

Copyright  
by  
Rui Qi  
2018

The Dissertation Committee for Rui Qi  
certifies that this is the approved version of the following dissertation:

## **Computational Studies of Protein-Ligand Recognition**

Committee:

Pengyu Ren, Supervisor

Kevin Dalby

Ron Elber

Hsin-Chih (Tim) Yeh

# Computational Studies of Protein-Ligand Recognition

by

Rui Qi

## DISSERTATION

Presented to the Faculty of the Graduate School of  
The University of Texas at Austin  
in Partial Fulfillment  
of the Requirements  
for the Degree of

## DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2018

For my parents.

## Acknowledgments

I would like to express sincere gratitude to my supervisor and my life-long friend Dr. Pengyu Ren. I could not complete this thesis and my Ph.D. without his continuous support. He shares his knowledge and experience with me in both science and life. I will always remember to think about the big picture first, ask questions, and keep a health work-life balance no matter where I am and what I am working on.

I would like to thank Dr. Ron Elber, Dr. Kevin Dalby, and Dr. Tim Yeh for serving on my doctoral committee. Dr. Elber has lectured many biophysics and statistical mechanics courses; and by taking those classes, I've learned essential concepts and theories that benefited me a lot in research. I am very grateful to Dr. Tim Yeh for his quick response to joining the committee and helpful suggestions on my dissertation. The Ren lab has been closely collaborating with Dr. Dalby in the inhibitors design for the protein kinase. I am grateful to have worked with him and grateful for his mentorship and his friendship.

I would like to thank people in Ren lab: Dr. Qiantao Wang, Dr. Changsheng Zhang, Dr. Chengwen Liu, and Zhifeng Jing for their great help

in my research. I would also like to thank Dr. Chenfeng He, Dr. Sara Cheng, Dr. David Bell, Dr. Xiaojia Mu, Matthew Harger, and Brandon Walker for many helpful technical discussions and I greatly enjoyed working with them. We are family and friends who concern and take care of each other. We will cherish this friendship and, in the future, offer help as much as we can whether we are in academia or industry.

Finally, I would like to express my profound gratitude to my family. My parents have been my greatest supporters during my Ph.D. studies. Although they probably do not understand my research, they offer constant love and encouragement to pursue my dreams. I thank Zexi Liang, my best friend, the fellow student in math and physics, volleyball captain, and my boyfriend for taking care of me when I was under depression and after surgery. I am grateful for their love and support, and I dedicate this thesis to them.

# Computational Studies of Protein-Ligand Recognition

Publication No. \_\_\_\_\_

Rui Qi, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Pengyu Ren

Molecular recognition between biomolecules and ligands is very specific in living cells. The functions of all biochemical processes and cell mechanisms are dependent upon complex but specific non-covalent intermolecular interactions. As essential building blocks in protein and nucleic acid, phosphate groups are commonly found in nucleic acids, proteins, and lipids. Nearly half of known proteins have been shown to interact with ligands containing a phosphate group. Binding of a phosphoryl group is fundamental to a range of biological processes including metabolism, biosynthesis, gene regulation, signal transduction, muscle contraction, and antibiotic resistance. Phosphorylation is one of the most common forms of reversible posttranslational modification of protein and, nearly 30% of all proteins are phosphorylated on at least one residue in cells. However, phosphate binding sites are less well defined and fundamental principles of why and how proteins recognize phosphate groups are not yet fully understood.

Molecular modeling is a common tool for studying biomolecular structure, dynamics, interaction and function. Due to the complex electrostatics, high concentration of ions and intricate interactions with environment, however, the modeling and designing of highly charged drug-like molecules and nucleic acid derivatives are extremely difficult. This thesis will focus on the highly charged phosphate, including its different protonation states, and energetic and thermodynamic driving forces behind protein-phosphate recognition. This thesis work will also discuss the development of more sophisticated computational models, AMOEBA+, that are necessary for a better understanding and prediction of the physical properties of small organic molecules.

Four projects will be discussed in this dissertation: two projects on force field development, and two on applying molecular dynamic simulations to understand biological processes. These projects have led to new insights into understanding of physical and chemical principles and mechanisms underlying highly protein-phosphate binding and nucleic acid stability. In addition, this thesis work will enhance the capability to develop and apply computational and theoretical frameworks to model, predict and design proteins, therapeutics, and diagnostic strategies targeting phosphates, phosphate-containing biomolecules.



# Table of Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>List of Figures</b>	<b>xiii</b>
<b>Chapter 1. Molecular Modeling for Biomolecules</b>	<b>1</b>
1.1 Polarizable Force Field . . . . .	2
1.2 AMOEBA . . . . .	4
1.2.1 Polarization Framework . . . . .	5
1.2.2 Parameterization Using ForceBalance . . . . .	8
1.2.3 AMOEBA Nucleic Acid and Ion Models . . . . .	9
1.3 Recent Applications of AMOEBA . . . . .	10
1.3.1 Small Molecules . . . . .	11
1.3.2 Ions, Ion Channels and Protein-Ions Binding . . . . .	11
1.3.3 Interaction with Electric Field . . . . .	12
1.3.4 QM/MM . . . . .	13
1.4 The Next Generation: AMOEBA+ . . . . .	13
1.4.1 S101x7 SAPT2+ Database . . . . .	15
1.4.2 Charge Penetration Model . . . . .	17
1.4.3 Many-Body Interactions . . . . .	18
1.5 Overview of Thesis Work . . . . .	22
<b>Chapter 2. Development of Advanced Polarizable Force Fields for Water and Organic Molecules</b>	<b>24</b>
2.1 General van der Waals potential for common organic molecules	26
2.1.1 Introduction . . . . .	26
2.1.2 Computational Methods . . . . .	30
2.1.2.1 Van der Waals functions and mixing rules . . . . .	30

2.1.2.2	QM Database for Model Training and Testing . . . . .	33
2.1.3	Results and Discussion . . . . .	35
2.1.3.1	Atom Types and Parameters . . . . .	35
2.1.3.2	Model Training Performance on Dimer Interactions	37
2.1.3.3	Validation on Stacked Nucleobase and Heterodimers	41
2.1.3.4	Further Development: VdW 2017 Model . . . . .	49
2.1.4	Conclustions . . . . .	50
2.2	United Polarizable Multipole Water Model for Molecular Me- chanics Simulations . . . . .	52
2.2.1	Introduction . . . . .	52
2.2.2	Methodology . . . . .	55
2.2.2.1	Parameterization Dataset . . . . .	55
2.2.2.2	Parameter Optimization . . . . .	56
2.2.3	Computational Details . . . . .	59
2.2.3.1	Parameterization Calculations . . . . .	59
2.2.3.2	Validation calculations . . . . .	60
2.2.4	Results and Discussion . . . . .	63
2.2.4.1	Optimized Parameters . . . . .	63
2.2.4.2	Fitted gas phase water properties . . . . .	65
2.2.4.3	Fitted liquid water properties . . . . .	69
2.2.4.4	Validation of uAMOEBA . . . . .	71
2.2.4.4.1	Radial Distribution Function . . . . .	73
2.2.4.4.2	O-O-O Angle Distribution . . . . .	75
2.2.4.4.3	Self-diffusion Coefficient and Viscosity	76
2.2.4.5	Comparison between Coarse-grained and All-atom AMOEBA . . . . .	78
2.2.4.5.1	Transferability Validation . . . . .	79
2.2.4.5.2	Computational Efficiency . . . . .	80
2.2.5	Conclusions . . . . .	81

<b>Chapter 3. Applications of AMOEBA to Protein-ligand Recognition</b>	<b>83</b>
3.1 Phosphate Binding Mode in Phosphate binding protein . . . . .	86
3.1.1 Introduction . . . . .	86
3.1.2 Computational Methods . . . . .	88
3.1.2.1 Quantum Mechanics and Molecular Mechanics . . . . .	88
3.1.2.2 Absolute Binding Free Energy Calculations . . . . .	90
3.1.3 Parameterization Strategy and Model development . . . . .	94
3.1.3.1 Parameterization of 1H and 2H Phosphate Models . . . . .	94
3.1.3.2 Validation of the Phosphate Models . . . . .	95
3.1.4 Results and Discussion . . . . .	96
3.1.4.1 One Hydrogen Phosphate is the Dominant Form . . . . .	96
3.1.4.1.1 Hydrogen Bond Distances Between Phosphate and PBPs . . . . .	98
3.1.4.1.2 Calculated Standard Binding Free Energies . . . . .	101
3.1.4.2 The Critical Effect of Buffer Solution . . . . .	102
3.1.4.2.1 Apparent Dissociate Constant $K_D^{app}$ . . . . .	102
3.1.4.2.2 Calculated apparent binding free energies . . . . .	103
3.1.5 Conclusions . . . . .	105
3.2 Calculating Binding Free Energy of Host-Guest system . . . . .	105
3.2.1 Introduction . . . . .	105
3.2.2 Computational Methods . . . . .	107
3.2.2.1 Orthogonal space random walk (OSRW) . . . . .	107
3.2.2.2 Recent Development of Enhanced Sampling . . . . .	108
3.2.2.3 Simulation System . . . . .	109
3.2.3 Results: Calculated Binding Free Energies . . . . .	111
3.2.4 Discussion . . . . .	116
3.2.4.1 Enthalpy-Entropy Decomposition . . . . .	116
3.2.4.2 Hydrogen Bonding Analysis . . . . .	118
3.2.4.3 Configurational Entropy . . . . .	119
3.2.4.4 Convergence of the BAR and OSRW . . . . .	123
3.2.5 Conclusions . . . . .	125

<b>Chapter 4. Conclusion</b>	<b>128</b>
<b>Bibliography</b>	<b>130</b>

## List of Figures

1.1	Density of liquid water over temperature range of 250-370K at atmospheric pressure . . . . .	8
1.2	Transition from A-DNA to B-DNA in ethanol/water solution as captured by AMOEBA simulations . . . . .	9
1.3	Schematic view of molecules in the S101 dataset . . . . .	15
1.4	Plots of multipole electrostatic energy . . . . .	17
1.5	Plot of (a) the $E_{3B}$ and (b) $E_{4B}$ calculated using the three polarization models and QM methods . . . . .	20
1.6	Plots of the $E_{3B}$ distance dependence calculated from three polarization models and QM methods . . . . .	22
2.1	Plots of homodimer vdW energies in noble gas systems . . . . .	31
2.2	Differences between heterodimer vdW energies in noble gas system given by different combinations of vdW functional forms and mixing rules ( ij/ij) compared to SAPT 2+ results. . . . .	32
2.3	Molecule structures in the S108x7 database. . . . .	33
2.4	VdW types and parameters ( $\sigma$ and $\varepsilon$ ) for H, C, N, O, P, S, F, Cl, and Br. . . . .	36
2.5	Plots of the vdW energy calculated by vdW2016. . . . .	37
2.6	Plot of the fitting vdW energy calculated by vdW2016 model compared to the SAPT2+/CBS/scaled energy. . . . .	39
2.7	Differences between vdW energies given by vdW2016 potential, AMOEBA09 parameters, compared to SAPT2+/CBS/scaled results in the S108x7 fitting dataset. . . . .	40
2.8	Plot of the vdW energy calculated using the vdW2016, AMBER FF14, and CHARMM 36 models compared to the SAPT0/jun-cc-pVDZ QM results of ten stacked base pairs across their rotational and translational configurations. . . . .	42
2.9	Plots of the AG:CT stacking vdW energy. . . . .	44
2.10	Plot of the vdW energy values of the testing set S36x7. . . . .	46
2.11	Plots of the vdW energy surfaces of selected pairs. . . . .	48

2.12	RMS error (in kcal/mol) of vdW2017 model comparing to the SAPT2+ components for S108x7 dimer set. . . . .	50
2.13	Parameters for uAMOEBA water model. . . . .	63
2.14	Gas phase monomer properties of the uAMOEBA, AMOEBA14 and iAMOEBA models compared with experiment, evaluated at the energy-minimized geometry. . . . .	65
2.15	Electrostatic potential plotted on the vdW surface, with blue representing 0.05 h and red -0.05 h. . . . .	66
2.16	Cluster energy of gas phase geometry-optimized clusters ranging from size 2-20. . . . .	68
2.17	Thermodynamic properties of uAMOEBA liquid water as a function of temperature. Error bars indicate one standard error. . . . .	69
2.18	The oxygen-oxygen RDF curves of the uAMOEBA water model, compared with experimentally derived RDFs. . . . .	72
2.19	The oxygen-hydrogen RDF curves of the uAMOEBA water model, compared with experimentally derived RDFs. . . . .	73
2.20	The O-O-O angle distributions of uAMOEBA, AMOEBA03, TIP4P-Ew, and TIP5P water models. . . . .	75
2.21	Comparison of experimental and simulated liquid data from different water models (T = 298.15 K, P = 1 atm). . . . .	76
2.22	Dimer equilibrium interaction energy between water and small molecules. . . . .	78
2.23	Efficiency test for the uAMOEBA water model. . . . .	80
3.1	Thermodynamic cycle for calculating the binding free energy of phosphate-protein binding. . . . .	90
3.2	Apparent dissociation constant KD Calculations without and with buffer effect at pH 4.5 and 8.5. The pKa value is 7.21. . . . .	93
3.3	Parameterization schema of 1H and 2H phosphate models. . . . .	94
3.4	Model compounds for amino acid-phosphate interactions extracted from PDB. . . . .	95
3.5	Performance of AMOEBA interaction energies of 120 model compounds for amino acid-phosphate dimers compared to RIMP2/aug-cc-pVTZ, SAPT0/jun-ccpVDZ and SAPT2+/aug-cc-pVDZ results. . . . .	96
3.6	Plot of 12 hydrogen bond distances (XY) between 1H phosphate and heavy atoms in wild type (top) and D56N mutant (bottom) PBP receptors. . . . .	98

3.7	Illustration of phosphate binding with Asp56 in wild-type PBPs	99
3.8	Calculated standard binding free energy (kcal/mol) of 1H/2H phosphate with PBPs or buffer ligands. . . . .	101
3.9	Calculated apparent binding free energy (kcal/mol) of phosphate with PBP in wild type and D56N mutant protein at pH 8.5 and 4.5 and 50 mM sodium acetate/Tris acetate. . . . .	103
3.10	Predicted binding free energy as a function of experimental binding free energy (in kcal/mol). . . . .	111
3.11	Host-guest binding free energies. . . . .	113
3.12	Model deviation from experiment. RMS energy difference, and AUE (Average Unsigned Error) are in kcal/mol. . . . .	115
3.13	Host-guest binding enthalpies and entropies (kcal/mol). $STD(\Delta H)$ is the uncertainty of enthalpy. . . . .	116
3.14	Analysis of hydrogen bond numbers for guests C7, C8 and C10.	118
3.15	Configurational entropy computed from quasiharmonic analysis. $\alpha S_{h(solution)}$ is 495.61 cal/mol/K. . . . .	121
3.16	Standard deviation of $F_\lambda$ as a function of for different coupling schemes. . . . .	123
3.17	Correlation between uncertainties of binding free energies and net charge for each system. . . . .	125

# Chapter 1

## Molecular Modeling for Biomolecules

Molecular modeling and simulation can be a powerful tool for quantitative understanding of the driving forces underlying molecular recognition,[87, 120] and accelerating drug discovery and guiding molecular design by predicting ligand interactions with biomolecular targets.[85, 140] Numerous potential energy methods have been proposed to compute binding free energy, increasing in complexity from empirical docking methods to quantum mechanics (QM) calculations.[85] Empirical docking methods[151] are frequently used for library screening and though they allow for fast calculation, they do not maintain high accuracy of the potential energy function, nor do they allow for sufficient sampling of binding conformations. QM calculations of binding free energy[10, 93, 196] are limited to small, predetermined binding sites. Bridges between docking methods and QM are semi-empirical force-field methods using Molecular Dynamics (MD) or Monte Carlo sampling schemes to generate many configurations and energies.[163, 202]

In force fields, the potential energy of the system is computed from the analytical functions of the atomic coordinates. Classical force fields such as AMBER,[50] CHARMM,[290] OPLS-AA,[254] or GROMOS[242] typically



represent intermolecular interactions by a van der Waals (vdW) term and electrostatic term depends on fixed point charges. This representation is computationally efficient and sufficiently accurate for many applications. However, the potential energy is limited by not capturing electrostatic responses to environmental stimulus, referred to as the polarization effect.[230] Additionally, modeling electrostatics as point charges neglects the intricate yet substantial effect of charge distribution,[313] which can be properly captured by higher order multipole moments.[247] Therefore, tremendous efforts have been made to develop advanced representations of electrostatics ranging from fluctuating charges,[225] Drude oscillators,[12, 167] up to fully polarizable force-fields such as AMOEBA (atomic multipole optimized energetics for biomolecular applications).[248, 265]

This chapter includes the current development and application of polarizable force field. It will cover an introduction to polarizable forcefield, the polarizable framework used in AMOEBA, an automatic approach for parameterization, nucleic acid and Ion model. Besides, it will include the applications of AMOEBA on small molecules, protein-ion binding, interaction with the electric field and hybrid QM/MM method. Last but not the least important, the overall structure and outline of the dissertation will be presented.

## 1.1 Polarizable Force Field

There has been much effort devoted to improving the potential-energy functions or force field (FF) used in MD simulations. It is believed in biology

that amino acid sequences carry the structure information, which then determines the function. The potential energy surface defines the physical driving forces underlying biomolecular structure and interactions. Force fields usually consist of several empirical energy terms including short-ranged bonded interactions and non-bonded interactions such as repulsion, dispersion and electrostatics. Electrostatics is both important and computationally expensive due to its long-range nature. To facilitate simulations of biomolecules with modest computational power, traditional force fields use fixed point charge placed at atomic centers to represent the electrostatic interactions. The limitations of the fixed point-charge force fields have been well recognized.[45, 182, 190, 220, 246] One significant approximation in traditional force fields is the omission of polarization, i.e. the response of the charge distribution to environment. This is problematic when applying the same set of charge parameters to different environments, such as aqueous solution, protein cavity, cell membrane and heterogeneous interfaces, where the charge distribution should change accordingly. Another approximation is the atom-centered point-charge model, whereas the realistic charge distribution should be smooth and anisotropic. To capture anisotropic features such as  $\pi$ -holes, lone pairs and  $\pi$ -bonding, it is necessary to adopt higher-order multipolar electrostatics models [95, 131, 154, 156, 288] and/or adding off-center sites.[95] The effect of having atomic multipoles beyond fixed charges is of the same magnitude as the effect of polarization, suggesting that both should be included in force field development.[131]

## 1.2 AMOEBA

Over the past decades, several polarizable force fields have been developed for biological systems, including AMBER,[43, 300] AMOEBA,[227, 232] CHARMM Drude, CHARMM fluctuating charge,[56, 258] SIBFA, GEM,[95] and ABEEM.[175, 326] As a physics-grounded force field and different from the simple fixed-point-charge (partial charge) force fields AMOEBA depicts molecular polarizability and electrostatic potential terms by using mutual atomic dipole-dipole induction along with permanent atomic point multipoles (monopole, dipole, and quadrupole).[265] These results in a more accurate description of molecular energetics in biological applications.

$$U = U_{bond} + U_{angle} + U_{torsion} + U_{(out-of-plane)} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \quad (1.1)$$

$$U_{bonded} = \sum k_r (r - r_{eq})^2 + \sum k_a (\theta - \theta_{eq})^2 + \sum \frac{V_n}{2} [1 + \cos(n\varphi - \varphi_n)] + \sum k_\chi \chi^2 \quad (1.2)$$

AMOEBA is the first general-purpose polarizable force field that has been utilized in MD simulations of protein-ligand binding and calculation of absolute and relative binding free energy.[17, 135, 136] AMOEBA potential energy is written as a sum of valence and nonbonded contributions (Equation 1.1). The first four terms describe the intramolecular valence interactions: bond stretching, angle bending, torsional rotation and out-of-plane torsion terms (Equation 1.1). The last three terms in Equation 1.3, Equation 1.4 and Equation 1.5 are the non-bonded van der Waals (vdW) and permanent

electrostatic ( $U_{ele}^{perm}$ ) and polarization contributions ( $U_{ele}^{ind}$ ).

$$U_{vdW} = \varepsilon_{ij} \left( \frac{1 + \delta}{\rho_{ij} + \delta} \right)^{n-m} \left( \frac{1 + \gamma}{\rho_{ij}^m + \gamma} - 2 \right) \quad (1.3)$$

$$U_{ele}^{perm} = \sum (M_i)^T T_{ij} M_j \quad (1.4)$$

$$U_{ele}^{ind} = -\frac{1}{2} \sum_i (\mu_i^{ind})^T E_i^{dir} \quad (1.5)$$

The vdW interaction is described by the buffered-14-7 vdW formula (Equation 1.3) is the potential well depth and is the ratio between the actual separation of i-j sites and the minimum energy distance, described as  $\rho_{ij} = R_{ij}/R_{ij}^0$ . [104] The electrostatic potential energy is evaluated from the permanent molecular dipole and quadrupole moments, in which  $M_j$  is the poly tensor of permanent multipoles (charge, dipole, and quadrupole) (Equation 1.4) and molecular isotropic dipole polarizability, which will be discussed in the AMOEBA polarization framework section. All electrostatic energy and force terms, including polarization, are calculated using the particle-mesh-Ewald (PME) approach. [73, 255]

### 1.2.1 Polarization Framework

Polarization is explicitly treated by mutual induction of dipoles at polarizable sites (located at atomic centers). A point dipole moment is induced at each polarizable site according to the electric field experienced by that site:

$$\mu_i^{ind} = \alpha_i (E_i^{dir} + E_i^{mut}) \quad (1.6)$$

where  $\alpha_i$  is the atomic polarizability on site  $i$ ;  $E_i^{dir}$  is the direct electric field generated by permanent multipoles of other sites;  $E_i^{mut}$  is the mutual field

generated by induced dipoles of other sites. The  $E_i^{dir}$  and  $E_i^{mut}$  are expressed as

$$E_i^{dir} = \sum T_{ij} M_j \quad (1.7)$$

$$E_i^{mut} = \sum_{i \neq j} T_{ij}^{11} \mu_j^{ind} \quad (1.8)$$

where the  $T_{ij}$  in Equation 1.7 is the multipole-multipole interaction matrix;  $M_j$  is the poly tensor of permanent multipoles; [289] In Equation 1.8,  $T_{ij}^{11}$  is the dipole-dipole interaction matrix and  $\mu_j^{ind}$  is the induced dipole moment of site j. The induced dipole on each polarizable site is solved iteratively to obtain the converged dipole values. With self-consistent field (SCF) converged induced dipole, the polarization energy can be obtained through

$$U_{ele}^{ind} = -\frac{1}{2} \sum_i (\mu_i^{ind})^T E_i^{dir} \quad (1.9)$$

To ensure the finite nature of the intermolecular induction effect, Thole used a damping scheme in which a smeared charge distribution replaces one of the point dipoles, and thus dipole interactions are damped. [279] As a result, the dipole interaction energy approaches a finite value instead of becoming infinite as the atomic separation approaches zero. Tholes scheme is very successful in reproducing dipole molecular polarizability tensors for a broad range of organic molecules using element-based isotropic atomic polarizabilities. [279, 289] This scheme has been adopted by AMOEBA force field to model polarization energy. [162, 289] The electric fields due to both the permanent multipoles and the induced dipoles are damped using the same function in the current

AMOEBA model. This is accomplished by modifying the interaction T matrices in the corresponding orders and higher order T matrices can be found in reference.[174]

$$T_\alpha = -\lambda_3 \frac{R_\alpha}{R^3} \quad (\alpha = x, y, z) \quad (1.10)$$

The form of  $\lambda_3$  that the current AMOEBA uses is:

$$\lambda_3(r) = 1 - e^{-aw^3(r)} \quad (1.11)$$

where  $u(r) = r_{ij}/(\alpha_i\alpha_j)^{\frac{1}{6}}$  is the scaled distance between sites i and j;  $r_{ij}$  and  $\alpha_i$  are the real distance and atomic polarizability, respectively. The factor a is the dimensionless width parameter of the smeared charge distribution, and effectively controls the damping strength. The universal damping factor was determined to be 0.39 for both the direct and mutual part in the current AMOEBA by considering the molecular polarizabilities and total associate energy of water clusters up to hexamers.[162]

## 1.2.2 Parameterization Using ForceBalance

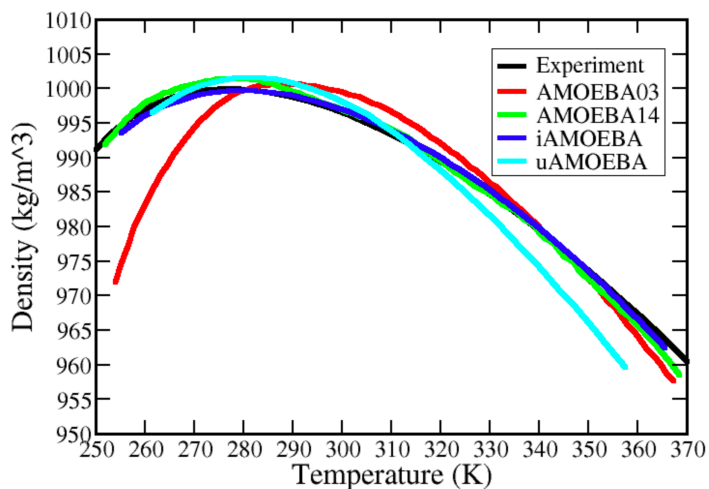


Figure 1.1: Density of liquid water over the temperature range of 250-370 K at atmospheric pressure. The data were reproduced from the original papers by using WebPlotDigitizer (<https://automeris.io/WebPlotDigitizer>).

An automatic and systematic approach for the parameterization of AMOEBA using the ForceBalance package[303] has also been explored. Overall the AMOEBA water model reparameterized (AMOEBA14)[162] using ForceBalance[303] better reproduces high-level quantum mechanical (QM) data and experimental condensed-phase properties compared to the original AMOEBA03. Variations of the functional form were devised to improve the computational speed, including the direct polarization (iAMOEBA)[303] and united atom models (uAMOEBA).[235] Both iAMOEBA and uAMOEBA, parameterized using ForceBalance, have comparable accuracy to AMOEBA03 for predicting gas-phase and liquid properties. As an example, the liquid densities over a

wide temperature range predicted by different water models are compared in Figure 1.1.

### 1.2.3 AMOEBA Nucleic Acid and Ion Models

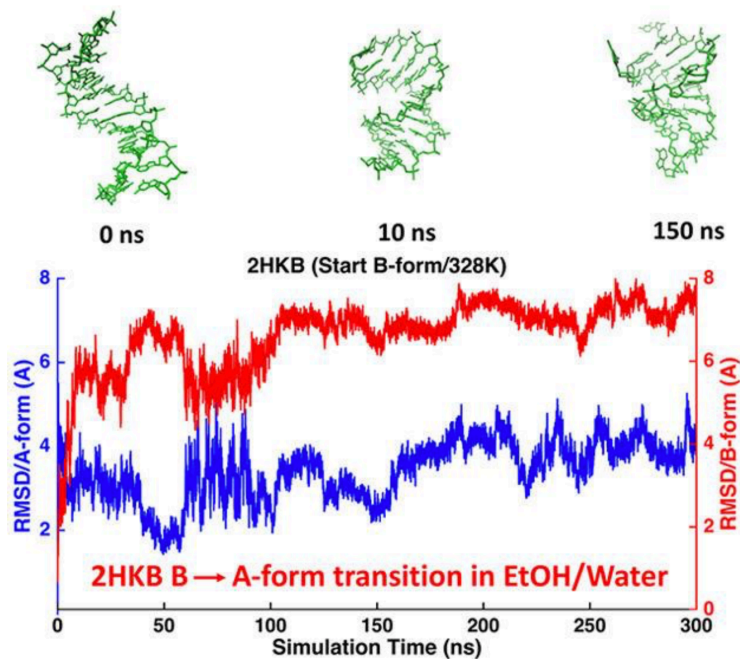


Figure 1.2: Transition from A-DNA to B-DNA in ethanol/water solution as captured by AMOEBA simulations.<sup>57</sup>

Recently, Zhang et al. developed the AMOEBA force field for DNA and RNA.<sup>[335]</sup> The force field was extensively validated through 35 microseconds of MD simulations. The simulated solution and crystal structures of DNA duplexes, RNA duplexes, and hairpins agree with NMR structures with  $RMSDs < 2.0\text{\AA}$ . Notably, the interconversion between A- and B-form DNAs



was observed in ethanol-water mixtures, (Figure 1.2) indicating a balanced description of the stabilities of different forms. Clavagura and coworkers developed the AMOEBA force field for Fe(II) and the heme cofactor in ferrous and ferric form.[263, 319] The parameters were validated for energy calculation of larger clusters and MD simulations of cytochromes, showing good agreement with DFT and NMR data. To match the energy components from ab initio calculations, Xia et al. incorporated an explicit charge-transfer term into the AMOEBA force field for Fe(III).[323] For the transition metal ions  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$ , AMOEBA-VB model was derived.[324] This model generates correct ion-ligand geometry and energetics for both QM gas-phase clusters and the coordination of the first solvation shell structure of their aqueous solutions. To better model the water ligand exchange rate around  $\text{Mg}^{2+}$ , Kurnikov and Kurnikova [159] treated the polarizability of AMOEBA water as variables according to the distance between water and  $\text{Mg}^{2+}$ .

