach. Journal of Medicinal Chemistry, 1999. **42**(5): p. 791-804.

336. Carlsson, J. and J. Aqvist, Absolute and relative entropies from computer simulation with applications to ligand binding. Abstracts of Papers of the American Chemical Society, 2005. **229**: p. U783-U784.

337. Hermans, J. and L. Wang, 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, 1997. **119**(11): p. 2707-2714.

338. Luo, H.B. and K. Sharp, On the calculation of absolute macromolecular binding free energies. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(16): p. 10399-10404.

339. Luo, R. and M.K. Gilson, Synthetic adenine receptors: Direct calculation of binding affinity and entropy. Journal of the American Chemical Society, 2000. **122**(12): p. 2934-2937.

340. Swanson, J.M.J., R.H. Henchman, and J.A. McCammon, Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophysical Journal, 2004. **86**(1): p. 67-74.

341. Baginski, M., F. Fogolari, and J.M. Briggs, Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. Journal of Molecular Biology, 1997. **274**(2): p. 253-267.

342. Olano, L.R. and S.W. Rick, Hydration free energies and entropies for water in protein interiors. Journal of the American Chemical Society, 2004. **126**(25): p. 7991-8000.

343. Warshel, A. and N. Singh, A comprehensive examination of the contributions to the binding entropy of protein-ligand complexes. Proteins-Structure Function and Bioinformatics, 2010. **78**(7): p. 1724-1735.

344. Ponder, J.W. and D.A. Case, Force fields for protein simulations, in *Protein Simulations*2003, Academic Press Inc: San Diego. p. 27-+.

345. Grossfield, A., P.Y. Ren, and J.W. Ponder, Single ion solvation thermodynamics from simulations. Biophysical Journal, 2003. **84**(2): p. 94A-94A.

346. Wu, J.C., et al., Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. Journal of Chemical Theory and Computation, 2010. **6**(7): p. 2059-2070.

347. Shi, Y., et al., Multipole Electrostatics in Hydration Free Energy Calculations. Journal of Computational Chemistry, 2011. **32**(5): p. 967-977.

348. Pappu, R.V., A.N. Drozdov, and A. Grossfield, Role of solvent in determining conformational preferences of alanine dipeptide in water. Journal of the American Chemical Society, 2004. **126**(8): p. 2574-2581.

349. Kubo, M.M., E. Gallicchio, and R.M. Levy, Thermodynamic decomposition of hydration free energies by computer simulation: Application to amines, oxides, and sulfides. Journal of Physical Chemistry B, 1997. **101**(49): p. 10527-10534.

350. Smith, D.E. and A.D.J. Haymet, Free-Energy, Entropy, and Internal Energy of Hydrophobic Interactions - Computer-Simulations. Journal of Chemical Physics, 1993. **98**(8): p. 6445-6454.

351. Tsunekawa, N., et al., A study of water-water interactions in hydrophobic association by a molecular dynamics simulation with an optimized umbrella sampling method. Journal of Chemical Physics, 2002. **116**(15): p. 6725-6730.

352. Lu, N., D.A. Kofke, and T.B. Woolf, Staging is more important than perturbation method for computation of enthalpy and entropy changes in complex systems. Journal of Physical Chemistry B, 2003. **107**(23): p. 5598-5611.

353. Ren, P., C. Wu, and J.W. Ponder, Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. Journal of Chemical Theory and Computation, 2011. **7**(10): p. 3027-3034.

354. Stone, A.J. and M. Alderton, Distributed Multipole Analysis - Methods and Applications. Molecular Physics, 1985. **56**(5): p. 1047-1064.

355. Yoda, T., Y. Sugita, and Y. Okamoto, Comparisons of force fields for proteins by generalized-ensemble simulations. Chemical Physics Letters, 2004. **386**(4-6): p. 460-467.

356. Zhou, R.H., Trp-cage: Folding free energy landscape in explicit water. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(23): p. 13280-13285.

357. Boresch, S. and M. Karplus, The role of bonded terms in free energy simulations: 1. Theoretical analysis. Journal of Physical Chemistry A, 1999. **103**(1): p. 103-118.

358. Boresch, S. and M. Karplus, The role of bonded terms in free energy simulations. 2. Calculation of their influence on free energy differences of solvation. Journal of Physical Chemistry A, 1999. **103**(1): p. 119-136.

359. Allen, M.P. and D.J. Tildesley, Computer simulation of liquids1989, Oxford: Clarendon Press. 193-195.

360. Chipot, C. and A. Pohorille, Free Energy Calculations. Theory and Applications in Chemistry and Biology. Springer Series in Chemical Physics, ed. C. Chipot and A. Pohorille2007: Springer.

361. Wyczalkowski, M.A., A. Vitalis, and R.V. Pappu, New Estimators for Calculating Solvation Entropy and Enthalpy and Comparative Assessments of Their Accuracy and Precision. Journal of Physical Chemistry B, 2010. **114**(24): p. 8166-8180.

362. Syme, N.R., et al., Comparison of Entropic Contributions to Binding in a "Hydrophilic" versus "Hydrophobic" Ligand-Protein Interaction. Journal of the American Chemical Society, 2010. **132**(25): p. 8682-8689.

363. Peter, C., et al., Estimating entropies from molecular dynamics simulations. Journal of Chemical Physics, 2004. **120**(6): p. 2652-2661.

364. Aqvist, J. and J. Carlsson, Absolute Hydration Entropies of Alkali Metal Ions from Molecular Dynamics Simulations. Journal of Physical Chemistry B, 2009. **113**(30): p. 10255-10260.

365. Lee, J.K., Statistical bioinformatics: a guide for life and biomedical science researchers2010, Hoboken, New Jersey: John Wiley & Sons, Inc.

366. Martin, S.F., et al., Binding of flexible and constrained ligands to the Grb2 SH2 domain: structural effects of ligand preorganization. Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 1101-1115.