### 1.3 Recent Applications of AMOEBA

AMOEBA force fields have been widely used to simulate water, organic molecules and proteins.[97, 186, 210, 248, 266, 316] Recent applications of AMOEBA on biomolecular systems include small molecules,[33] ions, [185, 260] protein-ion and protein-ligand binding,[121, 195, 218, 234] diffusion and permeation of small molecule,[184, 281] ion channels,[227] interaction with electric field, [72, 177] and hybrid QM/MM method.[177]

### 1.3.1 Small Molecules

Ionic liquid systems have received much attention because of their excellent thermal and electrochemical stability and good solvation properties.[148,272] Due to their charged nature, they are studied with MD simulations employing polarizable force fields. Busch et al. studied a highly concentrated aqueous solution of proline using neutron diffraction experiments and MD simulations employing AMOEBA and CHARMM force fields. Detailed structural analysis revealed the existence of proline dimers, which explains well the experimental observation. Compared to the non-polarizable CHARMM force field, the polarizable AMOEBA simulation gives better agreement with the EPSR fits to the diffraction data, which is similar to ab initio (CPAIMD) methods.[33]

### 1.3.2 Ions, Ion Channels and Protein-Ions Binding

The association of  $\text{Mg}^{2+}$  and  $\text{H}_2\text{PO}_4^-$  in water may give insights into our understanding of Mg and phosphate-containing biomolecules, e.g. DNA, RNA, and ATP. A recent simulation study shows that the binding free energy between  $\text{Mg}^{2+}$  and  $\text{H}_2\text{PO}_4^-$  determined by AMOEBA simulations (-2.23 kcal/mol) closely match the experimental value (-1.7 kcal/mol).[185] Another recent quantum calculation which used a mixed explicit/continuum solvent model gave a value of -3.3 kcal/mol, while non-polarizable force field over-predicted the binding free energy by a factor of ten.[260] These results again emphasized the importance of polarization in highly charged systems.

By explicitly introducing the multipole terms and polarization into the electrostatic potentials, the permeation free energy barrier of  $K^+$  through the gA channel is considerably reduced compared to the overestimated results obtained from the fixed-charge model. Moreover; the estimated maximum conductance, without any corrections, for both  $K^+$  and  $Na^+$  passing through the gA channel is much closer to the experimental results than any classical MD simulations, demonstrating the power of AMOEBA in investigating the membrane proteins.[227] Several recent studies have been focused on capturing the interactions of ions with proteins and nucleic acids. Using the AMOEBA polarizable-force field, many-body effects were shown to be essential for ion-selectivity in  $Mg^{2+}$  and  $Ca^{2+}$  protein complexes.[121]

### 1.3.3 Interaction with Electric Field

Electronic polarization is essential for modeling the interaction with the electric fields, such as in the simulations of THz spectra. AMOEBA force field was used to simulate the THz spectra of two zwitterionic amino acids (glycine and valine) in aqueous solution. After the detailed check of the THz spectral assignments, the mode-specific spectral decomposition into intramolecular solute motions, and solutewater cross-correlation modes, the authors found promising agreement of AMOEBA and ab initial molecular dynamics (AIMD) data for both systems.[72]

### 1.3.4 QM/MM

Polarizable force fields have been applied to the hybrid QM/MM method to better describe the environment of the QM region.[81, 103, 173, 178, 179, 332] The methods have been implemented in software interfaces, such as Gaussian/TINKER,[179] Psi4/TINKER[103], Q-Chem/CHARMM.[332] The use of polarizable force fields improves both ground-state energy and structure[103, 179] and excited-state spectral properties.[177, 179, 332] Loco et al. used QM/MM simulations with B3LYP and AMOEBA to study the color tuning in Carotenoid pigment-crustacyanin complexes.[177] It was found that the polarizable force field and MD simulations are necessary to obtain quantitative predictions of the spectrum. The high color tunability of the pigment-protein complex was explained by the bond length alternation in the long-chain carotenoids modulated by the dynamical protein environment.

## 1.4 The Next Generation: AMOEBA+

Classical molecular mechanics force fields typically model interatomic electrostatic interactions with point charges or multipole expansions, which can fail for atoms in close contact due to the lack of a description of penetration effects between their electron clouds. These short-range penetration effects can be significant and are essential for accurate modeling of intermolecular interactions. The current AMOEBA force field and most other widely used fixed-charge force fields utilize a less repulsive van der Waals potential to compensate the charge penetration contribution in short range. The prob-

lem with this approach, however, is that the resulting vdW parameters are less transferable.[209, 277] To improve the accuracy and transferability and mitigate the tedious parameterization process, the next-generation AMOEBA force field focuses on calibrating each energy component to high-level QM energy decomposition such as Symmetry-Adapted Perturbation Theory (SAPT) including electrostatic, induction/polarization, repulsion, and dispersion energies (Equation 1.12), and using automated optimization methods[303] for parameterization at large scales.

$$E_{SAPT2+} = E_{elst}^{SAPT} + E_{exch-r}^{SAPT} + E_{disp}^{SAPT} + E_{ind}^{SAPT} \quad (1.12)$$

For electrostatic interactions, the point charge or multipole model fails at close distances where electron clouds overlap. In this situation charge penetration (CP) effect must be considered. By utilizing empirical smearing functions either for charge-charge interactions only[306] or higher order multipoles,[238] the charge-penetration correction can be accurately captured. For polarization, the Thole damping function used in AMOEBA[232] was improved to better capture the explicit many-body interactions for a range of molecules at different intermolecular distances.[174] The polarization model also offers a way to separate the polarization energy from the charge-transfer energy in a physically consistent way. For vdW interactions, the buffered-14-7 potential used in AMOEBA is re-parametrized by targeting the SAPT exchange-repulsion and dispersion energy.[236] This vdW model will be dis-

cussed in Chapter3.<sup>1</sup>

### 1.4.1 S101x7 SAPT2+ Database

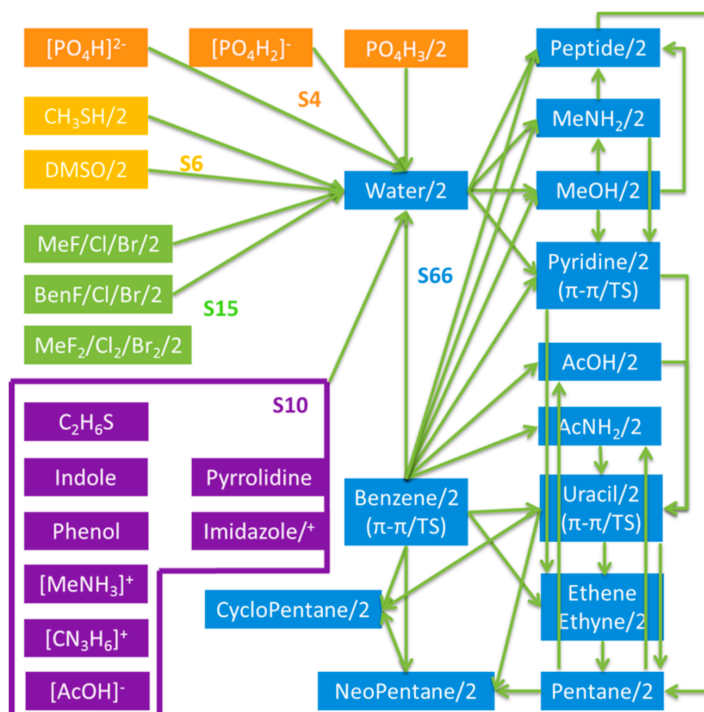


Figure 1.3: Schematic view of molecules in the S101 dataset. The arrows connect two molecules that form a dimer; “/2” represents the existence of a homodimer; “/+” indicates both neutral and ionized molecules are included.

<sup>1</sup>The first two subsections’ work (1.41 and 1.42) were previously published.[9] Q. T. Wang and J. A. Rackers developed and validated the model. C. F. He and I constructed QM database, which contains the total interaction energy as well as each energy component of 707 dimer. C. Narth, L. Lagardre, N. Gresh, J. W. Ponder and J.P. Piquemal and P. Y. Ren helped revised the paper. Besides, the many body interactions section (1.43) were previously published .[174]. C. W. Liu developed the model, I constructed the training data and helped writing the paper.

The S101x7 contains the detailed energy decomposition results of 39 common motifs and functional groups in biomolecules at various intermolecular distances.[304] The first 66 dimer structures are taken from the S66 database from Hobza et al.[250] This database covers the commonly encountered elements in biochemistry: H, C, N, O, S, P, and halogen atoms (F, Cl, and Br) in 11 chemical types including alcohol, acid, amide, amine, sulfoxide, sulfide, alkane, alkene, alkyne, haloalkane, and phosphate ions. The small molecules included in S101x7 are generally carriers of the functional group of interest while the larger ones are actual biomolecular building blocks. For example, the database includes both the neutral and ionized amino acid side chain analogs, peptide bond model N-methylacetamide and uracil.(Figure 1.3) Each dimer complex was placed at seven separations (0.70, 0.80, 0.90, 0.95, 1.00, 1.05, 1.10 times of the equilibrium distances) using the same definitions of the intermolecular distance vectors from the S66x8 database. The lower bound is at very short separations, 0.7 times the equilibrium, which is rarely investigated but is fundamental to the study of the short-range charge penetration effect of electrostatic and exchange-repulsion of van der Waals components. In the S101x7 database, the newly added pairs were optimized using MP2/cc-pVTZ method with counterpoise correction and the interaction energy was decomposed using SAPT2+/aug-cc-pVD(T)Z method[117, 222] in PSI4 program.[285] The results contain electrostatics, exchangerepulsion, induction, and dispersion components.[222] Different configurations of the same dimers are included in the data set to take into account orientation effects. These QM

energy decomposition results can be used as a reference for parameter training force fields to represents protein-phosphate binding and understand molecular recognition.

### 1.4.2 Charge Penetration Model

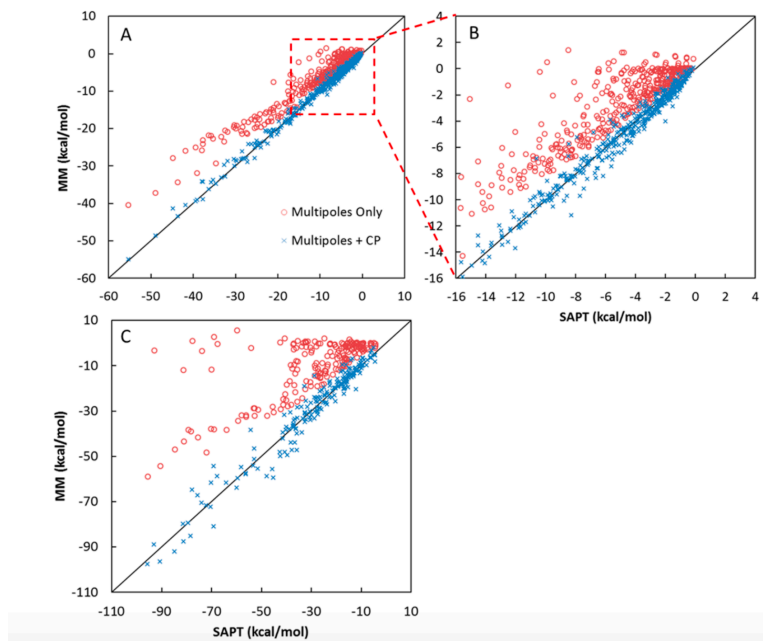


Figure 1.4: Plots of multipole electrostatic energy (kcal/mol) against the reference SAPT2+/aug-cc-pVTZ calculation for (A) near-equilibrium (0.90, 0.95, 1.00, 1.05, and 1.10) complexes taken from the S101x7 dataset, (B) expanded plot of the boxed region in A, and (C) short-range (0.70 and 0.80) complexes in the S101x7 dataset. The uncorrected AMOEBA point multipole energy (multipoles only) is shown in red circles, and the charge penetration corrected point multipole energies using the valence- $\alpha$  parameter set (multipoles + CP) are denoted by blue crosses.



By comparing to the electrostatic energy terms in S101x7 database, a general model has been developed to incorporate the charge penetration effect into the AMOEBA force field.[305] By replacing the Coulomb electrostatic interaction with the charge penetration corrected model, this model significantly improves the agreement between point multipoles and quantum mechanical electrostatic energies from SAPT2+ decomposition even at short inter-molecular distances. (Figure 1.4) The use of 18 pairs of charge penetration parameters for 9 chemical elements reflects the robustness and transferability of this model. The charge penetration correction is short-ranged and rapidly converges to the classical Coulomb interaction beyond 6 – 7 Å. Thus, it can be completely incorporated into the real space of Ewald summation without any additional computational cost in reciprocal space. Because simulations including penetration correction are clearly feasible, there is ongoing work dedicated to the optimization of parallel scaling the coupled penetration/smooth particle mesh Ewald approach. In addition, higher order penetration corrections (charge-dipole and charge-quadrupole penetration) are also possible and have been implemented in models such as SIBFA.[94]

### 1.4.3 Many-Body Interactions

The non-additive many-body interactions are significant for structural and thermodynamic properties of condensed phase systems. Recently, the many-body interaction energy of a large number of common organic/biochemical molecular clusters have been examined, which consist of 18 chemical species

and cover nine common organic elements, using the MllerPlesset perturbation theory to the second order (MP2). The capability of Thole-based dipole induction models has been evaluated to capture the many-body interaction energy. Three models were compared: the original model and parameters (model 0) used by AMOEBA force field, a variation of the original model (model 1), with 0.34 damping parameters which have been re-optimized to MP2 data and a third model (model 2) where the damping function form applied to the permanent electric field is modified as  $\lambda'_3(r) = 1 - \exp^{-au(r)^{\frac{3}{2}}}$ . [174]

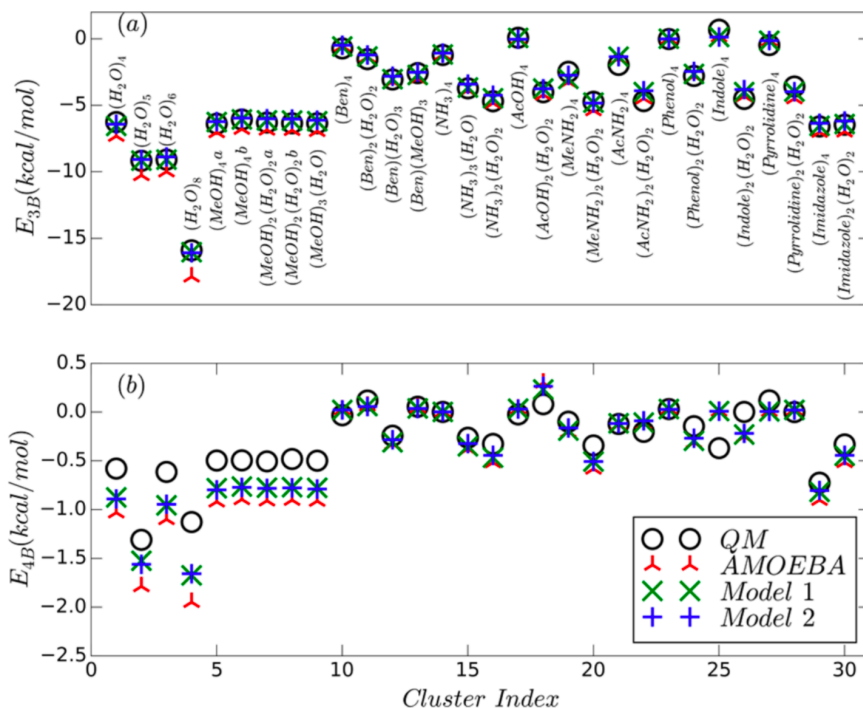


Figure 1.5: The plot of (a) the  $E_{3B}$  and (b)  $E_{4B}$  calculated using the three polarization models and QM methods (MP2/aug-cc-pvtz for Water-4568 and RI-MP2/cc-pvtz for Tetramers set). Each (MeOH)<sub>4</sub> and (MeOH)<sub>2</sub>(H<sub>2</sub>O)<sub>2</sub> cluster has two different structures.

The many-body interactions ( $E_{3B}$  and  $E_{4B}$ ) of organic molecular clusters have been well captured by classical polarization models. Overall, the current AMOEBA model and two new models are able to reasonably describe the three- and four-body energy for a wide range of organic molecular clusters in different configurations. (Figure 1.5 and Figure 1.6). In these simple models, universal parameters controlling the damping strength perform well for all organic species tested in this study. Comparing to the current AMOEBA,

as expected, the two new models that have been explicitly fitted to the  $E_{3B}$  show better agreement with the MP2 results. Model 2, where the damping function for direct induction due to the permanent field was modified, best reproduced the distance dependence of the  $E_{3B}$ . (Figure 1.5) In physical sense, these results clearly show that instead of using the smeared charge distribution given by the damping function of the current AMOEBA, a different charge distribution is needed to well capture the distance dependence of many-body interactions.

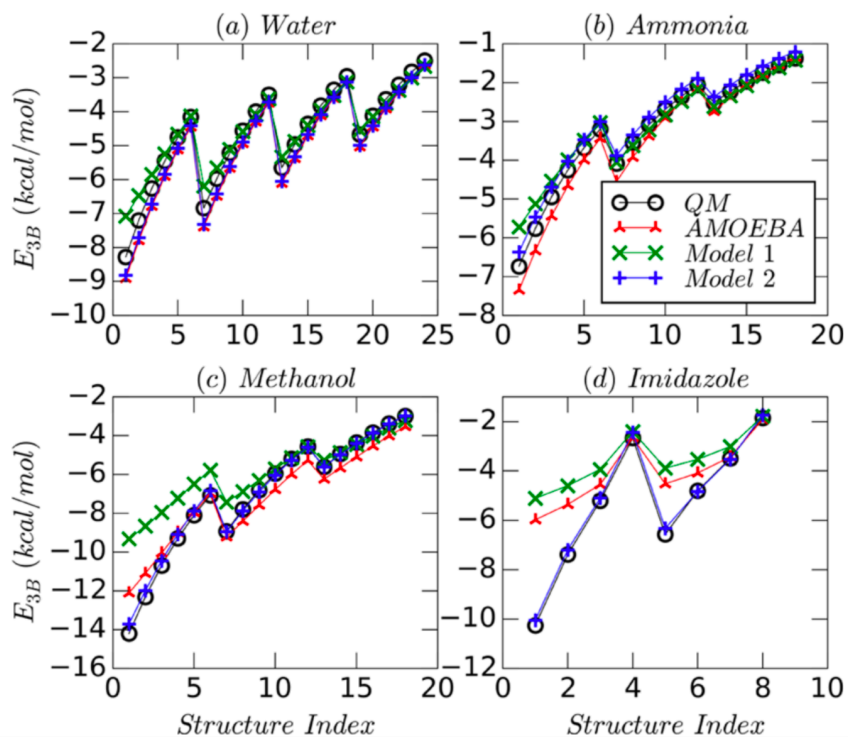


Figure 1.6: Plots of the  $E_{3B}$  distance dependence calculated from three polarization models, MP2/aug-cc-pvtz (for water) and RI-MP2/cc-pvtz (for other molecules) for selected trimer systems: (a) Water, (b) Ammonia, (c) Methanol and (d) Imidazole. The right-most structure indices represent the equilibrium structure for each trimer. The left side of x-axis indicates the smaller intermolecular distances than the right.

## 1.5 Overview of Thesis Work

The rest of this Thesis is organized as follows: Chapter 2 will cover the development of advanced polarizable force fields for water and organic molecules including the development of general van der Waals (vdW) po-

tential for common organic molecules and the coarse-grained water model united AMOEBA (uAMOEBA). Chapter 3 will describe the application of AMOEBA on elucidating the phosphate binding mode in phosphate binding protein (PBP) and analyzing the binding thermodynamics of series of host-guest systems vary from ligand to ligand. The conclusion is drawn in Chapter 4.

## Chapter 2

# Development of Advanced Polarizable Force Fields for Water and Organic Molecules

Molecular recognition is central to biomolecular processes such as signal transduction, metabolism, and gene expression.[115, 141] In drug discovery processes, one primary goal is to identify small ligands that can selectively bind to a macromolecular target with high affinities and favorable pharmacological properties.[139, 153] Computational modeling holds the promise of accelerating drug discovery and guiding molecular design by predicting ligand interactions with biomolecular targets.[140] Although computational modeling and simulation can provide a quantitative understanding of the mechanism underlying molecular recognition,[163] effective sampling methods and accurate force fields continue to be the main challenges for reliable prediction of molecular interactions. Accurately modeling intermolecular interaction is essential for organic/biochemical systems,[60] which motivates the necessity of accurate energy potentials in classical simulations. This chapter will cover the development of general van der Waals potential for common organic molecules, and the coarse-grained water model.<sup>1</sup>

---

<sup>1</sup>The vdW work (2.1) were previously published.[236] I developed and validated the vdW model, constructed both training and testing dataset, and wrote the paper. Q. T. Wang and

The first project presents a systematic development of a new van der Waals potential (vdW2016) for common organic molecules based on symmetry-adapted perturbation theory energy decomposition. The Buf-14-7 function, as well as Cubic-mean and Waldman-Hagler mixing rules were chosen given their best performance among other popular potentials. A database containing 39 organic molecules and 108 dimers was utilized to derive a general set of vdW parameters, which were further validated on nucleobase stacking systems and testing organic dimers. The vdW2016 potential is anticipated to significantly improve the accuracy and transferability of new generations of force fields for organic molecules.[236]

In addition, the second project reports the development of a united AMOEBA (uAMOEBA) polarizable water model. While providing comparable accuracy for gas-phase and liquid properties, uAMOEBA is computationally 3-5 times more efficient than the three-site AMOEBA03 model in MD simulations. In this coarse-grained polarizable water model, both electrostatic and van der Waals representations have been reduced to a single site located at the oxygen atom. Hydrogen atoms are retained only to define local frames for the molecular multipole moments and intramolecular vibrational modes. Good transferability from the gas to the liquid phase has been demonstrated over a wide range of temperatures, and from nonpolar to polar environments, due to

---

P. Y. Ren helped revising the paper. Besides, the united AMOEBA water model work (2.2) were previously published .[235]. L. P. Wang used the ForceBalanced software to optimize the model and I validated it using MD simulations and wrote the paper. V. Pande and P. R. Ren helped writing the paper.



the presence of molecular polarizability. The water coordination, hydrogen-bonding structure and dynamic properties given by uAMOEBA are similar to those derived from the all-atom AMOEBA03 model and experiments. Thus, uAMOEBA is an accurate and efficient alternative for modeling water.[235]

## **2.1 General van der Waals potential for common organic molecules**

### **2.1.1 Introduction**

Classical force fields such as AMBER,[38, 49] CHARMM, [24] OPLS-AA,[144] and GROMOS[242] typically model the electrostatic interactions with fixed atomic point charges and treat van der Waals interactions via simple functions, such as the Lennard-Jones (12-6) potential. As computationally efficient alternatives to quantum mechanical calculations, such classical point charge models have been widely used in molecular dynamic simulations of biological systems.[49, 144] Newer models that explicitly represent anisotropic atomic charge distributions and respond to surrounding changes via polarization[44, 230, 252] can potentially lead to improved transferability between different chemical and physical environments. For example, the point charge model can be replaced by atomic multipole moments, such as dipole and quadrupole, to capture the anisotropic nature of the electronic structures.[247] The polarization effects have been modeled by interactive atomic induced dipoles,[11] fluctuating charges[225] or Drude oscillators models.[12, 167] The AMOEBA force field, which employs atomic multipole moments and dipole

polarizabilities, is one typical example of such newer models.

Lennard-Jones (12-6) potential is one of the most common functions to describe the exchange-repulsion and dispersions interactions.[137] It has been adopted in a range of force fields[7, 46, 52, 171, 194, 216, 241] such as AMBER[49] and CHARMM[24]. This simple function contains only two parameters and it is faster to compute than exponential terms.[172] Other Lennard-Jones potential forms include the Lennard-Jones (9-6) function[14, 126] and the Lennard-Jones (12-10) function[216] which is an alternative to the usual Lennard-Jones (12-6) function in AMBER[49] to model hydrogen bonds. As a special case of the Buckingham potential function,[27] the exponential-6 potential is widely used in force fields[59, 71, 194] such as MM2,[4] MM3,[6] and MM4.[5] However, both Lennard-Jones and exponential-6 potentials account poorly for the high quality noble gas data while a simple buffered 14-7 (Buf-14-7) potential can accurately reproduce the noble gas potentials over a range of distances.[104] Besides the two interaction-specific parameters, the well depth and minimum-energy distance, the Buf-14-7 functional form is capable of adjusting the curvature of request depending on two shape parameters.

In addition to the functional forms, vdW interactions between unlike atoms (e.g. O and H) depend on the suitable combination of mixing rules to generate the well depth and distance parameters, unless pair-wise parameters are specified for all combinations. Current available force fields employ either geometric or arithmetic mean as the combining rules, which lead to large errors for mixed noble gas atom pairs.[296] More elaborate combination rules have

been proposed, involving additional parameters such as polarizability, ionization potentials, or dispersion force coefficients.[261] Newly derived combining rules, Waldman-Hagler (W-H)[296] and Halgren (HHG),[104] require no additional parameters other than the well depth and distance, also well reproduce the experimental values of noble gas interactions.

In classical force fields, electrostatic parameters, such as atomic charges, are normally computed from ab initio quantum mechanics directly.[262] The so-called vdW interaction in a force field is actually less well-defined and is essentially utilized to capture everything beyond the charge-charge interactions, including the charge penetration, charge transfer and perhaps some of the many-body polarization effects in condensed-phase. Additional difficulty in force field parameterization lies in the limitations of ab initio methods. A perfect ab initio force field will not reproduce experimental condensed-phase thermodynamic properties without explicit quantum corrections such as zero-point energy. Thus, deriving vdW parameters by fitting to gas-phase ab initio molecular interaction energy alone is insufficient if the resulting force fields were intended for condensed-phase systems, even with many-body polarization were explicitly accounted for. Given these limitations, the most effective approach, pioneered by Jorgensen and co-workers in developing OPLS and OPLSAA force fields,[51, 107, 147, 233, 254] has been to used derive vdW parameters directly against reliable and widely available liquid properties such as density and heat of vaporization.

The current AMOEBA force field and most other widely used fixed-

charge force fields address this issue by using a less repulsive van der Waals potential to compensate the missing charge penetration contribution in short range. The problem with this approach however, is that the resulting vdW parameters are less transferable.[209,277] Given the recent advances in computational chemistry and the computing power, the noncovalent interaction energy components can be computed using highly accurate quantum mechanical methods[223] such as symmetry-adapted perturbation theory (SAPT).[134] Recently, a general electrostatic model has been developed to incorporate the charge penetration effect based on the atomic multipole moments in all-atom AMOEBA.[305] This model significantly improves the agreement between point multipoles and quantum mechanical electrostatic energies from SAPT2+ decomposition even at short inter-molecular distances. This improvement leads to the need to revisit vdW potentials that are capable of capturing SAPT repulsion and dispersion energy.

This work presents the first systematic attempt to develop a new van der Waals potential, vdW2016, for common organic molecules based on SAPT energy decomposition. First, several common vdW functions were examined to determine the best functional form based on noble gas interaction energy data. A general set of parameters (element based) was then determined by fitting to the exchange-repulsion and dispersion energies for 756 dimers made of 39 common organic molecules with different configurations and separation distances (S108x7 database). This new vdW potential was further validated on nucleobase stacking systems as well as on a testing set of small organic dimers

(S36x7). Overall, the vdW2016 potential is expected to provide accurate prediction of vdW interaction energy for common organic molecules, including drug candidates.

## 2.1.2 Computational Methods

### 2.1.2.1 Van der Waals functions and mixing rules

The vdW functional form and mixing rules were determined from noble gas systems including helium (He), neon (Ne), argon (Ar), and krypton (Kr), where the non-bonded interactions were dominated by vdW interactions. Three typical vdW functional forms, Buf-14-7, Buckingham, and Lennard-Jones (12-6), were examined using the same sigma and epsilon extrapolated directly from the exchange-repulsion and dispersion energy of SAPT2+/aug-cc-pV5Z calculations (the dispersion energy has been scaled by 0.89). Different combinations of the vdW functions with three sets of mixing rules, Cubic-mean/HHG, W-H/W-H, Cubic-mean/W-H were examined against the SAPT2+/aug-cc-pV5Z potential surfaces. According to the best fitting results with the perfect match with SAPT (Figure 2.1) and the lowest RMSE (Figure 2.2), I chose the Buf-14-7 (Equation 2.1) as the functional form, and as mixing rules, Cubic-mean for  $\epsilon$  (Equation 2.2) and W-H for  $\sigma$  (Equation 2.3). The parameter  $\epsilon_{jj}$  is the potential well depth, and  $R_{ij}/\sigma_{ij}$  is the ratio between the actual separation of i-j atomic sites and the minimum-energy distance.[104]

$$E_{vdW} = \epsilon_{ij} \left( \frac{1 + \delta}{\frac{R_{ij}}{\sigma_{ij}} + \delta} \right)^7 \left( \frac{1 + \gamma}{\left(\frac{R_{ij}}{\sigma_{ij}}\right)^7 + \gamma} - 2 \right) \quad (2.1)$$

$$\sigma_{ij} = \frac{(\sigma_{ii})^3 + (\sigma_{jj})^3}{(\sigma_{ii})^2 + (\sigma_{jj})^2} \quad (2.2)$$

$$\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2} \frac{2(\sigma_{ii})^3(\sigma_{jj})^3}{(\sigma_{ii})^6 + (\sigma_{jj})^6} \quad (2.3)$$

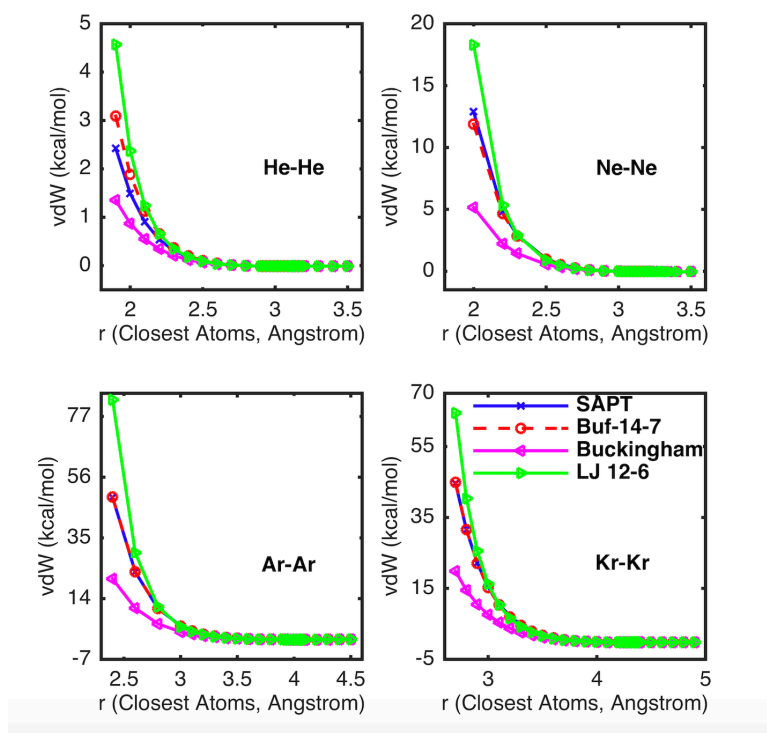


Figure 2.1: Plots of homodimer vdW energies in noble gas systems calculated by Buf-14-7 (red dash line with circle markers), Buckingham (magenta line with left-pointing triangle markers), and Lennard-Jones 12-6 (green line with left-pointing triangle markers) compared to QM reference SAPT2+ calculation (blue line with cross markers). Same sigma and epsilon parameters were used in all calculations: 3.11 Å and 0.014 kcal/mol for He; 3.31 Å and 0.048 kcal/mol for Ne; 4.07 Å and 0.160 kcal/mol for Ar; and 4.38 Å and 0.218 kcal/mol for Kr.

RMSE	Distance	Buf-14-7	Buckingham	LJ (12-6)
Cubic-mean/HHG	0.6-0.9	0.540	4.245	3.731
	0.9-1.1	0.005	0.007	0.007
	1.1-2.0	0.004	0.009	0.008
W-H/W-H	0.6-0.9	0.256	4.369	3.392
	0.9-1.1	0.004	0.005	0.005
	1.1-2.0	0.002	0.008	0.006
Cubic-mean/W-H	0.6-0.9	0.238	4.439	3.060
	0.9-1.1	0.003	0.005	0.005
	1.1-2.0	0.001	0.007	0.006

Figure 2.2: Differences between heterodimer vdW energies in noble gas system given by different combinations of vdW functional forms and mixing rules (ij/ij) compared to SAPT 2+ results. Same functional forms and parameters as above were used. Distances are ratios relative to the vdW minimum distances extrapolated from the SAPT2+ exchange-repulsion and dispersion energy.

### 2.1.2.2 QM Database for Model Training and Testing

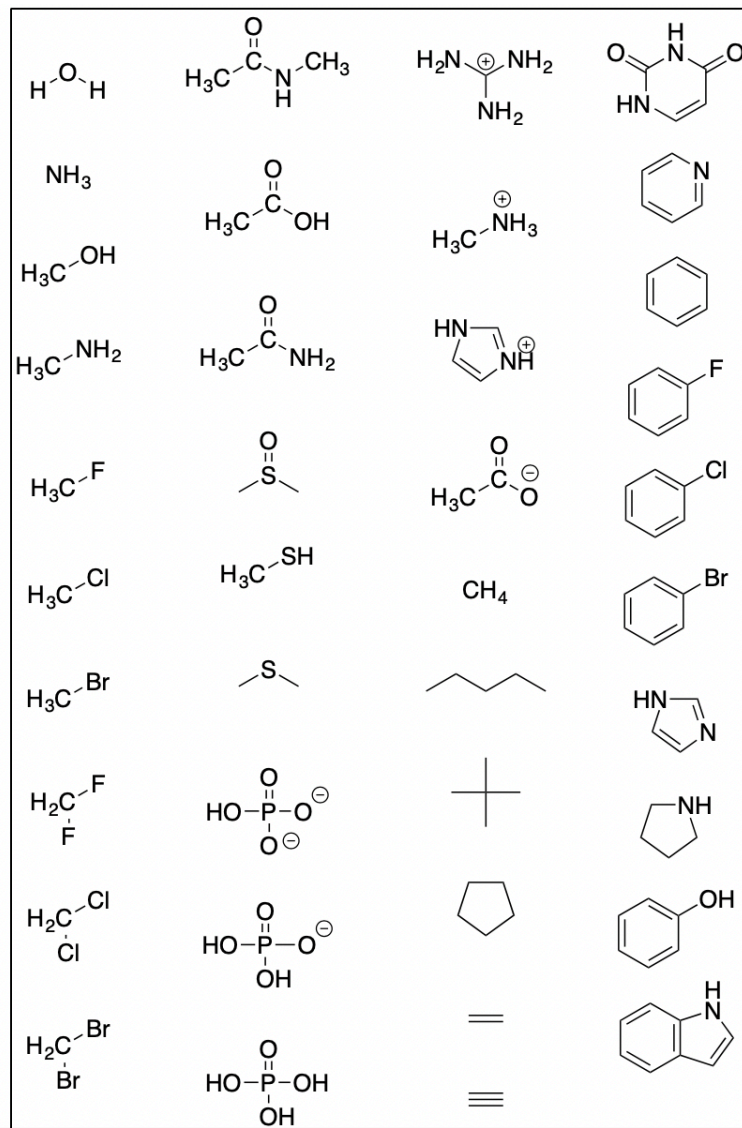


Figure 2.3: Molecule structures in the S108x7 database.



**S108 x7 fitting database.** The interaction energies of seven heterodimers between methane, amine, ethene and water were added to the S101x7 database (39 molecules and 101 dimers)[305] to construct the S108x7 fitting database.(Figure 2.3) The same structural optimization, distance generation and energy decomposition methods, including the extrapolation scheme and the scaling of dispersion energy, as in the S101x7 database construction were used to generate the energy profiles for newly added dimers. The parameters of the vdW model were optimized to match the sum of the dispersion and exchange energy in the S108x7 SAPT2+ database, which contains the total interaction energies as well as decomposed energy components for 108 pairwise interactions at seven distances (0.70, 0.80, 0.90, 0.95, 1.00, 1.05, 1.10 times of the equilibrium distances  $R$ ). The mean unsigned error (MUE), mean signed error (MSE), and root-mean-square error (RMSE) were calculated to examine the performance of different models at all 7 distances for a total of 756 dimers. This set includes 216 dimers at 2 short distances (0.70 $R$  and 0.8 $R$ ), and 540 dimers at 5 near equilibrium distances (0.9 $R$ , 0.95 $R$ ,  $R$ , 1.05 $R$  and 1.10 $R$ ).

**S36x7 SAPT0 testing dataset.** An S36x7 testing dataset was constructed containing newly generated non-aromatic heterodimer interaction energies between polar-nonpolar, polar-polar, and nonpolar-nonpolar molecules. Fifteen representative molecules were chosen, which covers 11 chemical types including alcohol, acid, amide, amine, sulfoxide, sulfide, alkane, alkene, alkyne, haloalkane, and phosphate ions in the S108x7 fitting database. Identical to the structural generation and optimization procedures used in the S108x7

database, the structures of the newly added pairs were optimized using MP2/cc-pVTZ method. Each dimer complex was studied for seven separation distances ranging from 0.70-1.10 times the equilibrium distance. For each of the resulting dimer configurations, the interaction energy was decomposed using SAPT0/jun-cc-pVDZ method[117,222] in PSI4 program.[285] For aromatic systems, existing data for nucleicbases were utilized in the testing.

### **2.1.3 Results and Discussion**

#### **2.1.3.1 Atom Types and Parameters**

An iterative process was applied to determine the vdW types and parameters: the universal shape parameters delta and gamma for Buf-14-7, and the interaction parameters sigma and epsilon for each atom type. I first kept the shape parameters fixed to optimize sigma and epsilon. Then, the delta and gamma were optimized with the fixed the interaction parameters generated previously. New vdW types were added after analyzing the systematic errors of the model compared to SAPT. The final optimized delta and gamma parameters in the Buf-14-7 vdW function are 0.273 and 0.025, respectively. For comparison, the previous AMOEBA[244] and MMFF94[105] force fields both used 0.07 and 0.12 for delta and gamma, respectively. The nonlinear least square optimization method, lsqnonlin, in Matlab[100] was applied to all fitting, including the optimization of delta and gamma parameters in the Buf-14-7 vdW function,  $\sigma$  and  $\epsilon$  parameters for each vdW type. The inputs of the lsqnonlin program were the objective function, initial parameters, and pa-

parameter limits. The objective function is the mean squared difference between calculated vdW and SAPT values of all structures.

<i>N</i>	vdW type	$\sigma$ (Å)	$\epsilon$ (kcal/mol)	<i>N</i>	vdW type	$\sigma$ (Å)	$\epsilon$ (kcal/mol)
1	H (nonpolar)	3.509	0.005	15	N (sp2)	3.816	0.109
2	H (nonpolar, Alkane)	3.475	0.005	16	N (aromatic)	3.691	0.131
3	H (polar, NH/N aromatic)	3.450	0.002	17	O (sp3, hydroxyl, water)	3.571	0.078
4	H (polar, OH)	3.541	0.002	18	O (sp2, carbonyl)	3.690	0.077
5	H (polar, SH)	3.550	0.005	19	O (O <sup>-</sup> in AcO <sup>-</sup> )	3.695	0.097
6	H (aromatic, CH)	3.385	0.003	20	O (O <sup>-</sup> in HPO <sub>4</sub> <sup>2-</sup> )	3.693	0.097
7	C (sp3)	3.839	0.036	21	O (O <sup>-</sup> in H <sub>2</sub> PO <sub>4</sub> )	3.691	0.085
8	C (sp3, Alkane)	3.816	0.012	22	O (O in H <sub>3</sub> PO <sub>4</sub> )	3.550	0.068
9	C (sp2, Ethene)	4.393	0.038	23	P (phosphate)	4.856	0.039
10	C (sp2, C=O)	4.129	0.012	24	S (sulfide, R-SH)	4.541	0.156
11	C (sp)	4.255	0.046	25	S (sulfur IV, DMSO)	4.413	0.130
12	C (aromatic, C-C)	4.420	0.025	26	F (organofluorine)	3.551	0.050
13	C (aromatic, C-X)	4.221	0.034	27	Cl (organochloride)	4.421	0.124
14	N (sp3)	3.753	0.141	28	Br (organobromine)	4.655	0.178

Figure 2.4: VdW types and parameters ( $\sigma$  and  $\epsilon$ ) for H, C, N, O, P, S, F, Cl, and Br. Reduction factors for all Hs are 0.93, and are 1.00 for other molecules. Delta and gamma are 0.273 and 0.025, respectively.

In Figure 2.4, all atoms in the S108x7 system were clustered into 28 vdW types. In addition to the 18 atom types used in the charge penetration paper,[305] ten more vdW types were added to better capture the vdW interactions. Polar H was divided into three types depending on whether it is connected to O, N or S. C in alkene (-C=C-) was separated from C in the carbonyl group. In aromatic rings, a separate type was used for C atoms bonded to non-carbon atoms (N and halogens). Both C and H in alkane molecules (sp<sup>3</sup> C only bonded to other sp<sup>3</sup> C or H) were separated from the general sp<sup>3</sup> C and non-polar H to two new classes. Additional types were assigned to oxygen ion in acetate and phosphate as well as sp<sup>2</sup> hybridized oxygen in phosphoric acid to capture their vdW interactions. S in sulfide and sulfoxide used different set of  $\sigma$  and  $\epsilon$ , while P, F, Cl, and Br have only one set of parameters each.

Good correlation has been found with a  $R^2$  value of 0.955 between the average for each element and the atomic radius calculated via SCF minimal basis-set functions.[47] This trend confirms the physical meaning of the sigma parameters is in relevance to the atomic diameters. A reduction factor of 0.93 was assigned to all H atoms in order to translate the vdW center of hydrogen toward the heavy atom along the chemical bond.[244]

### 2.1.3.2 Model Training Performance on Dimer Interactions

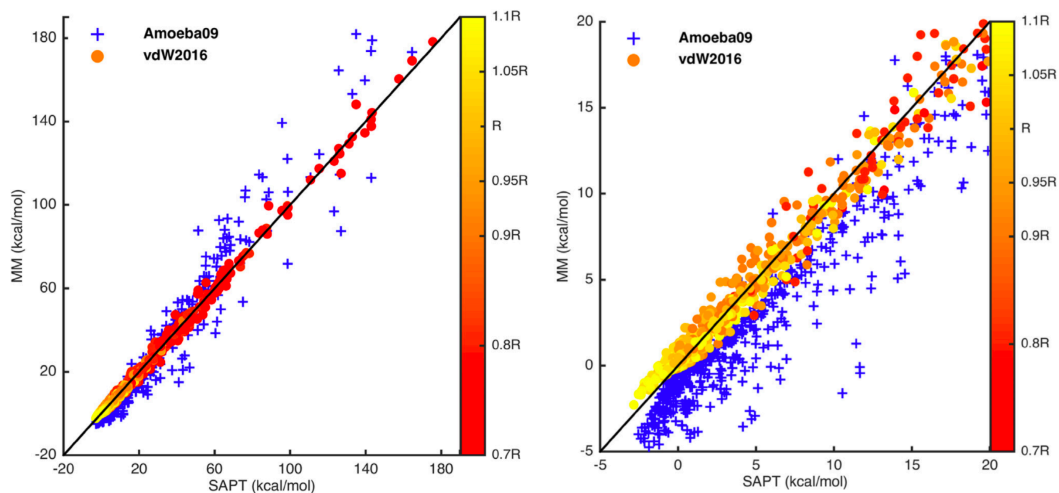


Figure 2.5: Plots of the vdW energy calculated by vdW2016 (circles filled with distances relative to the equilibrium distance  $R$ ) and AMOEBA09 model (blue plus signs) compared to the SAPT2+/CBS/scaled results. The AMOEBA09 vdW parameters were derived to reproduce the total dimer interaction energies and liquid thermodynamic properties in a force field without charge penetration corrections. The left plot contains all the 7 distances complexes in the S108x7 data set. The right plot expanded the top one and contains complexes with energy from -5 kcal/mol to 20 kcal/mol.

The van der Waals interactions calculated using the vdW2016 potential and the AMOEBA09[245] vdW component were compared to SAPT2+ results of the S108x7 dataset in Figure 2.5. By using a limited set of parameters, the new potential agrees with SAPT2+ with a correlation coefficient  $R^2$  of 0.996 across all dimers and distances. Figure 2.6 shows the 108 dimers energy profiles from both the vdW2016 potential and SAPT. The errors (blue line) of the new vdW potential are rather even across all the dimers. As expected, the AMOEBA09 vdW parameters, based on the same Buf-14-7 function but with different shape parameters and mixing rules, underestimates the short-range vdW energy in comparison with SAPT, although displaying a very reasonable correlation with QM results ( $R^2$  of 0.946). The AMOEBA09 vdW parameters were parameterized to capture the total interaction energy of molecule dimers and liquid thermodynamic properties. This underestimation of the repulsive energy was needed to compensate the missing charge penetration term in the original AMOEBA model. A similar trend is also seen for other fixed-charge force fields as discussed in later sections. In contrast, the vdW2016 potential has been specifically fitted to the sum of the dispersion and exchange-repulsion energy from SAPT.

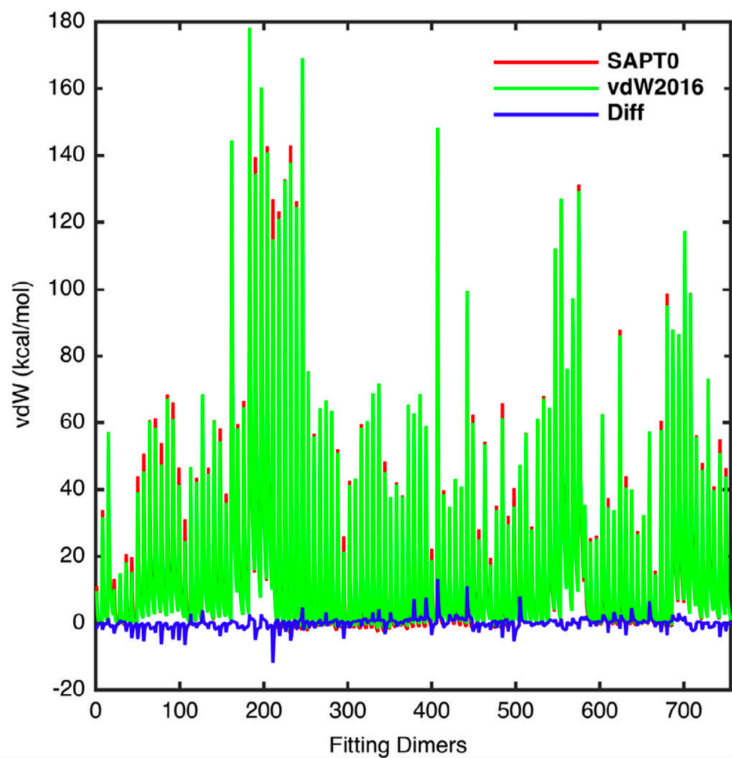


Figure 2.6: Plot of the fitting vdW energy calculated by vdW2016 model (green line) compared to the SAPT2+/CBS/scaled energy (red line). The blue line is the energy differences between the vdW2016 and QM results.

The statistical analysis of the vdW2016 model and the AMOEBA09 vdW component at different distances are listed in Figure 2.7. For 108 dimers at all 7 distances, the vdW interaction energy values given by the vdW2016 potential and SAPT2+ are in excellent agreement, with a MUE of 1.054 kcal/mol, a RMSE of 1.680 kcal/mol, and a MSE of -0.006 kcal/mol. This agreement is remarkable given that some of the vdW energies at close distances are over a hundred kcal/mol (Figure 2.6), i.e. 56.5% of the complexes have the energy

over 50kcal/mol and 14.8% are over 100 kcal/mol. For dimers in five distances near equilibrium, the MUE and the RMSE of the newly parameterized vdW potential are 0.658 and 0.835 kcal/mol, respectively, a more than 3-fold reduction from the values of the previous AMOEBA model, 2.125 and 2.780 kcal/mol, respectively. The good agreement persists when including dimers at short separations (0.70, 0.80) where the absolute vdW interaction energy is rather high (44.3 kcal/mol averaged over 216 dimers). Overall, the vdW2016 potential tends to slightly underestimate both the exchange-repulsion energy (less repulsive) at short separations given a MSE value of -0.345 kcal/mol, and the attraction energy (less attractive) near equilibrium distances, with a MSE value of 0.130 kcal/mol.

S108 fitting set	Statistics	New vdW model (kcal/mol)	AMOEBA09 vdW component (kcal/mol)
0.90R-1.10R (108 ×5 pairs)	MUE	0.658	2.125
	MSE	0.130	-1.970
	RMSE	0.835	2.780
0.70R-0.80R (108 ×2 pairs)	MUE	2.044	9.374
	MSE	-0.345	2.691
	RMSE	2.851	15.044
All distance 0.70R-1.10R (108 ×7 pairs)	MUE	1.054	4.196
	MSE	-0.006	-0.639
	RMSE	1.680	8.378

Figure 2.7: Differences between vdW energies given by vdW2016 potential, AMOEBA09 parameters, compared to SAPT2+/CBS/scaled results in the S108x7 fitting dataset.

It should be noted that the target SAPT2+ attractive dispersion energies have been scaled by 0.89 in the S108x7 database.[305] This scaling was obtained by comparing the SAPT2+ total interaction energy with high-level ab initio data (CCSD(T)/CBS)). Besides, another set of parameters has been derived by fitting to the non-scaled SAPT2+/CBS QM results with the same

functional form, shape parameters as above. After optimization, the  $\sigma$  and  $\varepsilon$  parameters were increased slightly with similar overall quality. In addition, damping function (Equation 2.4) constructed by Mooij et al.[203] with a simple  $c_{damp} = 3.54$  used by Weitao Yang[318] was tested to further adjust the short-range dispersion. With the same vdW interaction parameters but slightly different delta and gamma shape parameters (0.272 and 0.026, respectively), a same RMSE value of 1.68 kcal/mol was obtained with a slightly better repulsion at short separations (RMSE 2.850 kcal/mol) but somewhat worse dispersion at the equilibrium distance (RMSE 0.837 kcal/mol). Thus, the shape parameters, delta and gamma, in the Buf-14-7 are sufficient and the damping function is unnecessary for reproducing the vdW interaction energies of small organic molecules.

$$f_d(\sigma_{ij}) = \left( 1 - \exp \left[ -c_{damp} \left( \frac{R_{ij}}{\sigma_{ij}} \right)^3 \right] \right)^2 \quad (2.4)$$

### 2.1.3.3 Validation on Stacked Nucleobase and Heterodimers

The transferability of the vdW2016 model and parameters has been further validated on molecular dimers that have not been included in the fitting process. Interaction energies between stacked nucleobase dimers were used to test aromatic parameters. The additional S36x7 heterodimer interaction energies were used to validate other polar and non-polar vdW types.



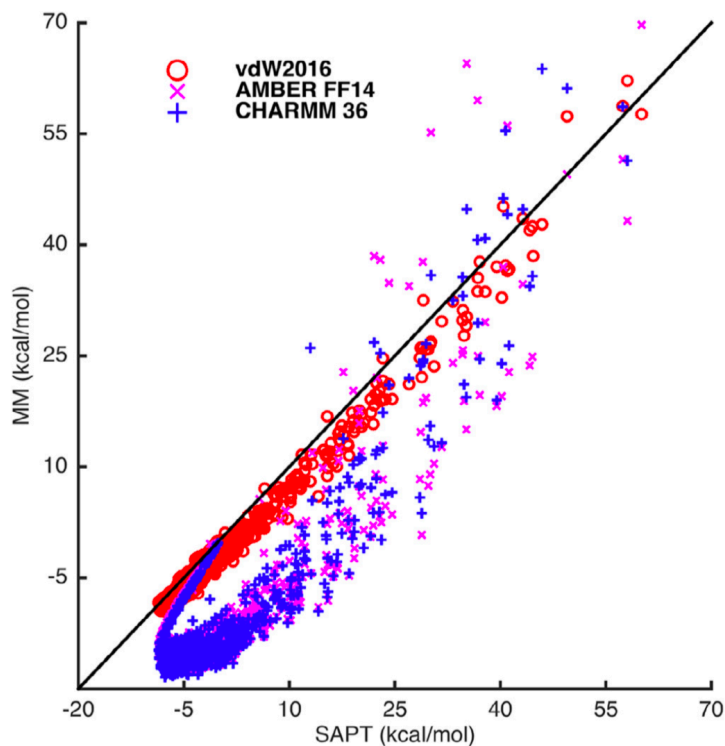


Figure 2.8: Plot of the vdW energy calculated using the vdW2016, AMBER FF14, and CHARMM 36 models compared to the SAPT0/jun-cc-pVDZ QM results of ten stacked base pairs across their rotational (Twist, Roll, Tilt) and translational (Rise, Slide, Shift) configurations. The geometries were taken from Sherrills work.<sup>55</sup>

**Validation on Stacked Nucleobase.** For aromatic systems, which are computationally expensive to obtain high level SAPT energy decomposition (e.g. SAPT2+), a previously published nucleobase stacking interaction energy database<sup>55</sup> was used in the validations. The QM energy decomposition was obtained using the SAPT0/jun-cc-pVDZ method. Only dimers with

total interaction energies less than 10 kcal/mol have been adopted in the testing. The sum of the SAPT0 repulsion and dispersion energy was compared with the energy calculated by the vdW2016 model. The comparison with AMBER FF[49] and CHARMM 36[113] from the previous publication has also been included, however, it should be noted again that the vdW terms in these force fields were derived by fitting to QM total interaction energy and/or liquid properties and implicitly include contributions such as charge penetration etc. Overall, the vdW2016 potential yields an excellent correlation with the SAPT0 energy ( $R^2 = 0.983$ ) for all base pairs in the translation and rotation configurations (Figure 2.8). It is also clear that the vdW energy from fixed-charge force fields systematically underestimates the vdW energy (too attractive) when compared to SAPT0 results, likely to compensate of the missing charge penetration contribution.

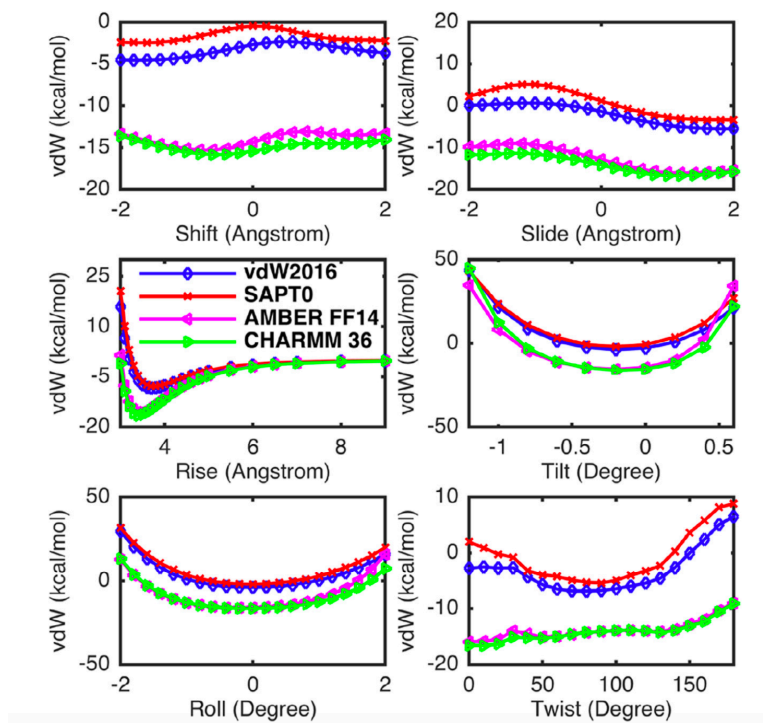


Figure 2.9: Plots of the AG:CT stacking vdW energy calculated using the vdW2016, AMBER FF14, and CHARMM models across rotational (Twist, Roll, Tilt) and translational (Rise, Slide, Shift) configurations.

The AG and CT stacking pairs were examined in detail in Figure 2.9. For some configurations, such as Shift, Slide, and Twist, a systematic shift was observed between the vdW2016 potential and the SAPT0, with the new vdW model predicting somewhat lower energy. This is due to the missing higher order terms in the SAPT0 level calculation while the vdW2016 potential was parameterized using SAPT2+ data. Although the SAPT0/jun-cc-pVDZ method performs qualitatively well for the mixed and dispersion dominated (MX/DD)

complexes with a mean absolute error (MAE) of 0.32 kcal/mol compared to the CCSD(T)/CBS method,[191, 222] the SAPT0 interaction energy is systematically shifted upwards (less negative) compared to SAPT2+ and yields positive mean signed errors in several systems.[222] For example, it was shown that the MSE of SAPT0, compared to CCSD(T)/CBS,[191] is 0.11 kcal/mol for the Non-Bonded Curves (NBC10) systems,[145, 191] 0.47 kcal/mol for the HSG set,[79] 0.08 and 0.05 kcal/mol for the mixed influenced and dispersion-dominated bound complexes in S22 set,[145] respectively. Thus, the errors would be reduced if the more expensive SAPT2+ method was used.

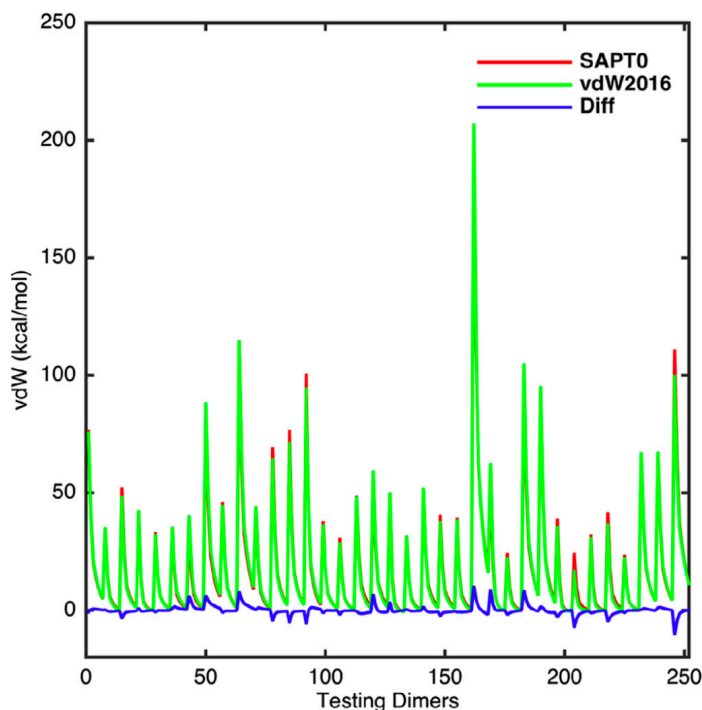


Figure 2.10: Plot of the vdW energy values of the testing set S36x7 calculated using the vdW2016 potential (green line) compared to the QM reference SAPT0/jun-cc-pVDZ (Red line). The blue line is the energy differences between the vdW2016 and SAPT0 results.

**Validation on Heterodimers.** Heterodimers of non-aromatic small organic molecules, which were not included in the fitting, were further examined to validate the transferability and accuracy of the vdW2016 potential. Fifteen representative molecules with 11 chemical types were chosen from the S108x7 database to construct the testing dataset (S36x7) covering interactions among alcohol, acid, amide, amine, sulfoxide, sulfide, alkane, alkene, alkyne, haloalkane, and phosphate ions. This dataset includes new cross interactions

that were not covered in the original S108x7 database used for parameterization. Again, the vdW SAPT0 energy decomposition was obtained on these new dimers. Overall, a strong correlation, with  $R^2$  value of 0.994 for 259 heterodimers, was found in this testing set. In contrast to the optimization process, where the new vdW potential tends to underestimate the dispersion attraction near equilibrium distance (MSE 0.13 kcal/mol), an overestimation was found (MSE of -0.11 kcal/mol) in the same distance range, which might be due to the lower level SAPT0 reference energy is less attractive than the higher level SAPT2+ references used in S108x7 dataset.

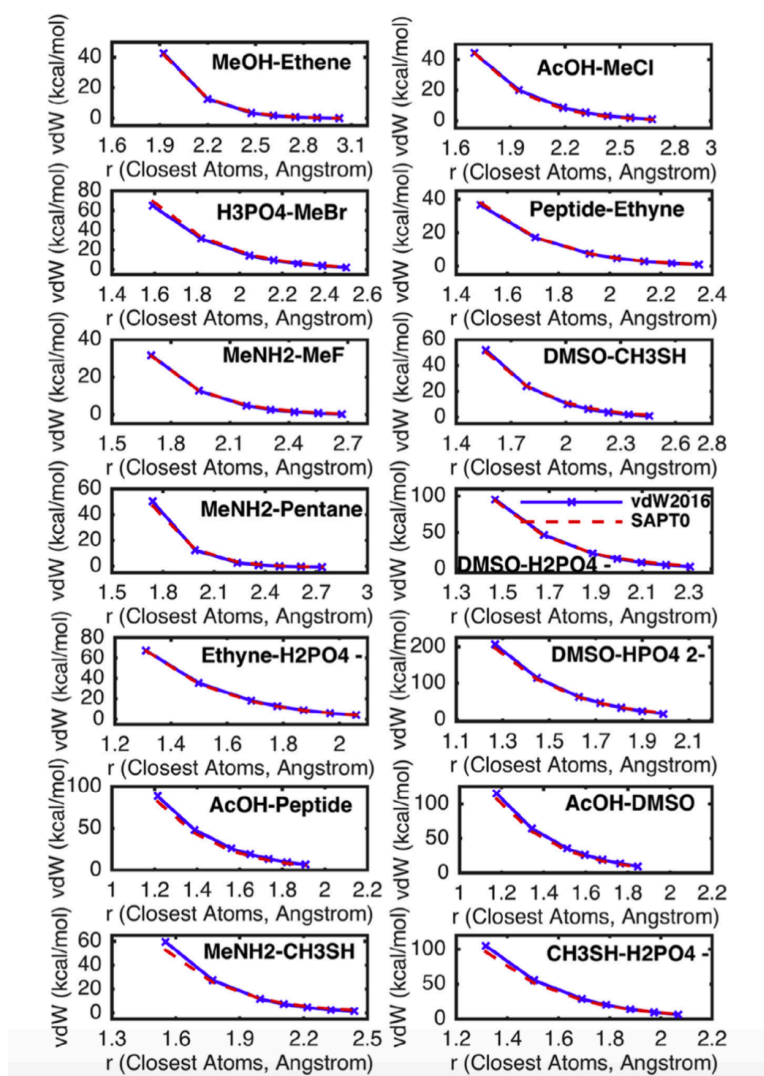


Figure 2.11: Plots of the vdW energy surfaces of selected pairs calculated using the new vdw2016 model (green line) and the reference SAPT0/jun-ccpVDZ (red line) in the S36x7 dataset.

All the testing vdW energies predicted by the vdw2016 model are compared with the SAPT0 results in Figure 2.10. Larger errors, either overestima-

tion or underestimation, are shown as the peaks and valleys in the blue line. Overall, the errors are rather even across all heterodimers and the shapes of the curve match the QM results well. Figure 2.11 shows the interaction energy surfaces of the selected pairs in details. Excellent agreement between the vdW model and SAPT0 can be observed for interactions between hydrogen bond donor (polar H) and acceptors (O, N or F), pairs made of alcohol, acid, amide, amine, haloalkane. For dimers with very short separation distances at around 1.3 angstrom, the vdW potential tended to overestimate the vdW energies in comparison with the SAPT0 results (See the plots of dimers DMSO-HPO4<sup>2-</sup>, AcOH-Peptide, AcOH-DMSO, MeNH<sub>2</sub>-CH<sub>3</sub>SH, and CH<sub>3</sub>SH- H<sub>2</sub>PO<sub>4</sub><sup>-</sup> in Figure 2.11). In addition, the T-shaped configuration of interactions between ethynes H and ethenes double bond are the most problematic in both fitting and testing sets. Anisotropic vdW treatments, such as the Gay-Berne potential,[89] may be necessary for C and H in unsaturated hydrocarbon.

#### **2.1.3.4 Further Development: VdW 2017 Model**

The vdW2016 model discussed above was parametrized on the total vdW energy, the sum of dispersion and exchange-repulsion of the SAPT2+ energy composition. Recently, our lab re-visit this model by fitting to each dispersion and exchange-repulsion components at the same time (vdW2017) Improvements have been observed in both dispersion and exchange-repulsion energy while the overall performance of the total vdW potential is slightly worse than vdW2016. (Figure 2.12) More details and updates can be found



at <http://biomol.bme.utexas.edu/ch38988/s36x7>.

Models	Dispersion		Exchange		Dispersion + Exchange	
	All	Near equilibrium	All	Near equilibrium	All	Near equilibrium
vdW2016	4.7	2.5	5.0	2.5	1.7	0.8
vdW2017	1.5	0.7	2.6	1.1	2.0	0.9

Figure 2.12: RMS error (in kcal/mol) of vdW2017 model comparing to the SAPT2+ components for S108x7 dimer set. VdW2016 model is also provided for comparison. All represents the dimers of 0.7 to 1.1-fold of the equilibrium distance while Near equilibrium represents the dimers of 0.9 to 1.1-fold of the equilibrium distance.

#### 2.1.4 Conclustions

The van der Waals interaction is one of the dominant noncovalent interactions in molecular systems. Accurate treatment of vdW interaction is crucial in molecular modeling. This work reportes the first systematic attempt to develop a new vdW potential, vdW2016, based on the SAPT2+ energy composition. Based on investigations of noble gas molecular interactions, the soft-core Buf-14-7 was found to best reproduce the QM (SAPT2+) repulsion and dispersion energy. The most effective mixing rules were found to be Cubic-mean for sigma and W-H for epsilon. A quantum mechanics interaction energy database (S108x7) was used to calibrate the vdW model, which contains the detailed energy components for 756 pairwise interactions between 39 common organic molecules at various inter-molecular distances. The parameters of the vdW2016 model were derived by fitting to the sum of the SAPT2+/CBS/scaled exchange-repulsion and dispersion energy in the S108x7

database. With 28 pairs of vdW parameters for 9 chemical elements, I am able to reproduce vdW interaction energy for all dimers in this database with a RMSE of 1.68 kcal/mol (0.84 kcal/mol if within 90-110% of the equilibrium distances). Subsequently, the accuracy and transferability of the vdW2016 potential were validated by comparing to the SAPT0 energy of an additional set of stacked nucleobase62 and newly generated cross-interacting dimers (S36x7). As a result, the vdW2016 potential yields strong correlations with  $R^2$  of 0.983 for all base pairs and 0.994 for 259 heterodimers compared to the SAPT0 energy.

Overall, the present systematic study shows that the accuracy and transferability of the new vdW2016 potential are excellent in comparison with the computationally rigorous SAPT data. However, it should be noted that the gas-phase dimer interaction energy data alone is insufficient for determining the vdW potential for condensed-phase applications. This new potential nonetheless serves as a starting point for future refinement using molecular dynamics simulations targeting liquid properties.[143, 144] Recent advancement of decomposition of intermolecular forces has made it possible to construct molecular mechanics force fields more systematically and rigorously, which hopefully will lead to higher chemical accuracy and more general transferability.

## 2.2 United Polarizable Multipole Water Model for Molecular Mechanics Simulations

### 2.2.1 Introduction

Water is an important solvent in living systems[40, 83] and many industrial applications.[101, 111, 129, 215, 292] There are a number of molecular mechanics models, such as the three-site TIPS[138], TIP3P,[142] SPC,[282] and SPC/E[20], the four-site TIP4P[142],TIP4P-Ew[118], TIP4P/2005[1], and the five site ST2[275], TIP5P[183] and TIP5P-E[251], that are commonly used in molecular simulation of water. These models use fixed atomic partial charges, with electrostatic energy evaluated in pairwise-additive fashion. Some models incorporate explicit electronic polarization to allow the charge distribution to respond to electrostatic environment and to further improve the reproduction of many water properties; these include Dang-Chang[53], Thole-Type-Model (TTM)[30–32, 77, 322], SWM-4DP[160], DPP2[158] and AMOEBA03.[244] The AMOEBA03 water model[74] was developed with a focus on capturing molecular polarizability, electrostatic potential, as well as the interaction energy from gas to condensed-phase, by utilizing permanent atomic monopole, dipole and quadrupole moments and mutual atomic dipole-dipole induction.[230, 243]

The inexpensive AMOEBA (iAMOEBA) model was introduced recently as a way to achieve improved computational efficiency.[302] In this model, the induced atomic dipoles are determined directly from the permanent multipole electric fields without further interactions between induced dipoles. Thus, it captures some 3-body effects in polarization while reducing the com-

putational cost relative to the fully self-consistent AMOEBA03 water model by a factor of 1.5-6. An alternative to improve the computational efficiency, without sacrificing the many-body effect, is to reduce the number of interaction sites within the model. Water models with single dipole moment representation have been developed since the 1980s.[23] Ichiye and her coworkers[127, 176] introduced the soft sticky dipole potential for liquid water, with one spherical repulsive potential, a short-range tetrahedral sticky potential and a point dipole at the center of mass. Compared with TIP3P and TIP4P models, it produced similar liquid water properties but with up to one order of magnitude speed-up. Later in 2010, the soft-sticky dipole-quadrupole-octupole water model was presented by the same group. It was suggested that the addition of octupole moments improved the dielectric constant (75 at 298K).[278] Carnie and Patey reported a polarizable dipole-tetrahedral quadrupole water model and a self-consistent mean field theory was applied to account for molecular polarizability.[36] Later, Kusalik and Patey added the octupole moments to their water model discussed above and observed strong preferential solvation of anions at infinite dilution, suggesting an important role of octupole moments in ion solvation. Jonsson et al.[312] introduced a one-site water potential based on electrostatic, induction, dispersion and short-range repulsion interactions. Also, multipoles up to quadrupole moment for polarizability and up to hexadecapoles for permanent electrostatics were included. Previously, one-site non-polarizable models have been explored, based on permanent molecular multipoles and the Gay-Berne potential, for molecular liquids including wa-

ter.[88, 90] Molinero and Moore proposed a water model that further omitted long-range intermolecular interactions. This model generally reproduces the bulk properties of liquid water except for the self-diffusion coefficient, which is too fast.[200] The faster diffusion rate is likely due to the lack of long-ranged electrostatic forces in the model.

In this paper, a new (nonbonded) one-site polarizable water model, uAMOEBA, is described. This model removes two nonbonded interaction sites on H atoms to speed up the energy and force calculations by a factor of 3 to 5 times over the previous three-site AMOEBA03 model in molecular dynamics when particle-mesh Ewald is used to treat long-range electrostatics.[73] Importantly, the full many-body polarization effects are retained via mutual induction of molecular dipoles. In this model, the permanent molecular multipole (dipole and quadrupole) moments, isotropic molecular dipole polarizability, and a single vdW interaction site are placed on oxygen. The remaining hydrogen atoms only carry atomic masses to define the local coordinate frames and the intramolecular geometry. For computational efficiency, uAMOEBA does not contain any octupole moment, which is a potential limitation of this model. The hydrogen atoms experience forces due to valence (bond, angle and Urey-Bradley) interactions. In addition, the torques experienced by the molecular dipole and quadrupole moments on oxygen are translated into forces on the hydrogen atoms in the same way as all-atom AMOEBA03.[244] The use of a single scalar polarizability is well justified, as the three components of water molecular polarizability (1.528, 1.415, 1.468)

are indeed similar in magnitude. The accuracy of uAMOEBA water model is demonstrated by evaluating a range of gas-phase and liquid properties and comparing the results with the existing AMOEBA03/ iAMOEBA models and experimental data. The current work aims to extend beyond the previous one-site water models by systematically examining gas-phase clusters of increasing size and important thermodynamic properties at a wide range of temperatures. In addition, the uAMOEBA model is applied to study interactions between water and other common organic molecules (modeled by all-atom AMOEBA03 [245]) to investigate its transferability.

## **2.2.2 Methodology**

### **2.2.2.1 Parameterization Dataset**

The data utilized for fitting the parameters was composed of a combination of experimentally determined liquid properties as well as high-level ab initio QM-derived properties. The liquid properties considered were density, enthalpy of vaporization, isothermal compressibility, isobaric heat capacity, thermal expansion coefficient, and dielectric constant. The temperature and pressure combinations were: 1 atmosphere at temperatures ranging from 249 K to 373 K (32 total), and 298 K at pressures from 1 kilobar to 4 kilobar (4 total).

The ab initio QM reference data included properties for systems ranging in size from the monomer to clusters of 23 water molecules. For the monomer, the molecular dipole, quadrupole, polarizability, vibration, and ge-

ometry were considered. The ab initio QM interaction energy and geometries for the ground state dimer, Smith dimer set (10 total),[269] trimer, tetramer, pentamer, eight hexamers,[16] two octamers,[321] five 11-mers,[29] five 16-mers, two 17-mers, and four 20-mers[76],[75] were utilized for calibration. In previous work,[302] over 42000 cluster (ranging from 2 to 22 water molecules) geometries were obtained from AMOEBA03 simulations of liquid water for temperatures ranging from 249 K to 373 K. Energy and gradients for the clusters were determined via RI-MP2[273],[274]/heavy-aug-cc-pVTZ[63] as implemented in Q-Chem 4.0.[264] This large compilation of theoretical data was included in the fitting of the model parameters.

### 2.2.2.2 Parameter Optimization

ForceBalance, an automatic optimization method, was applied to parameterize the uAMOEBA water model using the expanded data set described above. ForceBalance supports many different optimization algorithms, and the calculation in this work was carried out with the trust-radius Newton-Raphson (or Levenberg-Marquardt[168],[189]) algorithm and an adaptive trust radius.[57],[204] This algorithm requires the first and second derivatives of the objective function in the parameter space, which we derive from the first derivatives of the simulated properties using the Gauss-Newton approximation.

A major challenge in force field parameterization is the significant statistical noise in the objective function from the sampling of properties to be

matched to experiment. The parametric derivatives are challenging to evaluate because numerical differentiation requires running multiple simulations and evaluating small differences between statistically noisy estimates. Force-Balance uses thermodynamic fluctuation formulas to calculate accurate parametric derivatives of simulated properties without running expensive multiple simulations.[58, 301] For instance, the ensemble average of a generic observable  $A$  that does not depend explicitly on the force field parameters (for example, the density or an order parameter) can be expressed as follows:

$$\begin{aligned}\langle A \rangle_\lambda &= \frac{1}{Q(\lambda)} \int A(r, V) \exp(-\beta(E(r, V; \lambda) + PV)) \text{drdV} \\ Q(\lambda) &= \int \exp(-\beta(E(r, V; \lambda) + PV)) \text{drdV}\end{aligned}\tag{2.5}$$

where  $A$  is the observable,  $r$  a given molecular configuration in a periodic simulation cell,  $\lambda$  the force field parameter,  $E$  the potential energy,  $\beta \equiv (k_B T)^{-1}$  the inverse temperature,  $k_B$  the Boltzmann constant,  $T$  the temperature,  $P$  the pressure,  $V$  the volume,  $Q$  the isothermal-isobaric partition function, and the angle brackets with a  $\lambda$  subscript represent an ensemble average in the thermodynamic ensemble of the force field parameterized by  $\lambda$ . In practice, this integral is evaluated numerically using molecular dynamics or Monte Carlo simulation in the NPT ensemble. Since the expression for  $A$  depends on  $\lambda$  only through the potential energy  $E$ , we can differentiate Equation 2.5 analytically:



$$\begin{aligned}
\frac{d}{d\lambda} \langle A \rangle_\lambda &= \frac{1}{Q(\lambda)} \int A(r, V) \exp(-\beta(E(r, V; \lambda) + PV)) \left( -\beta \frac{dE(r, V)}{d\lambda} \right) dr dV \\
&\quad - \frac{1}{Q(\lambda)^2} \frac{dQ}{d\lambda} \int A(r, V) \exp(-\beta(E(r, V; \lambda) + PV)) dr dV \\
&= -\beta \left( \langle A \frac{dE}{d\lambda} \rangle_\lambda - \langle A \rangle_\lambda \langle \frac{dE}{d\lambda} \rangle_\lambda \right)
\end{aligned} \tag{2.6}$$

The potential energy derivative  $\langle dE/d\lambda \rangle$  is evaluated by numerically differentiating the potential energies at the sampled structures. Equation 2.6 provides a way to evaluate the parametric derivative of thermodynamic properties without running additional sampling simulations, though the derivative of any observable always manifests as a higher order correlation function and has a larger statistical error than the observable itself. This equation may be directly applied to obtain derivatives of ensemble-averaged observables with implicit parametric dependence through the thermodynamic ensemble, such as the density. Equation 2.6 is easily extensible to obtain derivatives of observables with explicit parameter dependence, such as the enthalpy; derivatives for higher-order thermodynamic response properties such as the dielectric constant are obtained using the chain rule. We omit the calculation of second parametric derivatives for reasons of computational cost and statistical noise, relying instead on the least-squares form of the objective function and the Gauss-Newton approximation to obtain the Hessian.

The problem of overfitting is treated by regularization via a penalty function, which corresponds to imposing a prior distribution of parameter

probabilities in a Bayesian interpretation. The prior widths reflect the expected variations of the parameters during the optimization. We used a Gaussian prior distribution, corresponding to a parabolic penalty function in parameter space centered at the original AMOEBA03 parameter values with the chosen force constants. Since the various parameters have different physical meanings (e.g. vdW well depth, O-H bond length), each parameter type was assigned its own prior width.

We ran the optimization until fluctuations from numerical noise prevented further improvement. The calculation converged to within the statistical error after about 15 nonlinear iterations, though we performed several optimizations with different choices of weights for the reference data and prior widths before arriving at the final answer.

### **2.2.3 Computational Details**

#### **2.2.3.1 Parameterization Calculations**

ForceBalance carried out the condensed phase simulations in the optimization by interfacing with OpenMM 6.1,[65, 66] a GPU-accelerated molecular dynamics software package with an extensively validated implementation of AMOEBA03, which provides a speedup of an order of magnitude over the reference implementation in TINKER 6.3. At each optimization step, the set of 36 simulations at different phase points (32 temperatures at 1.0 atm pressure plus 4 pressures at 298.15 K temperature), is performed simultaneously on multiple nodes in a GPU cluster; the Work Queue library allows ForceBal-

ance to act as a distributed computing server and coordinate many OpenMM simulations running on multiple compute nodes in different physical locations. Finally, the data from the finished simulations was analyzed using the multistate Bennett acceptance ratio estimator (MBAR),[267] which allows each simulation to contribute to the estimated properties of all other simulations. This combination of methods allowed us to optimize the condensed phase properties very accurately. Due to the non-overlapping features of the simulation codes, we combined OpenMM 6.1 and TINKER 6.3 during the optimization to evaluate quantities for comparison with the ab initio and gas phase reference data, using OpenMM to evaluate the potential energies and forces, and TINKER to evaluate the binding energies and monomer properties.

### 2.2.3.2 Validation calculations

The validation calculations were conducted using the TINKER 6.3 modeling package. PME summation was used to handle the electrostatic interactions (real-space cutoff at 7 Å) and the atom-based switching window was applied to restrict the vdW interactions with a cutoff of 12 Å. During the MD simulation of NPT ensembles, we utilized the Nose-Hoover algorithm to integrate the equation of motion and control pressure and temperature.[192]

**Diffusion.** The diffusion coefficient is typically computed from the slope of the mean-square displacement as a function of time, averaged over the MD trajectories of individual particles:

$$D_{\text{PBC}} = \lim_{t \rightarrow \infty} \frac{1}{6} \frac{d\langle |r(t) - r(0)|^2 \rangle}{dt} \quad (2.7)$$

Yeh and coworkers[329] showed that for a system of nearly 2000 water molecules in a cell with periodic boundary conditions (PBC), the diffusion coefficient could be underestimated by around 10%; correcting for such systematic errors is particularly crucial in comparisons of simulation to experiment when transport properties are used to assess interaction potentials.[205, 229, 315] A correction to the system-size dependence was proposed:

$$D_0 = D_{\text{PBC}} + \frac{2.837297k_B T}{6\pi\eta L} \quad (2.8)$$

where  $L$  is the length of the cubic simulation box,  $D_{\text{PBC}}$  is the diffusion coefficient calculated in the simulation,  $k_B$  the Boltzmann constant,  $T$  the absolute temperature, and  $\eta$  the shear viscosity.

To obtain the system-size independent diffusion coefficient  $D_0$ , I calculated the diffusion coefficients  $D_{\text{PBC}}$  under periodic boundary conditions with  $N = 216, 343, 512, 1000, 1600$  and  $2500$  water molecules. The production time of the simulations is 6 ns ( $N = 512$ ), 5 ns ( $N = 1000, 1600$ ), or 3 ns ( $N = 2500$ ). In practice, each MD trajectory was divided into 500ps blocks from which the water diffusion coefficient was evaluated. The final  $D_{\text{PBC}}$  was computed as the average over these blocks. The size-independent diffusion coefficient  $D_0$  was obtained by fitting a straight line to  $D_{\text{PBC}}$  vs.  $1/L$  and extrapolating to  $1/L = 0$ .

**Viscosity.** No significant system-size dependence of viscosity was observed in computer simulations of LJ fluids.[18, 333] Here I used the slope of Equation 2.8 to obtain the average viscosity. To estimate the error bar for

size-independent viscosity and self-diffusion coefficient, the size-dependent diffusion coefficient for each box size was computed from a randomly selected 500ps block of trajectory. These diffusion coefficients for the 6 box sizes were then combined to compute the size-independent viscosity and diffusion coefficient according to Equation 2.8. The above process was repeated for 16 times to calculate the standard error.

## 2.2.4 Results and Discussion

### 2.2.4.1 Optimized Parameters

Parameter name	Units	uAMOEBA	AMOEBA03	iAMOEBA
O–H equilibrium bond length	Å	0.949 9	0.957 2	0.958 4
O–H bond force constant	kcal/mol Å <sup>-2</sup>	557.55	529.60	557.63
H–O–H equilibrium angle	deg	105.95	108.50	106.48
H–O–H angle force constant	kcal/mol Å <sup>-2</sup>	48.93	34.05	49.90
H–H Urey–Bradley length	Å	1.516 8	1.553 7	1.535 7
H–H Urey–Bradley force constant	kcal/mol Å <sup>-2</sup>	-9.74	38.25	-10.31
Oxygen vdW sigma	Å	3.755 3	3.405 0	3.645 3
Oxygen vdW epsilon	kcal/mol	0.142	0.110	0.197
Hydrogen vdW sigma	Å	...	2.655	...
Hydrogen vdW epsilon	kcal/mol	...	0.013 5	...
Hydrogen vdW reduction factor	none	...	0.91	...
Oxygen charge	<i>e</i>	...	-0.519 66	-0.594 02
Oxygen dipole Z-component	<i>e</i> bohr	0.708 89	0.142 79	0.088 48
Oxygen quadrupole XX-component	<i>e</i> bohr <sup>2</sup>	2.080 11	0.379 28	0.226 18
Oxygen quadrupole YY-component	<i>e</i> bohr <sup>2</sup>	-2.088 28	-0.418 09	-0.322 44
Oxygen quadrupole ZZ-component	<i>e</i> bohr <sup>2</sup>	0.008 17	0.038 81	0.096 26
Hydrogen charge	<i>e</i>	...	0.259 83	0.297 01
Hydrogen dipole X-component	<i>e</i> bohr	...	-0.038 59	-0.093 91
Hydrogen dipole Z-component	<i>e</i> bohr	...	-0.058 18	-0.125 60
Hydrogen quadrupole XX-component	<i>e</i> bohr <sup>2</sup>	...	-0.036 73	0.187 54
Hydrogen quadrupole YY-component	<i>e</i> bohr <sup>2</sup>	...	-0.107 39	0.021 74
Hydrogen quadrupole XZ-component	<i>e</i> bohr <sup>2</sup>	...	-0.002 03	-0.036 35
Hydrogen quadrupole ZZ-component	<i>e</i> bohr <sup>2</sup>	...	0.144 12	-0.209 28
Oxygen polarizability	Å <sup>3</sup>	1.720 9	0.837 0	0.806 4
Hydrogen polarizability	Å <sup>3</sup>	...	0.496 0	0.504 8
Polarization damping factor	Å <sup>-1</sup>	0.390	0.390	0.236

Figure 2.13: Parameters for uAMOEBA water model.

The final parameters for the new uAMOEBA, AMOEBA03 and iAMOEBA models are compared in Figure 2.13. The first six rows contain the intramolecular parameters for uAMOEBA. The equilibrium bond length was set to 0.9499

Å, which is slightly shorter than the experimental value of 0.9572 Å.[82] The ideal bond angle parameter of  $105.9^\circ$  is reduced by  $1 - 2^\circ$  from the iAMOEBA and AMOEBA03 values. The slightly increased value from the experimental gas-phase angle of  $104.52^\circ$  is necessary to reproduce the experimental liquid properties such as the dielectric constant[101, 116], [244] (Figure 2.17). The three force constants for the valence terms were fit to reproduce the experimental gas-phase vibrational frequencies of the water monomer.[54] The bond, angle and the Urey-Bradley force constants for uAMOEBA are essentially unchanged from the iAMOEBA values.

In uAMOEBA the non-bonded interaction sites on the hydrogen atoms are removed. The repulsion-dispersion parameters (vdW radius and well-depth), permanent molecular multipole (dipole and quadrupole) moments, isotropic molecular dipole polarizability, and a single vdW interaction sites are assigned to oxygen, which is slightly shifted from the molecular center of mass.

The next five rows contain the vdW parameters, followed by twelve rows containing the permanent multipole parameters. uAMOEBA has a larger vdW radius and well depth compared to AMOEBA03, which is largely due to the removal of vdW interaction sites from hydrogen. With no partial charges in this model, the electrostatic representation relies on high order molecular moments, which are significantly different from the atomic multipole moments of AMOEBA03 or iAMOEBA. In this case, it is more meaningful to compare molecular properties as described below.

### 2.2.4.2 Fitted gas phase water properties

Property		Expt.	QM	uAMOEBA	AMOEBA14	iAMOEBA
Vibrational frequencies ( $\text{cm}^{-1}$ )	Antisymmetric stretch	3755	...	3754	3755	3755
	Symmetric stretch	3656	...	3658	3656	3656
	Bend	1594	...	1594	1594	1594
Dipole (D)	$d_z$	1.855	1.840	1.802	1.808	1.864
Quadrupole ( $\text{D} \text{ \AA}^2$ )	$Q_{xx}$	2.63	2.57	2.913	2.626	2.584
	$Q_{yy}$	-2.50	-2.42	-2.693	-2.178	-2.178
	$Q_{zz}$	-0.13	-0.14	-0.220	-0.045	-0.406
Polarizability ( $\text{\AA}^3$ )	$\alpha_{xx}$	1.528	1.47	1.721	1.767	1.816
	$\alpha_{yy}$	1.415	1.38	1.721	1.308	1.816
	$\alpha_{zz}$	1.468	1.42	1.721	1.420	1.816

Figure 2.14: Gas phase monomer properties of the uAMOEBA, AMOEBA14 and iAMOEBA models compared with experiment, evaluated at the energy-minimized geometry. The molecular multipole moments were evaluated at the center of mass of water molecule.[48]; bReference[293]; cReference[214]; dReference[68]; eReference[286]; fReference[96].

Recently, Abascal and Vega pointed out that water multipole moments, specifically quadrupole moments, are crucial for capturing water properties from the vapor and liquid to solid phases accurately in multi-site models.[291] In Figure 2.14, a comparison of the experimental, ab initio QM, and calculated molecular dipole moments, quadrupole moments and polarizability of an isolated water molecule at equilibrium geometry is given. The uAMOEBA water monomer possesses a molecular dipole of 1.801 Debye, similar to that of the all-atom AMOEBA14 water model with revised parameters (1.808 Debye). The  $yy$  and  $zz$  components of the uAMOEBA molecular quadrupole moments are in better agreement with experimental values than the previous AMOEBA14



or iAMOEBA model. The isotropic molecular polarizability of the uAMOEBA model is noticeably greater than the AMOEBA14 or experimental value but slightly lower than that of iAMOEBA.

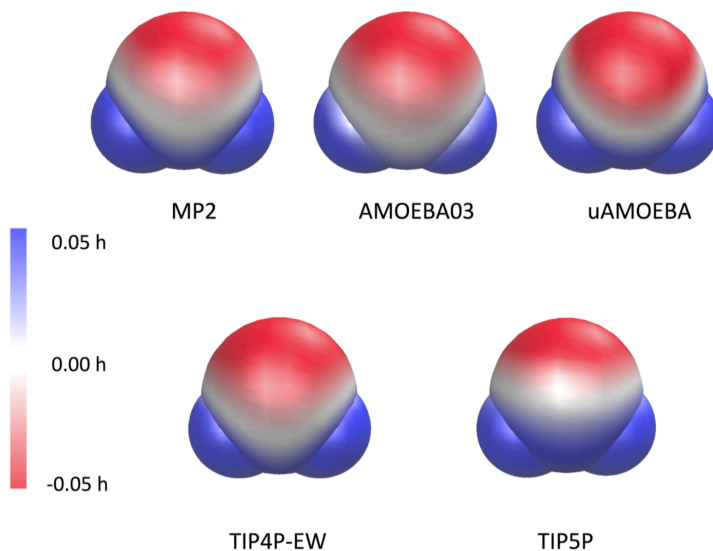


Figure 2.15: Electrostatic potential plotted on the vdW surface, with blue representing 0.05 h and red -0.05 h. The MP2 result was obtained using the 6-311++G (2d, 2p) basis set.

Figure 2.15 shows the plots of the electrostatic potential of different water models on the vdW surface. The water structure (O and H) used in calculations was based on the MP2/cc-pVTZ optimized geometry. In general, both uAMOEBA and AMOEBA03 compare well with the MP2/6-311++G(2d,2p) ESP. The similarity between the uAMOEBA and MP2 potentials around H is notable given that uAMOEBA has no electrostatic parameters on H. However, uAMOEBA is slightly more negative around the oxygen site; this is also

possibly a consequence of the missing hydrogen sites, which leads to larger multipole moments on oxygen and stronger electrostatic interactions at short range. This may also explain the slightly over-structured RDF plot around the second solvation shell, due to the stronger electrostatic interaction. The fixed charge TIP4P-EW model displays a very reasonable ESP surface while the TIP5P model seems not negative enough around the O. Nonetheless, due to the lack of explicit polarization, both TIP4P-EW and TIP5P models give a water dipole moment of 2.3 D, about 30% higher than the experimental or uAMOEBA/AMOEBA03 value for a gas-phase water monomer (Figure 2.14).

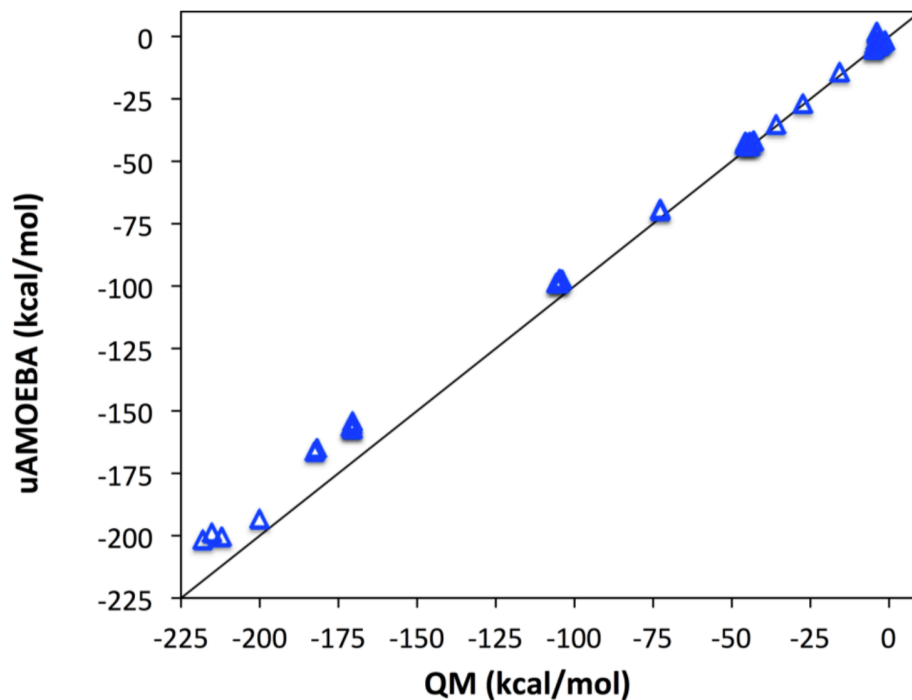


Figure 2.16: Cluster energy of gas phase geometry-optimized clusters ranging from size 2-20.

The interaction energies of water clusters ranging from dimers to clusters of 20 water molecules are shown in Figure 2.16. The predicted cluster energies are in generally good agreement with ab initio QM results, with a RMSE value of 0.85 kcal/mol per molecule and a correlation coefficient  $R^2$  of 0.988. The error increases for larger clusters. Despite the overall success, uAMOEBA has trouble with certain molecular orientations due to its isotropic, spherical nature. For example, it finds 1.056 and 1.841 kcal/mol for the Smith5 and Smith6 dimer interaction energies, while the experimental value is  $\sim -4$

kcal/mol. Besides, clusters provide a critical calibration for the increasing importance of polarization as one move from the gas-phase toward bulk phases. uAMOEBA is able to accurately reproduce the optimal structures of the water clusters; The RMSDs to the reference QM-optimized structures are around 0.15 Å for all of the clusters in the parameterization data set except a couple of hexamers and one eleven-mer.

### 2.2.4.3 Fitted liquid water properties

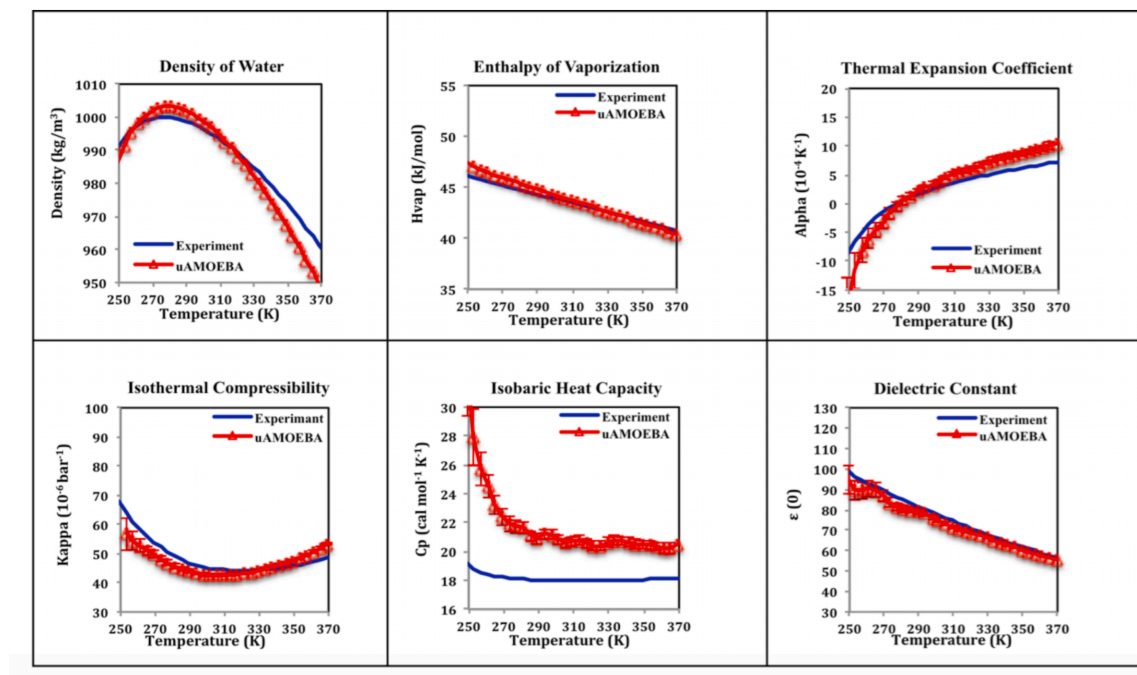


Figure 2.17: Thermodynamic properties of uAMOEBA liquid water as a function of temperature. Error bars indicate one standard error.

Figure 2.17 shows the temperature dependence of thermodynamic properties water simulated using uAMOEBA and compared to experiment: the density, enthalpy of vaporization, thermal expansion coefficient, isothermal compressibility, isobaric heat capacity and dielectric constant, which are included as part of the parameterization data set. Overall the coarse-grained uAMOEBA model is able to describe the liquid properties as well as the previous AMOEBA03 models and other non-polarizable and polarizable models in the literature (Figure 2.21).

The enthalpy of vaporization, thermal expansion coefficient, isothermal compressibility and dielectric constant all show excellent agreement with experiment over a range of temperatures after the parameter fit. The density of liquid uAMOEBA shows a correct maximum at 277 K. Small deviations in density of up to 1.4% are observed at very high temperatures (373.15 K). As the AMOEBA03 model, uAMOEBA overestimates the water heat capacity at room temperature by 3 cal/mol  $K^{-1}$ , similar to the other two polarizable models that have reported  $C_p$ , GCPM (22.5) and SWM6 (22.0). This deficiency is likely due to the approximated quantum correction to the heat capacity applied to the classical, flexible model.[118, 208, 244]

The dielectric constant of water is a critical property that is tightly coupled to the electrostatic model. However, the evaluation of dielectric constant by computer simulation is difficult due to the slow convergence near ambient conditions[119] and its dependence on the long-range interactions,[101] as well as the H-O-H angle.[116] Previous model shows that the non-polarizable

models tend to underestimate the dielectric constant (68 for SPC/E[116] and 62 for TIP4P-Ew[149]), likely due to the fixed atomic charges.[284] Our simulated value for the dielectric constant,  $78.4 \pm 1$ , matches perfectly with the experimental measurement of 78.5 at 298 K. The use of quadrupole moments and incorporation of many body polarization, even though at a coarse-grained molecular level, seems sufficient to capture the dielectric response of water.

In liquid phase, the instantaneous water dipole moment according to DFT simulations ranges between 2.6 and 3.0 Debye at room temperature.[124] The average molecular dipole moment predicted by uAMOEBA, including both the permanent and induced components, is  $2.80 \pm 0.19$  Debye, which is consistent with the 2.78 Debye given by the all-atom AMOEBA03 model. Due to the induced dipoles, the liquid phase principle molecular quadrupole moments, located at the water center of mass, changed slightly to (2.962, -2.645, -0.317).

#### **2.2.4.4 Validation of uAMOEBA**

In this section, the uAMOEBA model is validated by predicting several properties not used in parameterization, including the radial distribution function (RDF), O-O-O angle distribution, self-diffusion coefficient, viscosity and interaction energy with molecules other than water.

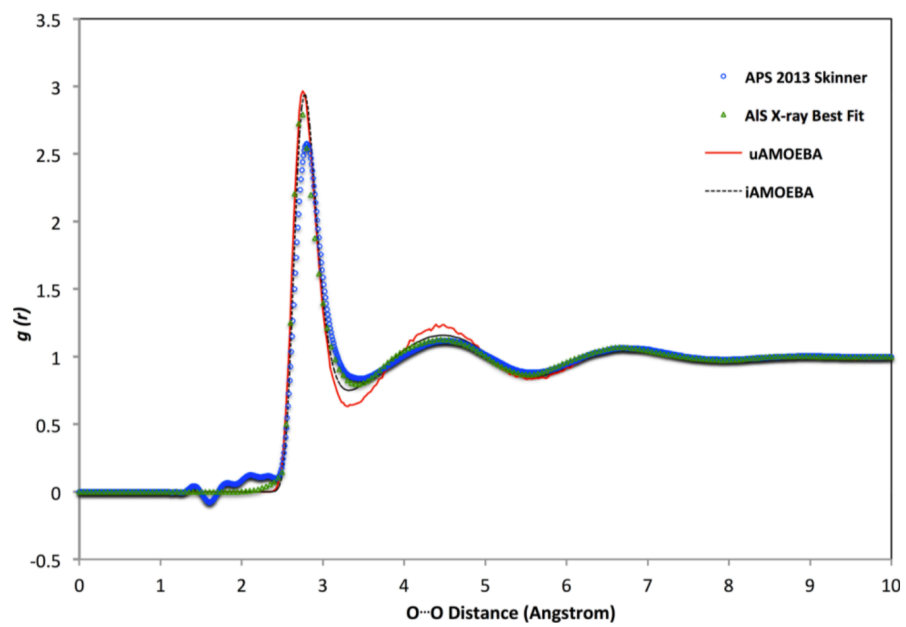


Figure 2.18: The oxygen-oxygen RDF curves of the uAMOEBA water model, compared with experimentally derived RDFs.

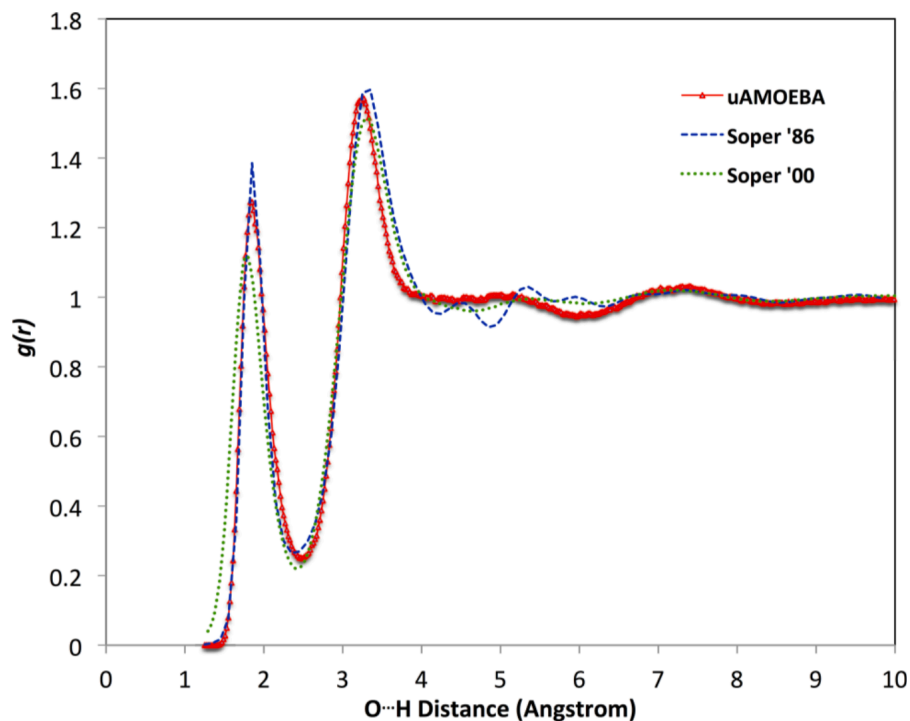


Figure 2.19: The oxygen-hydrogen RDF curves of the uAMOEBA water model, compared with experimentally derived RDFs.

**2.2.4.4.1 Radial Distribution Function** To characterize the liquid structure, the O-O and O-H radial distribution function were sampled from the NPT molecular dynamics simulations. As showing in Figure 2.18, the O-O RDF displays two well-defined peaks, similar to the experimentally derived RDFs from X-ray scattering data taken by Hura and co-workers using the Advanced Light Source (ALS)[268], more recently by Skinner and co-workers using the Advanced Photon Source (APS)[132] and simulations using the 2013 iAMOEBA model.[302] The positions of the first and second peaks of the



uAMOEBA  $g_{OO}(r)$  agree very well with the experimental data, especially the ALS data. Like iAMOEBA, the position of the first trough is slightly shifted to the left. The first peak height is almost identical to that of iAMOEBA, both similar to the ALS RDF. The second peak height is notably higher than the rest, and accordingly the first trough is 0.1 lower than the experimental RDF, which suggests that the second shell of uAMOEBA water is somewhat over structured. The positions of the first and second peaks of uAMOEBA  $g_{OH}(r)$  show excellent agreement with the experimental curve (Figure 2.19). The first peak appears around 1.9 Å, which matches the reported the hydrogen bond length (1.5–2.5 Å[271]). The first peak of the uAMOEBA  $g_{OH}(r)$  is lower than the Soper 1986 data[270] but higher than the Soper 2000 results.[295] Features in the first trough and second peak are similar to the Soper 1986 data.

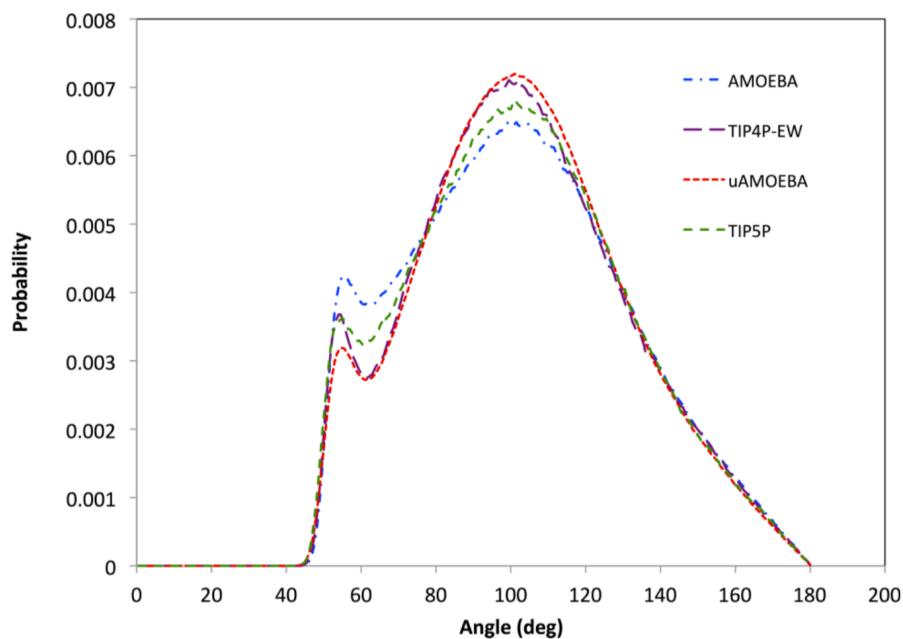


Figure 2.20: The O-O-O angle distributions of uAMOEBA, AMOEBA03, TIP4P-Ew, and TIP5P water models.

**2.2.4.4.2 O-O-O Angle Distribution** The O-O-O angle distributions sampled using uAMOEBA, AMOEBA03, TIP4P-EW, and TIP5P models are showing in Figure 2.21. From MD trajectories, I computed the O-O-O angle distribution within  $3.4\text{\AA}$  of each oxygen atom. Overall, the O-O-O angle distribution in these models suggests a tetrahedron-like structure, as the maximum probability appears around 101-105 degree. All models display a small shoulder at 55-58 degree, indicating a fifth atom entering the first solvation shell. This is strong evidence that the uAMOEBA model can describe hydrogen bonding as well as the other all-atom models. Note that the uAMOEBA

profile is particularly similar to that of TIP4P-Ew.

Property	Expt.	TIP3P	TIP4P-Ew	SWM4-NDP	TTM3-F	GCPM	SWM6	BK3	AMOEBA03	uAMOEBA
Enthalpy of vaporization/kcal mol <sup>-1</sup>										
$\Delta H_{vap}$	10.52	10.45	10.583	10.44	11.40	11.30	10.52	10.94	10.48	10.599 (7)
Density of water/g cm <sup>-3</sup>										
$\rho$	0.997	0.982	0.995	0.994	0.994	1.007	0.996	0.997	1.000	0.999 (1)
Isothermal compressibility/10 <sup>-6</sup> bar <sup>-1</sup>										
$\kappa$	45.3	57.4	48.1	...	...	...	...	44.4	66.0	42.7 (5)
Isobaric heat capacity/cal/mol K <sup>-1</sup>										
$C_p$	18.0	16.8	19.2	...	...	22.5	...	22.0	21.3	21.2 (3)
Static dielectric constant										
$\epsilon$	78.5	94.0	63.90	78.0	67.7	84.0	78.1	79.0	81.4	76.3 (12)
Thermal expansion coefficient/10 <sup>-4</sup> K <sup>-1</sup>										
$\alpha$	2.56	4.10	3.40	...	...	4.20	...	3.01	1.90	3.38 (26)
Self-diffusion coefficient/10 cm <sup>2</sup> s <sup>-1</sup>										
$D_0$	2.30	6.14	2.70	2.85	2.37	2.26	2.14	2.28	2.00	2.41 ± 0.03
Viscosity/cP										
$\eta$	0.896	0.321	0.573	0.66	...	...	0.87	0.95	1.08	0.72 ± 0.05

Figure 2.21: Comparison of experimental and simulated liquid data from different water models (T = 298.15 K, P = 1 atm). Numbers in parentheses include one standard error in terms of the least significant digit.

**2.2.4.4.3 Self-diffusion Coefficient and Viscosity** The size-independent diffusion coefficient  $D_0$  and viscosities  $\eta$  are summarized in Figure 2.21. Polarizable water models generally produce reasonable diffusion constants compared to nonpolarizable ones[101], except for SWM4-NDP that is slightly underpolarized. The strong dependence of the calculated diffusion constant on the system size was observed. With a small box of 216 water molecules, the simulated  $D$  using uAMOEBA is  $1.95 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ , which increases to  $2.21 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  when the simulation box contains 2500 molecules. By using extrapolation, the system-independent self-diffusion constant by uAMOEBA

is  $2.41 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ , in excellent agreement with the experimental value. For the viscosity, which is considered system size-independent, most models underestimate the viscosity except the BK3 and AMOEBA03 model. The deviation in viscosity given by the uAMOEBA model is similar to AMOEBA03 models and larger than those of SWM6, BK3 and iAMOEBA (0.85 cP).[302]

### 2.2.4.5 Comparison between Coarse-grained and All-atom AMOEBA

Dimer	Dimer structure	uAMOEBA $\Delta E_{\text{int}}$	AMOEBA03 $\Delta E_{\text{int}}$	SAPT2+/CBS $\Delta E_{\text{int}}$
(a) Water-water		-5.16 (0.13)	-4.99 (0.06)	-4.92
(b) Peptide-water		-5.35 (0.10)	-5.16 (0.13)	-5.14
(c) MeOH-water		-5.61 (0.07)	-4.83 (0.05)	-5.03
(d) Benzene-water		-3.14 (0.25)	-3.16 (0.23)	-3.25
(e) Water-MeNH <sub>2</sub>		-6.69 (0.13)	-8.53 (0.08)	-7.18
(f) Water-peptide		-6.48 (0.35)	-8.28 (0.26)	-8.04
(g) Water_pyridine		-6.46 (0.10)	-7.26 (0.05)	-7.15
(h) Water-MeOH		-4.84 (0.24)	-6.77 (0.19)	-5.61

Figure 2.22: Dimer equilibrium interaction energy between water and small molecules. Compared results from uAMOEBA, AMOEBA03 water model and SAPT2+/CBS (calculated using PSI4[285]). For the water-water dimer, AMOEBA03 or uAMOEBA was used for both molecules. For the other heterodimers, the water was modeled with either uAMOEBA or AMOEBA03 (as labeled in the 3rd and 4th column) while the other molecule was always modeled with AMOEBA03. The structural RMSD values in parentheses represent the structural different from the MP2/cc-pVTZ optimized structures.[9]

**2.2.4.5.1 Transferability Validation** To investigate the transferability of the uAMOEBA model and its ability to interact with the all-atom AMOEBA03 model, I have computed the dimer equilibrium interaction energy between water and several small molecules (Figure 2.22). For the water-water dimer, AMOEBA03 or uAMOEBA was used for both molecules. For the other heterodimers, the water was modeled with either uAMOEBA or AMOEBA03 (as labeled in the 3rd and 4th column in Figure 2.22) while the other molecule was always modeled with AMOEBA03. These comparisons allow us to understand the potential differences between the AMOEBA03 and uAMOEBA water models as solvent. Starting from QM optimized structures, each dimer was optimized using force fields to obtain the corresponding interaction energy. These molecules were chosen to test performance of uAMOEBA water model as the hydrogen bond donor, acceptor and interacting with aromatic benzene. The overall trend given by uAMOEBA model is in good agreement with the SAPT2+/CBS data (calculated using PSI4[285]). The correlation coefficient ( $R^2$ ) between uAMOEBA and SAPT2+/CBS results is 0.83 while the correlation coefficient between all-atom AMOEBA03 and SAPT2+/CBS is 0.92. As a hydrogen bond acceptor, uAMOEBA performs equally well compared to allatom AMOEBA03, even in the more complicated peptidewater interaction (-5.35 for uAMOEBA, -5.16 for AMOEBA03 and -5.14 kcal/mol for SAPT2+/CBS). In addition, uAMOEBA accurately captures the OH- $\pi$  interaction when facing the aromatic molecules. I also tested the hybrid water dimer where one of the water is described AMOEBA03 while the other is

uAMOEBA. The dimer interaction energy is -4.11 kcal/mol when uAMOEBA is the H-bond donor in the dimer, and -5.90 kcal/mol when uAMOEBA is the H-bond acceptor. The average of the two is -5.0 kcal/mol, matching very well with AMOEBA03 or QM values. Overall the uAMOEBA water model performs reasonably when replacing the AMOEBA03 model in the hybrid uAMOEBA-AMOEBA03 application even though there are no electrostatic or vdW parameters on the hydrogen atoms at all.

No. of molecules	AMOEBA03 models/1000 force evaluations	uAMOEBA models/1000 force evaluations	Ratio
216	59.6	12.3	4.8
512	95.6	22.4	4.3
1000	97.3	35.2	2.8
1600	155.4	59.9	2.6
2500	245.2	88.2	2.8
4000	386.0	132.1	2.9

Figure 2.23: Efficiency test for the uAMOEBA water model. For all simulations, the vdW cutoff was set to 12 Å and Ewald real-space cutoff was 7 Å. For the first 3 systems of 512 molecules or less, no neighbor-list was used for the vdW calculation. Computer hardware: Intel E5-2697 v2 @ 2.70GHz, 8 threads.

**2.2.4.5.2 Computational Efficiency** The main motivation for developing uAMOEBA is to improve computational efficiency by reducing the number of nonbonded interaction sites. I compared the simulation time between the uAMOEBA model and the all-atom AMOEBA03 model in 1000 steps of gradient evaluation (Figure 2.23). The same simulation settings were used for both

models and the only difference is the model parameters that distinguish the one-site model from the three-site model. The efficiency ratio is defined by the simulation time from the all-atom AMOEBA03 model divided by that of the one-site model. For relatively small systems containing less than 1000 atoms, the speed up is almost a factor of 5, while for large water box the improvement is about a factor of 3; this is expected as the computational cost of the PME method scales as  $N \log(N)$ .

### 2.2.5 Conclusions

Advancement in molecular simulation relies on accurate potential models and efficient sampling methods. This work presents the development of a coarse-grained polarizable water model, uAMOEBA, where all nonbonded interactions, including the polarizability, are placed on the oxygen atom. The parameters of this model are determined from a wide range of *ab initio* and experimental data using the automated ForceBalance procedure. The model and parameters are validated by comparing with additional *ab initio* and experimental results, including liquid structural properties, self-diffusion coefficient, shear viscosity, and interaction energies with other small organic molecules. Overall, uAMOEBA shows good transferability between gas and liquid phases, polar and nonpolar environments, most likely because of the incorporation of molecular polarizability. The water structural and dynamic properties given by uAMOEBA are in very good agreement with those derived from all-atom AMOEBA03 model and experiments. The dimer interaction energy between



AMOEBA03 small molecules and uAMOEBA water are mixed together also show a satisfying trend in comparison with all-AMOEBA03 and SAPT results. Meanwhile, the computational efficiency is improved by a factor of three compared to atomistic AMOEBA03. uAMOEBA has been expected to be a useful solvent model in simulations of biological systems such as proteins and nucleic acids and it can be readily combined with the existing all-atom polarizable protein force field.[265]

## Chapter 3

# Applications of AMOEBA to Protein-ligand Recognition

Molecular recognition between biomolecules and ligands is very specific in living cells. The performance of many biochemical processes and cellular mechanisms are dependent upon complex but specific non-covalent intermolecular interactions underlying a multitude of functions.[114] For example, proteins utilize negatively charged anions, particularly phosphate-derived groups, in signal transduction, metabolism, biosynthesis, gene regulation, and many other biological functions.[327] While there is an abundance of experimental structure information about protein-ion interactions, general understanding of the recognition mechanism for specific ions remains elusive. For example, it is unclear how proteins recognize and utilize the anionic phosphate groups that are ubiquitously involved in signaling proteins such as kinases and phosphatases.[115]

Due to the lack of effective sampling methods and accurate force fields that accurately represent molecular interactions,[163] computational and theoretical studies of ion-protein systems are challenging and our ability to make accurate predictions of ion binding sites, using statistical or physical ap-

proaches, is very limited.[297] In addition, progress regarding phosphate and anion binding to proteins has been limited; likewise, the energetic and thermodynamic driving forces of these binding events remain poorly understood.[259] The complex electrostatic interactions between anions and other biomolecules cannot be captured by the simple fixed-charge model prevalent in traditional force fields. For example, the binding free energy of  $\text{Mg}^{2+}$  to phosphate (Pi) in the solution given by fixed-charged model is overestimated by order of magnitude compared to the experiment. Such large error is mainly due to not explicitly accounting for the polarization of the phosphates or charge penetration/charge transfer in response to  $\text{Mg}^{2+}$ .[212] Quantum mechanical calculation can provide accurate descriptions of such interactions but is mostly restricted to small molecule compounds in predetermined binding sites. In this chapter, I will present the application of AMOEBA force field on highly charged molecule phosphate-protein binding and host-guest systems.<sup>1</sup>

In the first project, the controversial mechanism of PBP-phosphate recognition is discussed. Based on the similar binding affinities at acidic and basic pHs, it is believed that the hydrogen network in the binding site is flexible to adapt to different protonation states of phosphates. However, only hydrogen (1H) phosphate was observed in the sub-angstrom X-ray structures.

---

<sup>1</sup>The vdW work (3.1) were previously published.[234] I optimized the two phosphate models and constructed testing dataset to validate them against QM dimer interaction energies. I also run MD simulation to study the effect of buffer agent. Z. F. Jing, K. Dalby and P. Y. Ren helped revising the paper. Besides, the host-guest binding (3.2) were previously published .[17]. I run all the simulations, analyzed the effect of Hydrogen bonds and helped writing the paper. D. R. Bell and Z. F. Jing analyzed the motion of the ligands, the contribution of Entropy and Enthalpy. P. R. Ren helped revising the paper.

To address this inconsistency, I performed molecular dynamics simulations using the AMOEBA polarizable force field. Structural and free energy data from simulations suggested that 1H phosphate was the preferred bound form at both pHs. The binding of dihydrogen (2H) phosphate disrupted the hydrogen-bond network in the PBP pocket, and the computed affinity was much weaker than that of 1H phosphate. Furthermore, I figured that the discrepancy in the studies described above is resolved if the interaction between phosphate and the buffer agent is taken into account. The calculated apparent binding affinities are in excellent agreement with experimental measurements. Our results suggest the high specificity of PBP for 1H phosphate and highlight the importance of the buffer solution for the binding of highly charged ligands.

In the second project, a series of host-guest systems previously used in the SAMPL4 blind challenge is investigated by using molecular simulations and the AMOEBA polarizable force field. The free energy results computed by Bennetts acceptance ratio (BAR) method using the AMOEBA polarizable force field ranked favorably among the entries submitted to the SAMPL4 host-guest competition.[311] In this work, I conduct an in-depth analysis of the AMOEBA force-field host-guest binding thermodynamics by using both BAR and the orthogonal space random walk (OSRW) methods. The binding entropy-enthalpy contributions are analyzed for each host-guest system. For systems of inordinate binding entropy-enthalpy values, I further examine the hydrogen bonding patterns and configurational entropy contribution. The binding mechanism of this series of host-guest systems varies from ligand to

ligand, driven by enthalpy and/or entropy changes. Convergence of BAR and OSRW binding free energy methods are discussed. Ultimately, this work illustrates the value of molecular modeling and advanced force fields for the exploration and interpretation of binding thermodynamics.

## **3.1 Phosphate Binding Mode in Phosphate binding protein**

### **3.1.1 Introduction**

Inorganic phosphate is one of the most important nutrients for organisms. It is not only used in the biosynthesis of cellular components, such as ATP, nucleic acids, phospholipids, and protein, but also an integral part of many biological processes, including metabolism, gene regulation, and muscle contraction.[35, 98, 146, 287, 317, 327] Protein phosphorylation is a key mechanism for regulating transmembrane and intracellular signal transduction and affects every basic cellular process.[123, 180] Due to the biological importance of Pi, the transport of Pi into cells and the maintenance of proper Pi homeostasis are critical. The phosphate binding protein (PBP), an initial receptor of the phosphate-specific transport systems (Pst),[157, 240] binds phosphate with high specificity against competing anions such as sulfate.[110, 164, 237, 328] This high specificity is explained by a rich network of 12 hydrogen bonds (H-bonds) between phosphate and the binding site.[110, 240, 309] Another distinct feature of the PBP binding site is a low-barrier hydrogen-bond (LBHB) between phosphate and the Asp side chain. While its effect on protein phosphate

binding is not as significant as indicated by gas-phase calculations,[91, 308] the LBHB is responsible for discriminating arsenate and other similar tetrahedral oxyanions.[70, 240, 307].

Despite numerous studies on PBPs, questions remain about the binding mechanism. It is believed that PBP can bind to either 1H or 2H phosphate based on the observation that *E. coli* PBP has similar affinities for phosphate at pH 4.5 and 8.5.[70, 91, 109, 170, 240, 294, 309] It was suggested that the binding pocket is capable of binding to both 1H and 2H phosphate with reorientation of hydrogen atoms/H-bond network only.[240] However, sub-angstrom X-ray crystallography studies of a topologically similar protein PfluDING (*P. fluorescens* PBP) showed that only 1H phosphate was present in the binding pocket at both pH 4.5 and 8.5, suggesting that PBP should follow a similar binding mechanism.[26, 70, 307] It seems that the conserved experimental affinities and the 1H-specific binding mode in sub-angstrom crystal structures cannot be reconciled with each other.

Molecular dynamics simulations based on classical force fields have been a valuable tool for the understanding of protein-ligand binding.[61, 135, 198, 325] The popular fixed-charge force fields have been applied to model neutral ligands and monovalent ions.[61, 99, 199, 219] Highly charged species are more challenging in simulation and often treated by computationally demanding quantum mechanical methods.[62, 152, 239] Polarizable force fields offer an efficient way to model electrostatic interactions for highly charged species with high fidelity.[231, 256, 323] Here, I carried out molecular dynamics simu-

lations using the AMOEBA polarizable force field[265, 276] to investigate the phosphate binding mode of PBPs.

### 3.1.2 Computational Methods

#### 3.1.2.1 Quantum Mechanics and Molecular Mechanics

*Ab initio* quantum mechanics calculations were performed using Gaussian 09 and PSI4 program[117, 222, 285] with the following methods and basis sets. For permanent multipole fitting, the structures of the 1H and 2H phosphate models were first optimized at the MP2/6-31G\* level, then single point energy calculations at the MP2/aug-cc-pVTZ level were performed. Atomic multipole moments were initially assigned from QM electron density calculated at the MP2/6-311G\*\* level via Stones distributed multipole analysis.[106] For torsion fitting, structures with different dihedrals angles at the rotatable bonds were optimized at HF/6-31G\* level, followed by single point energy calculation at the MP2/6-311++G\*\* level. The molecular polarizability of 1H and 2H phosphate models were calculated using two methods, B97XD and MP2, each with two basis sets aug-cc-pVTZ and 6-311++G\*\*. For a series of distances, the interaction energies of 1H and 2H phosphate with water, amide dimers were calculated at MP2/aug-cc-pVTZ, MP2/aug-cc-pVQZ in PSI4 and extrapolated to the CBS level. For different protonation states, dimers structures of phosphate interacting with residue models, cropped from PDB structures, were optimized at the MP2/6-31G\* level in Gaussian, with fixed heavy atom positions to mimic the interactions distances in binding

pockets. The interaction energies were calculated via SAPT0/jun-cc-pVDZ, SAPT2+/aug-cc-pVDZ and RIMP2/aug-cc-pVTZ method in PSI4.

All molecular mechanics force field based calculations were performed using TINKER 6 Software.[276] The POTENTIAL program was used to fit electrostatic potentials around molecules. The ANALYZE program was used to calculate the total potential energy and its components, for all individual interactions and electrostatic moments of molecules, dimers and protein-ligand complexes. The MINIMIZE program was used to relax protein-ligand system. The POLARIZE program was used to compute molecular polarizabilities based on atomic polarizability parameters in the AMOEBA force field. Other useful tools including XYZEDIT, PDBXYZ, and CRYSTAL were used in setting up the MD simulation systems. All parameters for water and protein were adopted from the current AMOEBA force field.[135]



### 3.1.2.2 Absolute Binding Free Energy Calculations

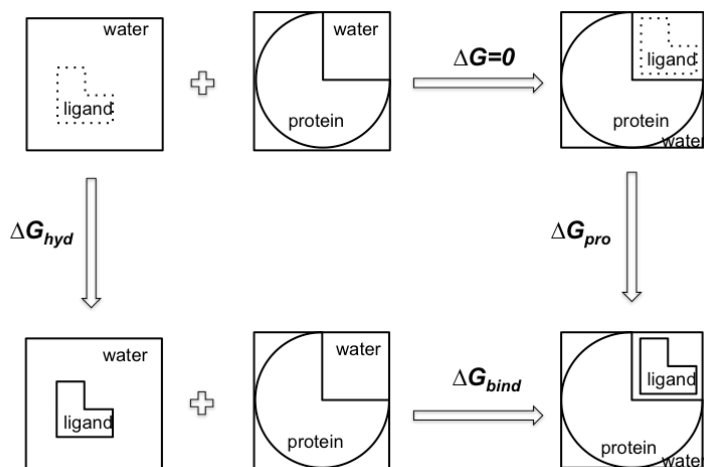


Figure 3.1: Thermodynamic cycle for calculating the binding free energy of phosphate-protein binding.

$$\Delta G_{bind} = \Delta G_{pro} - \Delta G_{hyd} \quad (3.1)$$

The absolute binding free energy ( $\Delta G_{bind}$ ) of the phosphate protein binding was calculated by the double-decoupling method. This involves disappearing the phosphate in water and in protein-phosphate complex.[135] Hence, the binding free energy can be defined as the difference between the decoupling free energies in water ( $\Delta G_{hyd}$ ) and protein environments ( $\Delta G_{pro}$ ).[22, 84] (Equation 3.1, Figure 3.1) When the ligand is completely decoupled from its environment, the ligand could jump out of the pocket, prolonging convergence. To prevent this situation, a constraint with constant spring constant, of 15 kcal/(mol/Å<sup>2</sup>)

was added in the simulations, and the effect of the constraint was subsequently removed by a correction term ( $\Delta G_{correction}$ ) defined by Equation 3.2.[84, 283]

$$\Delta G_{correction} = RT \ln \left[ C^o \left( \frac{2\pi RT}{k} \right)^{3/2} \right] \quad (3.2)$$

All molecular dynamics simulations were run using the Tinker-OpenMM package with a RESPA integrator[34], Bussi thermostat[21], Berendsen barostat,[19] and 3.0 fs time step with hydrogen-mass repartition (heavy-hydrogen keyword) on GPUs. The van der Waals interactions used a 12.0 Å cutoff, while the electrostatic interactions used a 7.0 Å cutoff. The simulation schema started with a total of 3 ns of NVT MD simulations, then the system was gradually heated from 50K to 298K exponentially. In order to relax the water in the box, position-restraints were added on protein-ligand complex during the initial heating steps. Next, NVT simulations were performed at 298K for 3 ns while gradually turning off all the position restraints on protein-ligands from a restraint constant value of 100 to 0 using an interval of 10. NPT simulations at 298K were conducted for 2ns to compute the average density or box size. The average box size from these NPT simulations was used in production MD simulations.

$$U = U(\lambda, \mathbf{R}), U_0 = U(0, \mathbf{R}), U_1 = U(1, \mathbf{R}) \quad (3.3)$$

$$\Delta G = \sum_{i=1}^{N-1} [G(\lambda_{i+1}) - G(\lambda_i)], \lambda_1 = 0, \lambda_N = 1 \quad (3.4)$$

A mixed potential was defined to calculate the free energy difference between the end states that were connected analytically (Equation 3.3). The free energy changes from one state to the other is thus given by Equation 3.4. In order to obtain free energy estimates, the Bennett acceptance ratio method[108] was applied, which required a certain value of fixed order parameter  $\lambda$  to connect the two end states. In this study the perturbation schedules contained scaling down electrostatics and scaling down van der Waals with  $\lambda$  in a series of steps: 1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.4, 0.0. For each  $\lambda$  value at 298K, I carried out 10ns NVT simulation for calculating  $\Delta G_{solv}$  and 6ns simulations for  $\Delta G_{bind}$ . Therefore, I analyzed 22 trajectories in total to obtain the absolute binding free energy for each phosphate protein system.

In this work, the structures of the PBP-phosphate complexes for the wild type and the D56N mutant were taken from the PDB (PDB code: 1IXI and 1IXH). The structures were then solvated in periodic boxes of 73.27 x73.27 x91.59 Å<sup>3</sup> with NaCl added to yield 0.15 M salt concentration. The systems were relaxed and heated before free energy simulations. The binding free energies were calculated by the double-decoupling scheme and the Bennet acceptance ratio method.[80, 135] All molecular dynamics simulations were run with the Tinker-OpenMM program on GPU.[80] The equations in Equation 3.2 were used to calculate the apparent KD that without and with buffer effect at pH 4.5 and 8.5. The pKa value is 7.21.

1. No effect of buffer solution, all are using wild type binding as examples.

$$K_D^{app} = \frac{[PBP] \times ([HPO_4^{2-}] + [H_2PO_4^-])}{([HPO_4^{2-}:PBP] + [H_2PO_4^-:PBP])} = \frac{\left(1 + \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)}{\left(\frac{[HPO_4^{2-}:PBP]}{[HPO_4^{2-}] \times [PBP]} + \frac{[H_2PO_4^-:PBP]}{[H_2PO_4^-] \times [PBP]} \times \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)}$$

$$pH = pKa + \log\left(\frac{[HPO_4^{2-}]}{[H_2PO_4^-]}\right)$$

$$\frac{[HPO_4^{2-}:PBP]}{[HPO_4^{2-}] \times [PBP]} = \exp\left(-\frac{dG_{bind}}{RT}\right)$$

$$\text{At pH 8.5, } K_D^{app} = \frac{(1+0.05)}{(1.755 \times 10^8 + 5.14 \times 10^4 \times 0.05)} = 6nM$$

$$\text{At pH 4.5, } K_D^{app} = \frac{(1+512.96)}{(1.755 \times 10^8 + 5.14 \times 10^4 \times 512.9)} = 2.55 \mu M$$

2. Considering the effect of buffer solution:

$$\text{At pH 8.5, } K_D^{app} = \frac{[PBP] \times ([HPO_4^{2-}] + [H_2PO_4^-] + [HPO_4^{2-}:Tris-H^+] + [H_2PO_4^-:Tris-H^+])}{([HPO_4^{2-}:PBP] + [H_2PO_4^-:PBP])} = \frac{\left(1 + \frac{[H_2PO_4^-]}{[HPO_4^{2-}]} + \frac{[HPO_4^{2-}:Tris-H^+]}{[HPO_4^{2-}]} + \frac{[H_2PO_4^-:Tris-H^+]}{[H_2PO_4^-]} \times \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)}{\left(\frac{[HPO_4^{2-}:PBP]}{[HPO_4^{2-}] \times [PBP]} + \frac{[H_2PO_4^-:PBP]}{[H_2PO_4^-] \times [PBP]} \times \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)} = \frac{(1+0.05+9.57+0.34*0.05)}{(175482669.6+51403.3 \times 0.05)} = 60.6 nM$$

$$\text{At pH 4.5, } K_D^{app} = \frac{[PBP] \times ([HPO_4^{2-}] + [H_2PO_4^-] + [HPO_4^{2-}:Na^+] + [H_2PO_4^-:Na^+])}{([HPO_4^{2-}:PBP] + [H_2PO_4^-:PBP])} = \frac{\left(1 + \frac{[H_2PO_4^-]}{[HPO_4^{2-}]} + \frac{[HPO_4^{2-}:Na^+]}{[HPO_4^{2-}]} + \frac{[H_2PO_4^-:Na^+]}{[H_2PO_4^-]} \times \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)}{\left(\frac{[HPO_4^{2-}:PBP]}{[HPO_4^{2-}] \times [PBP]} + \frac{[H_2PO_4^-:PBP]}{[H_2PO_4^-] \times [PBP]} \times \frac{[H_2PO_4^-]}{[HPO_4^{2-}]}\right)} = \frac{(1+512.96+0.21+0.03*512.96)}{(175482669.6+51403.3 \times 512.96)} = 2.6 \mu M$$

Figure 3.2: Apparent dissociation constant  $K_D$  Calculations without and with buffer effect at pH 4.5 and 8.5. The  $pK_a$  value is 7.21.

### 3.1.3 Parameterization Strategy and Model development

#### 3.1.3.1 Parameterization of 1H and 2H Phosphate Models

1. Bond Angle parameters were adopted from DMP model <sup>23</sup>
2. Electrostatics parameters ( <i>Distributed multiple analysis and potential field fitting at MP2/aug-cc-pVTZ</i> )
3. Polarization parameters ( <i>QM molecular polarizability at <math>\omega</math>B97XD/6-311++G**</i> )
4. Torsion parameters ( <i>Conformational energy surface at MP2/6-311++G**</i> )
5. Van der Waals parameters ( <i>QM interaction energy with water at MP2/CBS</i> )
6. Validation of the parameters QM interactions with common protein residues models at RIMP2/aug-cc-pVTZ, SAPT0/jun-cc-pVDZ, and SAPT2+/aug-cc-pVDZ levels.

Figure 3.3: Parameterization schema of 1H and 2H phosphate models.

The parameterization schema is showing in Figure 3.3. The phosphate models are represented by 4 types of atoms: phosphorus (P), double-bonded oxygen (O), single-bonded oxygen (Os), and hydrogen (H) on the single-bonded oxygen. 1H phosphate model contains a -2e charge while 2H phosphate model carries -1e charge. The electrostatic and polarization parameters were determined from the QM electrostatic potential and molecular polarizabilities at MP2/aug-cc-pVTZ and  $\omega$ B97X-D/6-311++G\*\* level, respectively. The torsional parameters were derived to reproduce the QM conformational energy profile at MP2/6-311++G\*\* level. The vdW parameters were optimized to capture the phosphate water interaction energy at different orientations calculated at the MP2/CBS level. 1H and 2H phosphate models share the same vdW parameters. To reproduce the interaction energy of the short H-bond

between phosphate and carboxylate group, pairwise vdW parameters between the carboxylate O atom and phosphate H and O atoms were necessary. The pairwise vdW parameters were optimized by fitting to interaction energies calculated by MP2/CBS and PCM. Different pairwise vdW parameters were used among 1H and 2H models. As validation, the final force field parameters of phosphate were applied to compute the total interaction between phosphate and protein residue side chain models. The parameters for Tris-H<sup>+</sup> were derived by using POLTYPE. Gaussian 09[224] and Psi4[334] were used for the QM calculations.

### 3.1.3.2 Validation of the Phosphate Models

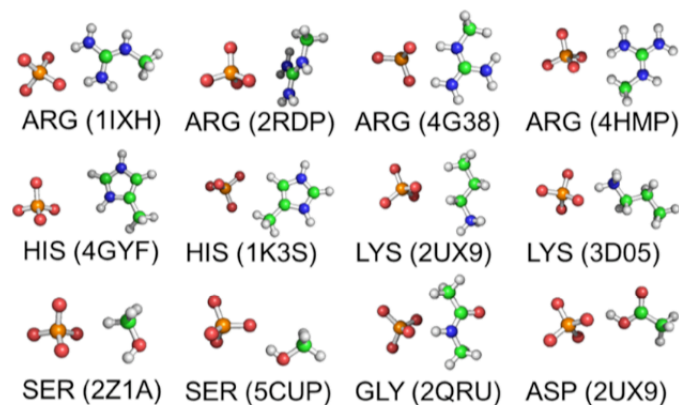


Figure 3.4: Model compounds for amino acid-phosphate interactions extracted from PDB. All 10 possible protonation states for phosphate were considered.

As validation, the final force field parameters of phosphate were applied to compute the total interaction between phosphate and protein residue

side chain models (Figure 3.4). In Figure 3.5, the total interactions of residue-phosphate calculated using AMOEBA force field were compared to QM results at different levels including RIMP2/aug-cc-pVTZ, SAPT2+/aug-cc-pVDZ, and SAPT0/jun-cc-pVDZ. Good agreement between the AMOEBA and QM results was obtained with a correlation coefficient  $R^2$  value of 0.99 across all dimers configurations and phosphate protonation states. The RMSE values are 3.4, 3.8, and 4.4 kcal/mol when comparing the total interaction energy calculated by AMOEBA and QM at RIMP2/aug-cc-pVTZ, SAPT2+/aug-cc-pVDZ, and SAPT0/jun-cc-pVDZ levels.

statistics	RIMP2	SAPT2+	SAPT0
MUE (kcal/mol)	2.613	3.068	3.232
MSE (kcal/mol)	-0.528	0.110	-1.534
RMSE (kcal/mol)	3.404	3.816	4.411
$R^2$	0.997	0.997	0.996

Figure 3.5: Performance of AMOEBA interaction energies of 120 model compounds for amino acid-phosphate dimers compared to RIMP2/aug-cc-pVTZ, SAPT0/jun-cc-pVDZ and SAPT2+/aug-cc-pVDZ results. MUE, MSE and RMSE stand for mean unsigned error, mean signed error and root mean squared error, respectively.

### 3.1.4 Results and Discussion

#### 3.1.4.1 One Hydrogen Phosphate is the Dominant Form

The molecular mechanism of phosphate-PBP binding in E.coli has been hypothesized based on X-ray structures without the protons on phosphate.[91,

110, 240] The phosphate is bound through 12 H-bonds formed with 5 backbone NH groups, 4 hydroxyl groups of serine and threonine, 2 NH groups of the Arg135 sidechain, and one oxygen of the Asp56 sidechain. 1H phosphate was described as the acceptor for 11 H-bonds and the donor for one H-bond, in which 1H phosphate shared its proton with the carboxylate sidechain of Asp56 to form low-barrier hydrogen-bond (Figure 3.7A). In the 2H phosphate binding mode, Luecke et al.[240] assumed that no drastic structural rearrangement of the binding pocket is needed. The only difference from 1H binding was that the hydroxyl group of Ser38, as the only one favorably positioned to have a proton bound to an O lone pairs, flipped to accept the second H on phosphate while donating its proton to Asp56 (Figure 3.7B). The binding pocket in the D56N mutant PBP is similar to the wild type, except for a normal hydrogen bond between Asn56 and phosphate.[91]



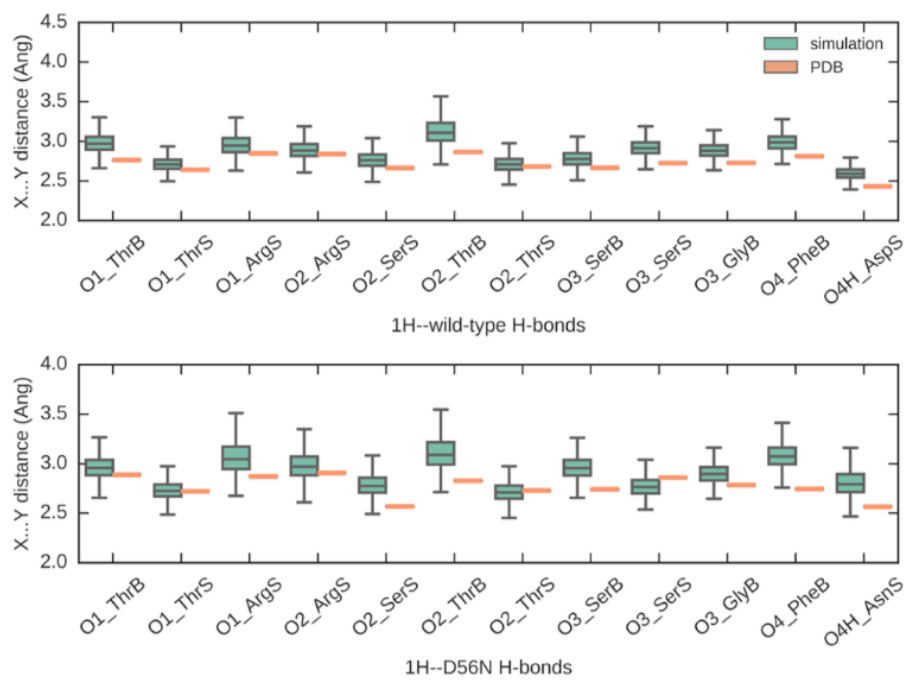


Figure 3.6: Plot of 12 hydrogen bond distances (XY) between 1H phosphate and heavy atoms in wild type (top) and D56N mutant (bottom) PBP receptors. The orange symbols represent the distances in crystal structures and the green boxes represent the average distances in simulations with standard deviations. X axis lists all interacting residues in PBP. Suffix S stands for sidechain and B stands for backbone. In simulations, 1H phosphate reproduced and maintained 12 hydrogen bonds in both binding pockets and the H-bond distances are consistent with the experimental values.

### 3.1.4.1.1 Hydrogen Bond Distances Between Phosphate and PBPs

In the 1H phosphate binding simulations, 12 H-bonds for the wild type and the mutant were maintained over the entire 18-ns simulations (Figure 3.6). The mean distances are  $2.87 \pm 0.14 \text{ \AA}$  and  $2.91 \pm 0.14 \text{ \AA}$  for wild type and the mutant PBP binding pocket, respectively, compared to experimental H-bonds

distances of  $2.72 \pm 0.11 \text{ \AA}$  and  $2.77 \pm 0.11 \text{ \AA}$ . The H-bond between Asp56 and 1H phosphate is the shortest among all H-bonds, which was captured by the simulation. On the other hand, the binding of 2H phosphate led to significant rearrangement of the H-bond network and an unstable binding pocket. In the 2H phosphate-D56N mutant complex, the mean distances of all H-bond pairs increased to  $3.23 \pm 0.37 \text{ \AA}$ . I observed relatively weak binding (discussed below) compared to 1H phosphate, with water molecules entering the binding pocket during the simulations. These water molecules replaced the binding with the Arg135 side chain and pushed Ser139, Ser 38, Gly 140, and Phe11 away from the phosphate.

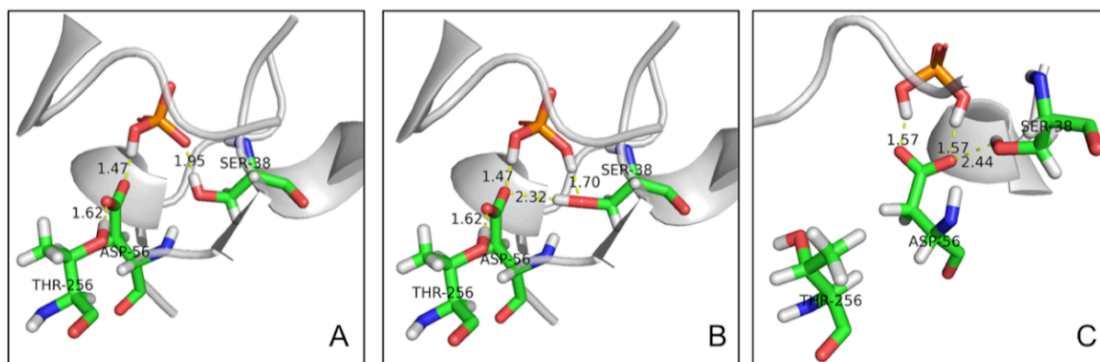


Figure 3.7: Illustration of phosphate binding with Asp56 in wild-type PBPs. A. 1H phosphate binding mode in current simulation; B. Hypothesized 2H binding mode involves flipping of the hydroxyl group of Ser38 to Asp56 to accept the second proton on phosphate;[240] C. 2H phosphate bidentate binding with Asp56 in current simulation. During the simulations, 1H phosphate kept the same binding mode as crystal structure while 2H phosphate preferred bidentate binding with Asp56, which is different from the monodentate binding hypothesized in literature.[26]

It was found that the 2H phosphate preferred bidentate binding with Asp56 in wild-type protein, in contrast to the hypothesized monodentate binding mentioned above (Figure 3.7B and C). Asp56 rotated toward the phosphate and formed two stable hydrogen bonds with mean distances of 2.58 and 2.63 Å throughout the 18-ns simulation. The bidentate binding brings the negative-charge center closer to the phosphate and disrupted the H-bonds between the phosphate and other residues, both of which would weaken the binding in the pocket. As for the 2H phosphate binding with the mutant protein, I observed water entering the pocket from the Arg135 side. As a result, the distances of all native H-bond pairs increased to  $3.43 \pm 0.42$  Å.

Based on the simulations of phosphate binding with wild-type and mutant PBPs, the crystal structures for complexes could be well maintained when the PBPs were bound to 1H phosphate. However, 2H phosphate in the PBPs led to significant disruption of the H-bond network, and the hypothesized binding mode, which involves flipping of the hydroxyl group of Ser38 from phosphate to Asp56 (Figure 3.7B), is unlikely. This is consistent with the sub-angstrom X-ray structure analysis.[26, 70, 307] In the sub-angstrom X-ray crystallography studies of PfluDING, only one proton on oxygen atoms of the phosphate was observed at pH 4.5 and 8.5. PfluDING and E. coli PBP are topologically similar with a structural RMSD value of 1.34 Å over 228 C $\alpha$  carbons and 0.23 Å over all heavy atoms in the binding pocket. Indeed, the residues in the binding pocket are conserved except that the Phe backbone is replaced by that with Leu, and the H-bond distances in the binding pocket

are almost the same.[26] Thus, the phosphate-PBP binding should have the same mechanism i.e. 1H phosphate binds at both pHs. The result highlights the high specificity arising from the extensive network of H-bonds which can distinguish not only between 1H phosphate and sulfate, but also between 1H and 2H phosphates.

	1H	2H
WT	-11.2 (0.4)	-6.4 (0.3)
D56N	-11.1 (0.4)	-5.1 (0.3)
Tris	-3.1 (0.3)	-1.1 (0.3)
Na <sup>+</sup>	-0.8 (0.3)	0.4 (0.3)

Figure 3.8: Calculated standard binding free energy (kcal/mol) of 1H/2H phosphate with PBPs or buffer ligands. The uncertainties are shown in parentheses.

**3.1.4.1.2 Calculated Standard Binding Free Energies** To determine the thermodynamic preference for 1H *vs.* 2H phosphate, I calculated the free energy changes for transferring 1H/2H phosphate from water to PBPs using MD simulations (Figure 3.8). The 1H phosphate binding free energies are 5-6 kcal/mol lower than those of 2H phosphate, suggesting that 1H is the dominant form bound to PBPs. This again agrees with sub-angstrom crystallography studies.[26, 70, 307] It should be noted that constant-pH MD is an alternative way to determine the protonation state of titratable groups.[122, 201] In addition, the binding free energies of 1H phosphate are similar for the wild type and the mutant protein, consistent with experimental measurements. This

similarity is a result of the compensation between the stabilization effect of the LBHB and the changes in total charge of the binding pocket. In the simulations, the stabilization of the LBHB was modeled by the special vdW interactions, which contributed -2.9 kcal/mol to the binding free energy between 1H phosphate and the wild-type PBP. In the mutant, the LBHB is replaced by a normal H-bond, while the replacement of Asp with Asn increases the electrostatic potential of the binding site and reinforce the interaction with the negatively charged phosphate ion. This explains why the D56N mutant with altered electrostatic potential can have a similar binding affinity to that of the wild type.

### 3.1.4.2 The Critical Effect of Buffer Solution

**3.1.4.2.1 Apparent Dissociate Constant  $K_D^{app}$**  It remains unclear why the experimentally measured binding affinities are similar under acidic and basic conditions. The measured quantity is the apparent dissociation constant  $K_D^{app}$ , which concerns the ratio between the total concentrations of phosphate in free and bound forms,  $K_D^{app} = \frac{[PBP] \times ([HPO_4^{2-}] + [H_2PO_4^-])}{([HPO_4^{2-}:PBP] + [H_2PO_4^-:PBP])}$ . Because the calculated binding free energy of 2H phosphate is  $\sim 5$  kcal/mol smaller than that of 1H phosphate,  $[H_2PO_4^- : PBP]$  much smaller than  $[HPO_4^{2-} : PBP]$  at pH 4.5. As a result,  $K_D^{app}$  depends on the ratio between 1H and 2H forms in solution,  $K_D^{app} \approx K_{D,HPO_4^{2-}:PBP} / \frac{[HPO_4^{2-}]}{([HPO_4^{2-}] + [H_2PO_4^-])}$ . The fraction of  $[HPO_4^{2-}]$  is close to 1 at pH 8.5, and reduces to  $\sim 1/500$  at pH 4.5. Thus, the apparent  $K_D^{app}$  will increase from 6 nM at pH 8.5 to 2.5  $\mu M$  at pH 4.5, which corre-

sponds to a change of +3.6 kcal/mol in binding free energy at 298 K. However, this is different from the experimental trend in binding free energy, which only increase by 1.5-1.7 kcal/mol when pH changes from 8.5 to 4.5. It is interesting to note that the calculated binding affinity at pH 4.5 is in excellent agreement with experimental measurement.

pH	expt.		calc. without buffer effect		calc. with buffer effect	
	WT	D56N	WT	D56N	WT	D56N
8.5	-8.85	-8.76	-11.2 (0.5)	-11.1 (0.5)	-9.8 (0.5)	-9.7 (0.5)
4.5	-7.37	-7.05	-7.6 (0.4)	-7.5 (0.4)	-7.6 (0.4)	-7.5 (0.4)

Figure 3.9: Calculated apparent binding free energy (kcal/mol) of phosphate with PBP in wild type and D56N mutant protein at pH 8.5 and 4.5 and 50 mM sodium acetate/Tris acetate. The uncertainties are shown in parentheses.

**3.1.4.2.2 Calculated apparent binding free energies** This discrepancy can be explained by the buffer solution, which has a significant effect on the binding free energy, especially for highly charged species.[62] In previous experimental work, Tris acetate and sodium acetate buffer solutions were used to maintain the pH at 8.5 and 4.5, respectively.[91] I calculated the binding free energy between 1H/2H phosphate and Tris-H<sup>+</sup> or Na<sup>+</sup> (Figure 3.8). The calculated standard binding free energy between 1H phosphate and Tris-H<sup>+</sup> is -3.1 kcal/mol and the corresponding  $K_D$  is 5.2 mM. At experimental Tris acetate concentrations of 50 mM,  $[HPO_4^{2-}]$  is  $\sim 10$  times lower than

$[HPO_4^{2-} : Tris - H^+]$ . Therefore, at pH 8.5, this buffer agent will compete with PBP to bind the 1H phosphate and reduce the apparent binding affinity. This matches the observation that the binding affinity decreases with increasing ionic strength.[294] In acidic buffer solution, neither 1H nor 2H phosphate is bound to  $Na^+$ , so the buffer solution does not affect  $K_D^{app}$ . Using the binding free energies of PBPs and buffer agents with phosphate and the concentration of the buffer solution, I computed the apparent binding free energies under the experimental conditions (Figure 3.9 and Figure 3.2). For both wild-type and mutant PBPs, the binding free energy at pH 4.5 and 8.5 agrees with experimental data within  $\sim 1$  kcal/mol.

Clearly, the interpretation that 1H and 2H phosphate should bind to PBP with similar affinities is due to neglecting the effect of buffer solution. The effect of buffer solution on the binding affinity measurement has been well recognized,[228] but not routinely incorporated in the experimental analyses. The current study demonstrates that molecular simulations can complement experimental measurement in delineating different contributions in the binding process.

In biology, pH is maintained at 7.4 by carbonic acid and bicarbonate, which are neutral or negatively charged and should not bind strongly to phosphates. Therefore, the biological buffer should not affect the specificity for 1H phosphate, i.e. 1H phosphate is the predominant bound state. This again points out the importance of considering the binding of buffer agents if the buffer used in biochemical experiment is different from the biological buffer.

### 3.1.5 Conclusions

In summary, this work resolved a long-standing controversy about the phosphate binding mode (1H vs. 2H) in PBPs through molecular dynamics simulations using the AMOEBA polarizable force field. It shown that 1H phosphate is the energetically favorable species in the binding pocket at different pHs, and simulated H-bond network agrees well with the crystal structure. Based on our simulations, 2H phosphate binds much more weakly and disrupts the H-bond network observed in the crystal structure. After considering the interaction of phosphate with the agents in the buffer solution, for both WT and mutant PBP, the calculated binding free energies are in good agreement with experimental data, i.e. the binding affinity does not vary much with pH although only 1H phosphate is bound to PBPs. Our results highlight the importance of the buffer solution when interpreting the binding affinity data for highly charged species. In addition, molecular simulations can bridge the gap between different experimental techniques and provide new insight for protein-ligand binding.

## 3.2 Calculating Binding Free Energy of Host-Guest system

### 3.2.1 Introduction

Molecular recognition is fundamental to biological processes and is utilized in applications ranging from therapeutics to chemical sensors[280]. Understanding the importance of molecular recognition, the interactions involved



are exceedingly complex and dependent upon a high degree of order between the solutes and the solvent for binding. Computer prediction of binding affinity holds potential to accurately capture thermodynamic information from different states as well as allow for the design of novel ligands.

Methods for binding free energy calculation can be classified according to depiction of either alchemical or physical pathways. The alchemical pathway uses alchemical, or non-physical intermediates to compute binding free energy, which is popular for its general applicability and efficiency. Physical pathways are preferable for large molecules and can give binding mechanism and kinetics[42, 102]. While traditional methods such as Bennetts acceptance ratio[108] have been successful, improvement in computational efficiency is desired for application to large systems and more sophisticated potential energy representations. To this end, many enhanced sampling methods have been developed[55, 211].

Host-guest systems are often used as a model for binding affinity prediction because of their modest size and high specificity among guests. By computing the free energy behaviour of relatively small molecules, inadequacies can be better determined and remediated for the rigorous and strenuous computation of binding free energies for large proteins. In the SAMPL3[213] and SAMPL4[193] host-guest binding competitions, the cucurbit[7]uril macrocycle was used as the host molecule. The cucurbit[n]uril macrocycle (CB[n]) is composed of n conjoined glycoluril subunits forming a cylindrical molecule approximately 9.1Å in height (for a thorough review of CB[n] chemistry, see

[298]). As with many macrocycles, such as cyclodextrin, CB[n] has been explored as a molecular container for drug delivery[133, 165, 337]. The glycoluril subunits position a ring of carbonyl groups at the two faces of the cylinder, while the inner region of carbon-nitrogen chains remains hydrophobic. Hence, guests of hydrophobic cores with cationic end groups can bind with high affinity to the CB[n] host.

This work reports the investigation of host-guest binding thermodynamics between a CB[7] host and a set of 14 small molecules. The guests range from linear hydrocarbons to cycloalkanes, species of norbornanes and adamantane. I use two free energy calculation methods and several thermodynamic inquiries to interpret experimental affinities. In particular, I dissect the roles of entropy and enthalpy in binding for each guest. For anomalous enthalpy/entropy values, the separate entropy contributions of water and the host-guest systems are investigated. The Binding affinities of the host-guest systems are both enthalpy- and entropy-driven. I further discuss the application and convergence of the OSRW and BAR binding free energy methods. Our results attest to the application of binding free energy simulation methods towards the understanding of experimental binding affinities.

## **3.2.2 Computational Methods**

### **3.2.2.1 Orthogonal space random walk (OSRW)**

OSRW is an enhanced sampling scheme for free energy calculation, which allows more effective sampling of conformational transitions in aque-

ous solution.[197, 226, 336, 338] It performs a random walk in two orthogonal dimensions. One dimension is along the order parameter representing an alchemical intermediate state that connects the two states of interest.[3, 155] The other dimension is along the orthogonal generalized force ( $F_\lambda = \partial U/\partial\lambda$ ), whose integral is the free energy (Figure 3.5). Once a state is sampled, a Gaussian distributed bias is added to discourage the system from revisiting that state. A complete explanation of the method as well as the requisite adjustments needed to employ a polarizable force field can be found in Abella et al.[188]

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3.5)$$

### 3.2.2.2 Recent Development of Enhanced Sampling

Several efforts have been made in recent years to accelerate the simulations of polarizable force fields. Multiple time step algorithms have been developed to allow for very large time steps in molecular dynamics simulations.[169, 206, 217] In the extreme case, the computation speed can be accelerated by 10 to 20 times.[206] Dual force field approach introduced by Schnieders and coworkers,[64] takes advantage of the sampling efficiency of the fixed-point charge model (OPLS-AA) and accuracy of polarizable force fields to compute the absolute crystal decomposition thermodynamics. A similar procedure was used by Shirts and coworkers to indirectly calculate the free energy of three benzene polymorphs by AMOEBA.[41] There have also been significant ad-

vances in thermodynamic and kinetic reweighting methods,[181, 314] which can in principle be combined with the dual-force field methods. An exciting new direction is to combine polarizable force fields with enhanced sampling methods such as orthogonal space sampling,[69] Markov state models and Milestoning,[112, 125] which will significantly extend the time and length scales of polarization force fields simulations to areas such as protein and nucleic acids conformational dynamics. These studies would provide crucial feedback to the force field development and insights into our understanding of the intermolecular forces and how they affect the structure and properties of biomolecular systems.

### 3.2.2.3 Simulation System

In this study, the absolute binding free energy values of 14 guests in the SAMPL4 CB[7]-guest system were calculated using the polarizable AMOEBA force field. Parameters for the molecules were derived by following the procedure previously described in Ren et al[248]. All molecular dynamics simulations were run using TINKER with a RESPA integrator[34], Bussi thermostat[21], and a 2.0 femtosecond time step. The vdW calculations had a 12.0 Å cutoff while the electrostatic calculations had a 7.0 Å cutoff. The Gaussian bias was deposited every 10 steps, with a height of 0.005 kcal/mol and widths of 4 kcal/mol for  $F_\lambda$  and 0.01 for  $\lambda$ . Additional simulations with a reduced height of 0.001 or 0.002 kcal/mol were also carried out for some guests. The production time of the OSRW is around 15-20 ns. All OSRW simulations were

conducted on Texas Advanced Computing Center (TACC) Stampede as well as a local computer cluster. For the BAR simulations, first the electrostatics were gradually scaled off with vdW interactions kept at full strength, and then the vdW interactions were scaled off. The numbers of steps for these two stages were 11-12 and 10-13 respectively. The total simulation time for each step was 1 ns and coordinates were saved every 1 ps for analysis. The correction  $G_{correction}$  was 6.245 kcal/mol and should be added to all binding free energy calculations for both BAR and OSRW. The binding enthalpy was obtained from the difference between the average energies in the binding and free states. This method has comparable accuracy with that of the vant Hoff method[320] and that of the BAR method.[25]

### 3.2.3 Results: Calculated Binding Free Energies

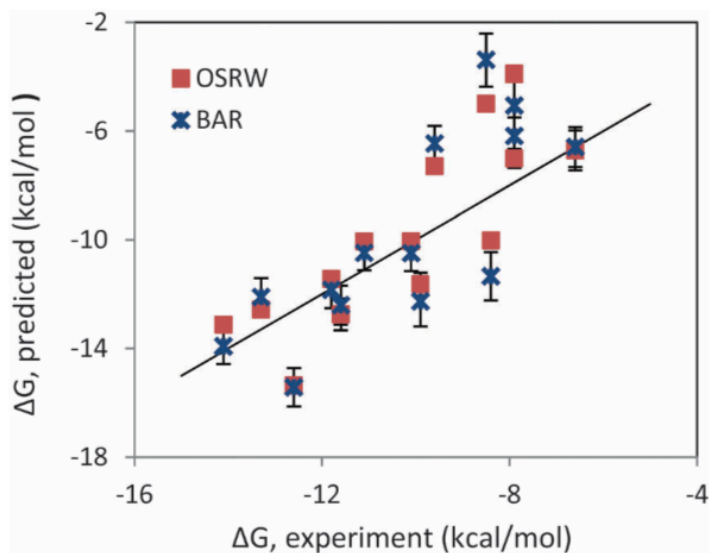


Figure 3.10: Predicted binding free energy as a function of experimental binding free energy (in kcal/mol). Line is  $y=x$ .

Figure 3.10 and Figure 3.11 both present binding free energy results from OSRW and BAR computations compared with experiment. Figure 3.10 shows that the OSRW and BAR free energies establish good correlation with experiment, having  $R^2$  correlation values of 0.69 (OSRW) and 0.62 (BAR). Besides, Figure 3.11, structures and energies of the guest ligands studied here are presented[193]. The host for all ligands is CB[7] as stated previously. For each ligand in Figure 3.11, the free energy values of experiment, OSRW, and BAR are presented explicitly. Reported in the SAMPL4 results, the absolute uncertainty of all experimental free energy values is  $\pm 0.1$  kcal/mol. The

BAR results are those that were previously reported in the SAMPL4 contest.[193] Lastly, Figure 3.12 presents errors and correlation metrics between OSRW/BAR and experimental values. Despite the duplicate runs and Gaussian height decrease necessary for the OSRW computations, the Kendall  $\tau$  coefficient for OSRW supports a strong agreement between OSRW and experiment.



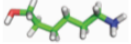
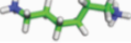










Guest	Guest structure	$\Delta G_{bind}$			
		Experimental	BAR	OSRW	OSRW (10 ns)
C1		-9.9	-12.27 ± 0.92	-11.64 ± 0.42	-10.43 ± 1.01
C2		-9.6	-6.46 ± 0.65	-7.29 ± 0.20	-5.50 ± 0.05
C3		-6.6	-6.59 ± 0.74	-6.71 ± 0.74	-5.17 ± 1.03
C4		-8.4	-11.34 ± 0.89	-10.02 ± 0.25	-8.31 ± 0.75
C5b		-8.5	-3.39 ± 0.97	-5.00 ± 0.11	-4.46 ± 0.26
C6		-7.9	-6.18 ± 0.69	-7.01 ± 0.35	-6.64 ± 0.07
C7		-10.1	-10.49 ± 0.66	-10.05 ± 0.09	-10.96 ± 0.78
C8		-11.8	-11.84 ± 0.68	-11.44 ± 0.09	-11.15 ± 0.76
C9		-12.6	-15.42 ± 0.71	-15.35 ± 0.29	-15.44 ± 0.65
C10		-7.9	-5.06 ± 0.91	-3.90 ± 0.26	-3.69 ± 0.68
C11		-11.1	-10.48 ± 0.64	-10.06 ± 0.25	-9.82 ± 0.34
C12		-13.3	-12.11 ± 0.70	-12.57 ± 0.03	-12.33 ± 0.16
C13		-14.1	-13.92 ± 0.65	-13.13 ± 0.13	-12.63 ± 0.52
C14		-11.6	-12.41 ± 0.72	-12.75 ± 0.59	-12.05 ± 0.58

Figure 3.11: Host-guest binding free energies. The OSRW column presents the average of results from the full length of simulations, while the OSRW (10 ns) column presents values cut off at 10ns. The host molecule for all structures is cucurbit[7]uril. All free energies are given in kcal/mol. The experimental free energies hold an uncertainty of  $\pm 0.1$  kcal/mol.

For ligand C5, positive binding free energies calculated from BAR led to the exploration of multiple ligand protonation states, denoted as C5 and C5b. In five ligand cases (C1, C3, C5b, C9, and C10), the OSRW computation displayed large fluctuations in free energy. Since the fluctuation is proportional



to the bias deposition rate, additional OSRW simulations were conducted with decreased Gaussian-height biases for each of these ligands. In theory, lowering the height of the Gaussian distribution will suppress fluctuations at the expense of slowing down dynamics. However, in this work, the OSRW computations with a lowered Gaussian height (LGH) bias converged at roughly the same simulation time as the original computations. Lastly, ligands C3 and C10 were duplicated in the OSRW computation due to poor convergence of the original simulations.

For Figure 3.10 and Figure 3.11, the final OSRW ligand binding free energy value is taken to be the average over all of the OSRW computations for that ligand, with some values excluded (explained below). Multiple independent OSRW simulations were run for each ligand. As mentioned above, for two ligands, OSRW computations were repeated with the original parameters. The averaging of the free energies includes the LGH and repeated computations with the original pair of simulations. Exceptions to this average method are ligands C5 and ligands C10. The binding free energy value for ligand C5 was taken to be the average of the ligand C5b\_LGH computation. The protonation state for ligand C5 reported by the BAR computations was similarly C5b. The 2.5 kcal/mol disagreement between original OSRW simulations for ligand C5b, as well as the nice agreement between the C5b\_LGH simulations (within 0.3 kcal/mol), supported our use of the C5b\_LGH data. For ligand C10, all of the OSRW free energy values were used in the average except the -0.76 value as it was in disagreement with all of the other five values by 2.5 kcal/mol.

I suspect that this low free energy value is an artefact of a slow-convergence binding energy computation.

Analysis of computed binding affinities from the SAMPL challenge allows for elucidation of binding thermodynamics as well as examination of computational predictions. In the official SAMPL4 host-guest paper, free energy values from BAR simulations using the AMOEBA polarizable force field were noted for good performance[193]. Our OSRW-computed free energies correlate with experimental values slightly better than the BAR results. Note that both methods use the exact same parameter sets and simulations parameters. However, long computational times needed for convergence of OSRW free energy were observed. Upwards of 20 ns of computation time in binding was required for some ligands, while in our previous work, the hydration free energy was able to converge in less than 4 ns[188]. For comparison, the BAR computations were performed for 1ns for each vdW and electrostatic window. One possible reason that the OSRW method applied here may be slow to converge is due to the underlying metadynamics procedure. Recently, the Orthogonal Space Tempering[197] method has been proposed to address this problem.

Method	RMS error	AUE	$R^2$	Kendall $\tau$
OSRW	1.92	1.51	0.69	0.74
BAR	2.26	1.73	0.62	0.58
OSRW (10 ns)	2.22	1.73	0.73	0.75

Figure 3.12: Model deviation from experiment. RMS energy difference, and AUE (Average Unsigned Error) are in kcal/mol.

I also investigated the OSRW results if the free energy computations were carried out for only 10 ns rather than continued to 15+ ns. In Figure 3.11 the 10ns OSRW binding free energies are presented in comparison to the experimental values. Surprisingly, the  $R^2$  correlation value and the Kendall correlation coefficient are high, supporting strong correlation between OSRW and experiment after just 10ns simulations (Figure 3.12). Despite this strong correlation, the individual ligand errors and the RMSE between experiment and OSRW are slightly higher than the final results.

### 3.2.4 Discussion

#### 3.2.4.1 Enthalpy-Entropy Decomposition

Guest	$\Delta H$	STD( $\Delta H$ )	$-T\Delta S$
C1	-14.91	13.87	2.64
C2	-17.39	13.54	10.93
C3	-18.58	14.79	12.00
C4	-6.62	14.02	-4.72
C5	3.03	15.24	-6.41
C5b	-5.08	13.96	1.53
C6	-12.48	14.60	6.30
C7	-26.99	13.47	16.56
C8	-28.56	14.30	16.72
C9	-3.69	13.41	-11.73
C10	45.20	13.71	-49.57
C11	1.33	13.37	-11.94
C12	-5.58	12.68	-6.67
C13	-7.53	14.19	-6.18
C14	4.28	12.63	-16.90

Figure 3.13: Host-guest binding enthalpies and entropies (kcal/mol). STD( $\Delta H$ ) is the uncertainty of enthalpy.

To gain insights into the molecular driving forces for binding, the enthalpy and entropy contributions of the binding free energy have been examined. Figure 3.13 lists the calculated binding enthalpy and entropy for each guest ligand. Although the binding free energies for different ligands are close, ranging from -15 to -5 kcal/mol, the binding enthalpies are vastly different. This is a good demonstration of the enthalpy-entropy compensation in host-guest binding. Due to the relatively short simulation time (1 ns), the uncertainties are on the order of 10 kcal/mol. Nevertheless, it can be seen that some of the recognitions are driven by enthalpy while others by entropy. Ligands 9, 12, and 13 have both favorable binding enthalpy and entropy. Extreme examples are ligand C10 for entropy-driven binding, and ligands C7 and C8 for enthalpy-driven binding. However, there appears to be no simple relationship between the binding thermodynamics of the ligand and its charge or geometry. Comparing C5 with C5b and C3 with C4, I find that the binding enthalpy does not correlate with the net charge. Ligands C7, C8 and C9 have the same functional groups and their binding affinities increase with ring size, but their entropy values differ. Enthalpy values of ligands C7 and C8 clearly indicate a dominant contribution of enthalpy, while for ligand C9 the enthalpy value is competitive with entropy.

### 3.2.4.2 Hydrogen Bonding Analysis

Guest	$N_{\text{solution}}$	$N_{\text{complex}}$	$N_{\text{complex}}^{\text{g-h}}$	$N_{\text{complex}}^{\text{g-w}}$
C7	2.690	3.554	1.617	1.937
C8	2.792	3.389	1.201	2.188
C10	5.709	5.452	1.325	4.127

Figure 3.14: Analysis of hydrogen bond numbers for guests C7, C8 and C10. The number of hydrogen bonds between guest-water in solution and between guest-host/water in host-guest complex are listed as  $N_{\text{solution}}$  and  $N_{\text{complex}}$  respectively. Further decompositions of hydrogen bond numbers between guest-host, and between guest-water in host-guest complex are given in  $N_{\text{complex}}^{\text{g-h}}$  and  $N_{\text{complex}}^{\text{w-h}}$ . The presenting hydrogen bond numbers are averaged by 1000 frames over 1 ns.

Further analyses were carried out to look into the binding mechanisms. To explain why guest ligands C7 and C8 are enthalpy-driven, I investigated the ligand hydrogen bonding formation in water and complexes. Figure 3.14 lists hydrogen bond (H-bond) numbers for ligands C7, C8 and C10 between guest-water in solution and between guest-host/water in the complex. The data are averaged over 1000 frames in 1 ns. Compared to ligands C7 and C8, ligand C10 formed more H-bonds when free in water and bound to the complex. Furthermore, I analysed the portion of H-bonds formed between guest-host and guest-water. The three ligands formed similar numbers of H-bonds with the host while ligand C10 has twice the H-bonds formed with the surrounding water than other ligands. This may be attributed to the structural differences: ligand C10 has 3 polar amine groups with two of them

exposed to the surroundings, attracting water and other polar groups. In contrast, ligands C7 and C8 have only one amine group each, leading to less intermolecular interaction. Noticeably, an increase of H-bonds in ligands C7 and C8 is found when moved from solution to the host-guest complex. On the other hand, the number of H-bonds formed by C10 decreases upon binding. The changes in H-bonds may explain why the binding of ligands C7 and C8 were found to be enthalpy-driven.

### 3.2.4.3 Configurational Entropy

The rotation of guest ligands C7, C8, and C10 inside the CB[7] host was measured to explore the entropic aspects of these ligands. Three atoms from each ligands aromatic ring were chosen to represent one plane, while three atoms from the host were chosen to produce a plane that bisects the host equally. The rotation of the guest plane with respect to the host plane was measured over the coordinates of 5ns trajectories. The potential of mean force (PMF) was also computed for the rotation angles. Similar to a study of an octa-acid host-guest system[196], the guest ligands here were determined to rotate almost freely with only small free energy barriers ( $\sim 0.5$  kcal/mol). Likewise, computation of the entropy using  $S = -k_B \times \sum p \ln(p)$  resulted in minute contributions ( $S_{lig(complex)rot} - S_{lig(free)rot}^{rot} \approx 0.05 kcal/mol$  at T=300 K).

The configurational entropies of host-guest complexes C7, C8, and C10 were computed using quasiharmonic analysis[8, 39]. In the quasiharmonic analysis method, the mass weighted covariance matrix of atomic fluctuations is

computed. Eigenvalues  $\lambda_i$  of this covariance matrix are then expounded to frequencies of collective motions,  $\omega_i = (RT/\lambda_i)^{1/2}$ . The estimated entropy  $S$  of the molecule is determined by Equation 3.6 where  $R$  is the gas constant,  $\hbar$  is Planck's constant, and  $T$  is temperature.

$$S = R \sum_{i=1}^{3N-6} \frac{\hbar\omega_i/RT}{e^{\hbar\omega_i/RT} - 1} - \ln(1 - e^{\hbar\omega_i/RT}) \quad (3.6)$$

The quasiharmonic entropy was computed using AMBER14[50]. For each molecule, all heavy atoms (C,N,O) were included in the covariance matrix. Figure 3.15 shows the quasiharmonic entropy values for the host-guest systems C7, C8 and C10. These values include entropies of the host-guest complex  $S_{hg}$ , the guest only in complex  $S_{g(complex)}$ , the host only in complex  $S_{h(complex)}$ , the guest in solution  $S_{g(solution)}$ , the host in solution  $S_{h(solution)}$ , and the entropic contribution of binding  $T\Delta S_{conf}$  where  $\Delta S_{conf} = S_{hg} - S_{h(solution)} - S_{g(solution)}$ . The quasiharmonic approximation maintains limitations involving the use of Cartesian coordinates and the presence of multiple steep energy wells[15]. Further, the quasiharmonic approximation is known to present an upper-bound to entropy primarily due to correlations between modes[13, 310]. However, several trends may be observed from the computed values. CB[7]-C10 complex has the highest entropic cost ( $T\Delta S_{conf}$ ) out of the three complexes computed.

Guest	$S_{hg}$	$S_{g(\text{complex})}$	$S_{h(\text{complex})}$	$S_{g(\text{solution})}$	$-T\Delta S_{\text{conf}}$
C7	364.38	74.53	302.73	80.99	5.69
C8	379.41	91.88	289.87	96.34	3.12
C10	366.28	87.67	294.41	94.59	6.61

Figure 3.15: Configurational entropy computed from quasiharmonic analysis.  $S_{h(\text{solution})}$  is 495.61 cal/mol/K.

Entropy values are given for  $S_{hg}$  the host-guest complex,  $S_{g(\text{complex})}$  the guest only in complex,  $S_{h(\text{complex})}$  the host only in complex, and  $S_{g(\text{solution})}$  the guest in solution.  $S_{hg}$ ,  $S_{g(\text{complex})}$ ,  $S_{h(\text{complex})}$ , and  $S_{h(\text{solution})}$  are computed from 5ns simulations while  $S_{g(\text{solution})}$  values are computed from 1ns simulations. All entropy values (except where marked) in cal/mol/K.  $\Delta S_{\text{conf}} = S_{hg}S_{h(\text{solution})} - S_{g(\text{solution})}$ .  $T\Delta S_{\text{conf}}$  computed at 300K, with units of kcal/mol.

Given that the enthalpy/entropy decomposition analysis suggested binding of guest C10 to be entropically favorable ( $T\Delta S_{\text{tot}} < 0$ ), the positive configurational entropy change computed here (Figure 3.15) indicates that the favourable binding entropy of ligand C10 is likely water driven. Binding of guests C7 and C8 resulted in approximately the same entropic cost. Although the values of  $S_{hg}$  and  $S_{g(\text{solution})}$  differ for guests C7 and C8, when combined, the values largely offset the differences. From analysis of guests C7 and C8, intramolecular atomic fluctuations of the aromatic carbon atoms inside the host are greater for C8 than for C7. This is consistent with intuition: the larger aromatic molecule of ligand 8 is slightly pressed inward by the host. This effect is evident in the  $S_{h(\text{complex})}$  values, where the host in guest C7 complex has



roughly 4 kcal/mol greater entropy (TS) than the host in guest C8 complex, which is strained due to ligand size. Similar to the C10 complex, guests C7 and C8 complexes have a positive (unfavourable) entropy contribution, and additional unfavourable entropic interactions from water likely increase the binding entropy to the values in Figure 3.13.

As noted above, there are discrepancies between OSRW and BAR results as well as between independent OSRW simulations for some ligands. To explain this, I observe that for an unbiased estimator, the uncertainty of a measured quantity is related to the sample distribution and the autocorrelation time as

$$\sigma(\bar{A}) = \sigma(A) \sqrt{\frac{2\tau}{t}} \quad (3.7)$$

where  $\tau$  is the integrated autocorrelation time and  $t$  is the total sampling time.  $t/2\tau$  is also interpreted as the effective number of independent samples. Equation 3.7 is valid for BAR. As for OSRW, since the underlying metadynamics does not converge asymptotically[207], Equation 3.7 should provide a lower bound for its error. The sample distribution depends on the hybrid Hamiltonian, i.e. the decoupling scheme for the alchemical transition, which is different in the OSRW and BAR simulations. The correlation time varies with the simulation method.

### 3.2.4.4 Convergence of the BAR and OSRW

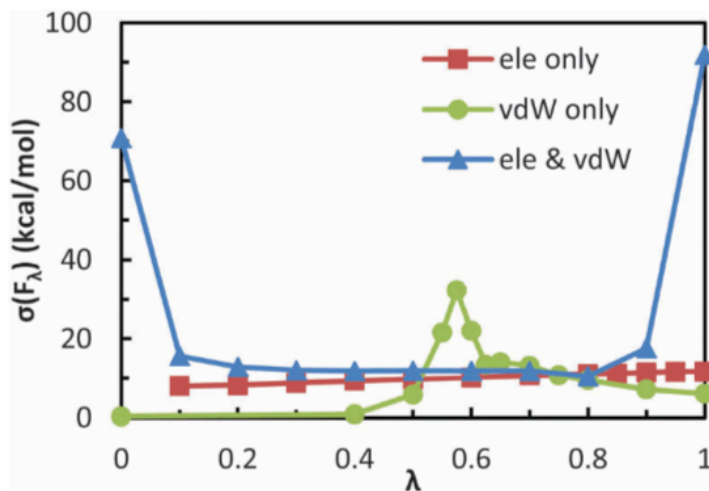


Figure 3.16: Standard deviation of  $F_\lambda$  as a function of  $\lambda$  for different coupling schemes. All analyses are based on the decoupling of guest C10 from its host-guest complex state. vdW only means that the vdW interaction is decoupled when there is no electrostatics. ele only means that the electrostatics is decoupled while vdW interaction is modelled at full strength. ele & vdW means that both electrostatics and vdW interactions are decoupled simultaneously as in the current OSRW implementation.

Generally, the correlation time in metadynamics should be shorter than that of a classical molecular dynamics simulation on the same Hamiltonian. However, it is difficult to compare the correlation time between OSRW and BAR because OSRW has an additional degree of freedom. So here I focus on the effect of the decoupling scheme on the convergence. Figure 3.16 shows the standard deviation of  $F_\lambda$  in different decoupling schemes. When only the vdW interaction is decoupled (scaled down), the distribution  $F_\lambda$  is very narrow

up to  $\lambda = 0.5$ .  $\sigma(F_\lambda)$  increases sharply and then falls to roughly 10 kcal/mol when  $\lambda$  goes from 0.5 to 0.6. When the electrostatics interaction is decoupled in the presence of vdW interaction,  $\sigma(F_\lambda)$  is nearly constant around 10 kcal/mol, which means that there is no dramatic change in phase space and that the evenly spaced  $\lambda$  points perform very well in distributing the simulation time. When vdW and electrostatics interactions are decoupled simultaneously,  $\sigma(F_\lambda)$  is significantly higher than when the two interactions are decoupled separately as  $\lambda$  approaches 0 and 1. In other words, decoupling both interactions together enlarges the available phase space. As a result, more independent samples are needed for  $\langle F_\lambda \rangle$  to converge at these two end states. In addition, I note that the correlation time in the fixed  $\lambda$  OSRW simulations is much longer than in the BAR simulations when  $\lambda = 0$  (but not when  $\lambda = 1$ ). Although there is no direct link to the dynamics of the OSRW simulations reported in Figure 3.11, it manifests that decoupling both interactions will create a rough energy landscape that makes sampling difficult. Therefore, the poor convergence of some of the OSRW simulations can be largely attributed to the decoupling scheme.

Guest	RMSE (kcal mol <sup>-1</sup> )	Charge
C5b	1.35	2
C4	1.34	2
C10	1.28	2
C1	1.05	2
C5	0.98	1
C6	0.90	1
C2	0.85	1
C13	0.80	1
C3	0.80	1
C14	0.68	1
C11	0.49	1
C12	0.46	1
C7	0.45	1
C8	0.41	1
C9	0.30	1

Figure 3.17: Correlation between uncertainties of binding free energies and net charge for each system. RMSE is the root mean square difference between OSRW results and the reference BAR results.

There is a positive correlation between the uncertainties of the OSRW simulations and the net charge of the system. Except for guest 3, all the OSRW results for systems with charge +1 agree well with those of BAR results, whereas large differences can be found for systems with charge +2 (Figure 3.17). This further supports our finding that decoupling vdW and electrostatics interactions together hinders the sampling. I expect that the problem will be less prominent for neutral systems.

### 3.2.5 Conclusions

In this work, binding free energies of the SAMPL4 host-guest system CB[7] with 14 guest molecules were computed with both BAR and OSRW

methods and AMOEBA polarizable force field. Overall the AMOEBA binding free energy values computed using both BAR and OSRW are in good agreement with experimental results. The binding thermodynamics of this series of host-guest systems varies from ligand to ligand. Some are driven by enthalpy changes while others by entropy gains. The guest ligands C7, C8 and C10 have been further examined, which display high enthalpy or entropy changes upon binding. The enthalpy-entropy decomposition suggests that the binding of guest C10 is entropy driven, while binding of guests C7 and C8 have large enthalpic contributions. Hydrogen bonding analysis showed that guest C10 formed several hydrogen bonding interactions with both water and host CB[7], largely due to the three hydrophilic amine groups. Guests C7 and C8 gain additional H-bonds upon binding while C10 loses H-bonds upon binding, consistent with the enthalpy-entropy decomposition results. Configurational entropy was computed for guests C7, C8, C10 and their complexes with the host using quasiharmonic analysis. The configurational binding entropy was determined to be relatively small for all guests, hinting at the substantial role of water molecules. Through analysis of intramolecular atomic fluctuations of guests C7 and C8, cyclic carbon atoms inside the host were found to fluctuate more for guest C8 than C7, intuitively a result of the larger ring of C8. Unlike ligand-protein binding, the guest molecules were observed to freely rotate inside the host ring. Convergence of the BAR and OSRW free energy calculation methods were compared. The current OSRW implementation encounters convergence problems at the low end of vdW and electrostatics

decoupling. Possible improvements can be achieved by separating the vdW and electrostatic decoupling, well-tempered metadynamics[207] and employing metadynamic alternatives[197]. Nonetheless, here, both BAR and OSRW methods are found to be adequate to determine the binding affinities for the model host-guest systems.

## Chapter 4

### Conclusion

Molecular dynamics simulation is indispensable tools for investigating physical properties of proteins, nucleic acids and designing new molecules and materials.[86, 257, 299] Due to recent advances in computing hardware and improved simulation methods, the time and length scales of molecular dynamics simulations have been greatly extended.[37, 67, 78, 112, 125] These advances not only lead to more reliable interpretation and predictions by computer simulations but also crucial for examining and improving the underlying physical models and simulation methods.

Polarizable force fields have grown steadily during the past few years in terms of computational efficiency, model accuracy and applications to biomolecular systems. Advances in GPU computing, polarization and simulation algorithms have provided access to the microsecond time scale with polarizable force fields, and the computational overhead compared to fixed-charge force field has been significantly reduced. The applications of polarizable force fields have provided many new insights. Recent studies using polarizable force fields have demonstrated the critical role of polarization for the stability of nucleic acids and proteins, base-pair flipping, ion distribution around DNA, diffusion

and permeation of small molecules. In general, simulations with polarization force fields agree better with experiments.

In this thesis, we have discussed the development of advanced polarizable force fields for water and organic molecules and the application of AMOEBA on phosphate-protein binding and hostguest systems. These studies would provide crucial feedback to the force field development and insights into the understanding of the intermolecular forces and how they affect the structure and properties of biomolecular systems.



## Bibliography

- [1] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *Journal of Chemical Physics*, 123(23), 2005.
- [2] Badi Abdul-Wahid, Li Yu, Dinesh Rajan, Haoyun Feng, Eric Darve, Douglas Thain, and Jesus A. Izaguirre. In *IEEE International Conference on e-Science*.
- [3] Jayvee R. Abella, Sara Y. Cheng, Qiantao Wang, Wei Yang, and Pengyu Ren. Hydration free energy from orthogonal space random walk and polarizable force field. *Journal of Chemical Theory and Computation*, 10(7):2792–2801, 2014.
- [4] N. L. Allinger. Conformational-analysis .130. mm2 - hydrocarbon force-field utilizing v1 and v2 torsional terms. *Journal of the American Chemical Society*, 99(25):8127–8134, 1977.
- [5] N. L. Allinger, K. S. Chen, and J. H. Lii. An improved force field (mm4) for saturated hydrocarbons. *Journal of Computational Chemistry*, 17(5-6):642–668, 1996.
- [6] N. L. Allinger, Y. H. Yuh, and J. H. Lii. Molecular mechanics - the mm3 force-field for hydrocarbons .1. *Journal of the American Chemical*

- Society*, 111(23):8551–8566, 1989.
- [7] V. S. Allured, C. M. Kelly, and C. R. Landis. Shapes empirical force-field - new treatment of angular potentials and its application to square-planar transition-metal complexes. *Journal of the American Chemical Society*, 113(1):1–12, 1991.
- [8] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics*, 115(14):6289–6292, 2001.
- [9] A. D. Andricopulo, L. B. Salum, and D. J. Abraham. Structure-based drug design strategies in medicinal chemistry. *Current Topics in Medicinal Chemistry*, 9(9):771–790, 2009.
- [10] Victor M. Anisimov and Claudio N. Cavasotto. Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Icksh2 domain. *Journal of Computational Chemistry*, 32(10):2254–2263, 2011.
- [11] J. Applequist, J. R. Carl, and K. K. Fung. Atom dipole interaction model for molecular polarizability - application to polyatomic-molecules and determination of atom polarizabilities. *Journal of the American Chemical Society*, 94(9):2952–+, 1972.
- [12] Christopher M. Baker, Victor M. Anisimov, and Alexander D. MacKerell. Development of charmm polarizable force field for nucleic acid

- bases based on the classical drude oscillator model. *The Journal of Physical Chemistry B*, 115(3):580–596, 2011.
- [13] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.
- [14] S. Barlow, A. L. Rohl, S. G. Shi, C. M. Freeman, and D. OHare. Molecular mechanics study of oligomeric models for poly(ferrocenylsilanes) using the extensible systematic forcefield (esff). *Journal of the American Chemical Society*, 118(32):7578–7592, 1996.
- [15] Riccardo Baron, Philippe H. Hnenberger, and J. Andrew McCammon. Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: Correction terms and convergence properties. *Journal of Chemical Theory and Computation*, 5(12):3150–3160, 2009.
- [16] D. M. Bates and G. S. Tschumper. Ccsd(t) complete basis set limit relative energies for low-lying water hexamer structures. *Journal of Physical Chemistry A*, 113(15):3555–3559, 2009.
- [17] D. R. Bell, R. Qi, Z. F. Jing, J. Y. Xiang, C. Mejias, M. J. Schnieders, J. W. Ponderc, and P. Y. Ren. Calculating binding free energies of host-guest systems using the amoeba polarizable force field. *Physical Chemistry Chemical Physics*, 18(44):30261–30269, 2016.

- [18] F. Bemani and R. Sadighi-Bonabi. Plasma core at the center of a sonoluminescing bubble. *Phys Rev E Stat Nonlin Soft Matter Phys*, 87(1):013004, 2013.
- [19] C. H. Bennett. Efficient estimation of free-energy differences from monte-carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.
- [20] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.
- [21] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [22] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus. Absolute binding free energies: A quantitative approach for their calculation. *Journal of Physical Chemistry B*, 107(35):9535–9551, 2003.
- [23] D. Bratko, L. Blum, and A. Luzar. A simple-model for the intermolecular potential of water. *Journal of Chemical Physics*, 83(12):6367–6370, 1985.
- [24] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm - a program for macromolecular

- energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.
- [25] B. R. Brooks, D. Janezic, and M. Karplus. Harmonic-analysis of large systems .1. methodology. *Journal of Computational Chemistry*, 16(12):1522–1542, 1995.
- [26] Ignasi Buch, Toni Giorgino, and Gianni De Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 108(25):10184–10189, 2011.
- [27] R. A. Buckingham. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences*, 168(A933):264–283, 1938.
- [28] P. Bui, D. Rajan, Badi Abdul-Wahid, Jesus A. Izaguirre, and D. Thain. In *Workshop on Python for High Performance and Scientific Computing (PyHPC)*.
- [29] S. Bulusu, S. Yoo, E. Apra, S. Xantheas, and X. C. Zeng. Lowest-energy structures of water clusters (h<sub>2</sub>o)<sub>11</sub> and (h<sub>2</sub>o)<sub>13</sub>. *Journal of Physical Chemistry A*, 110(42):11781–11784, 2006.
- [30] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. i. prominent features of the water dimer potential energy surface. *Journal of Chemical Physics*, 116(4):1479–1492, 2002.

- [31] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iii. reparametrization of an all-atom polarizable rigid model (ttm2-r) from first principles. *Journal of Chemical Physics*, 116(4):1500–1510, 2002.
- [32] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iv. a flexible, all-atom polarizable potential (ttm2-f) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *Journal of Chemical Physics*, 116(12):5115–5124, 2002.
- [33] Sebastian Busch, Christian D. Lorenz, Jonathan Taylor, Luis Carlos Pardo, and Sylvia E. McLain. Short-range interactions of concentrated proline in aqueous solution. *The Journal of Physical Chemistry B*, 118(49):14267–14277, 2014.
- [34] G. Bussi, D. Donadio, and M. Parrinello. Comp 8-canonical sampling through velocity rescaling. *Abstracts of Papers of the American Chemical Society*, 234, 2007.
- [35] Claudia Caltagirone and Philip A. Gale. Anion receptor chemistry: highlights from 2007. *Chemical Society Reviews*, 38(2):520–563, 2009.
- [36] S. L. Carnie and G. N. Patey. Fluids of polarizable hard-spheres with dipoles and tetrahedral quadrupoles - integral-equation results with application to liquid water. *Molecular Physics*, 47(5):1129–1151, 1982.

- [37] Kutzner Carsten, Pll Szilrd, Fechner Martin, Esztermann Ansgar, de Groot Bert L., and Grubmller Helmut. Best bang for your buck: Gpu nodes for gro-macs biomolecular simulations. *Journal of Computational Chemistry*, 36(26):1990–2008, 2015.
- [38] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.
- [39] C. E. Chang, W. Chen, and M. K. Gilson. Evaluating the accuracy of the quasiharmonic approximation. *Journal of Chemical Theory and Computation*, 1(5):1017–1028, 2005.
- [40] M. Chaplin. Opinion - do we underestimate the importance of water in cell biology? *Nature Reviews Molecular Cell Biology*, 7(11):861–866, 2006.
- [41] John D. Chodera, William C. Swope, Frank No, Jan-Hendrik Prinz, Michael R. Shirts, and Vijay S. Pande. Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures. *The Journal of Chemical Physics*, 134(24):244107, 2011.
- [42] Clara D. Christ, Alan E. Mark, and Wilfred F. van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, 31(8):1569–1582, 2010.

- [43] Piotr Cieplak, James Caldwell, and Peter 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*, 22(10):1048–1057, 2001.
- [44] G. A. Cisneros, J. P. Piquemal, and T. A. Darden. Generalization of the gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *Journal of Chemical Physics*, 125(18), 2006.
- [45] G. Andrés Cisneros, Mikko Karttunen, Pengyu Ren, and Celeste Sagui. Classical electrostatics for biomolecular simulations. *Chemical Reviews*, 114(1):779–814, 2014.
- [46] M. Clark, R. D. Cramer, and N. Vanopdenbosch. Validation of the general-purpose tripos 5.2 force-field. *Journal of Computational Chemistry*, 10(8):982–1012, 1989.
- [47] E. Clementi, D. L. Raimondi, and Reinhard.Wp. Atomic screening constants from scf functions .2. atoms with 37 to 86 electrons. *Journal of Chemical Physics*, 47(4):1300–, 1967.
- [48] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman. Dipole-moment of water from stark measurements of h<sub>2</sub>o, hdo, and d<sub>2</sub>o. *Journal of Chemical Physics*, 59(5):2254–2259, 1973.



- [49] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.
- [50] R.M. Betz D.S. Cerutti T.E. Cheatham III T.A. Darden R.E. Duke T.J. Giese H. Gohlke A.W. Goetz N. Homeyer S. Izadi P. Janowski J. Kaus A. Kovalenko T.S. Lee S. LeGrand P. Li T. Luchko R. Luo B. Madej K.M. Merz G. Monard P. Needham H. Nguyen H.T. Nguyen I. Omelyan A. Onufriev D.R. Roe A. Roitberg R. Salomon-Ferrer C.L. Simmerling W. Smith J. Swails R.C. Walker J. Wang R.M. Wolf X. Wu D.M. York D.A. Case, J.T. Berryman and P.A. Kollman. Amber 2015. Report, University of California, San Francisco, 2015.
- [51] M. K. Dahlgren, P. Schyman, J. Tirado-Rives, and W. L. Jorgensen. Characterization of biaryl torsional energetics and its treatment in opls all-atom force fields. *Journal of Chemical Information and Modeling*, 53(5):1191–1199, 2013.
- [52] W. Damm, A. Frontera, J. TiradoRives, and W. L. Jorgensen. Opls all-atom force field for carbohydrates. *Journal of Computational Chemistry*, 18(16):1955–1970, 1997.
- [53] L. X. Dang 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*, 106(19):8149–8159, 1997.
- [54] L. X. Dang and B. M. Pettitt. Simple intramolecular model potentials for water. *Journal of Physical Chemistry*, 91(12):3349–3354, 1987.
- [55] M. Zuckerman Daniel. Equilibrium sampling in biomolecular simulations. *Annual Review of Biophysics*, 40(1):41–62, 2011.
- [56] Joseph E. Davis and Sandeep Patel. Charge equilibration force fields for lipid environments: Applications to fully hydrated dppc bilayers and dmpc-embedded gramicidin a. *The Journal of Physical Chemistry B*, 113(27):9183–9196, 2009.
- [57] J. E. Dennis, D. M. Gay, and R. E. Welsch. An adaptive non-linear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7:348–368, 1981.
- [58] M. Di Pierro and R. Elber. Automated optimization of potential parameters. *Journal of Chemical Theory and Computation*, 9:3311–3320, 2013.
- [59] J. L. M. Dillen. An empirical force-field .1. alkanes. *Journal of Computational Chemistry*, 16(5):595–609, 1995.
- [60] Robert A. DiStasio, O. Anatole von Lilienfeld, and Alexandre Tkatchenko. Collective many-body van der waals interactions in molecular systems. *Proceedings of the National Academy of Sciences*, 109(37):14791, 2012.

- [61] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1):429–452, 2012.
- [62] Todor Dudev and Carmay Lim. Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chemical Reviews*, 114(1):538–556, 2014.
- [63] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *Journal of Chemical Physics*, 90:1007–1023, 1989.
- [64] Eric C. Dybeck, Natalie P. Schieber, and Michael R. Shirts. Effects of a more accurate polarizable hamiltonian on polymorph free energies computed efficiently by reweighting point-charge potentials. *Journal of Chemical Theory and Computation*, 12(8):3491–3505, 2016.
- [65] Peter Eastman, Mark S. Friedrichs, John D. Chodera, Randall J. Radmer, Christopher M. Bruns, Joy P. Ku, Kyle A. Beauchamp, Thomas J. Lane, Lee-Ping Wang, Diwakar Shukla, Tony Tye, Mike Houston, Timo Stich, Christoph Klein, Michael R. Shirts, and Vijay S. Pande. Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation*, 9(1):461–469, 2013.
- [66] Peter Eastman and Vijay S. Pande. Openmm: A hardware-independent

- framework for molecular simulations. *Computing in Science Engineering*, 12(4):34–39, 2010.
- [67] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Bernard R. Brooks, and Vijay S. Pande. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *bioRxiv*, 2016.
- [68] David S. Eisenberg and Walter Kauzmann. *The structure and properties of water*. Clarendon P., Oxford,, 1969.
- [69] Ron Elber. Perspective: Computer simulations of long time dynamics. *The Journal of Chemical Physics*, 144(6):060901, 2016.
- [70] Mikael Elias, Alon Wellner, Korina Goldin-Azulay, Eric Chabriere, Julia A. Vorholt, Tobias J. Erb, and Dan S. Tawfik. The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature*, 491(7422):134–137, 2012.
- [71] E. M. Engler, J. D. Andose, and P. V. Schleyer. Critical evaluation of molecular mechanics. *Journal of the American Chemical Society*, 95(24):8005–8025, 1973.
- [72] Alexander Esser, Saurabh Belsare, Dominik Marx, and Teresa Head-Gordon. Mode specific thz spectra of solvated amino acids using the

- amoeba polarizable force field. *Physical Chemistry Chemical Physics*, 19(7):5579–5590, 2017.
- [73] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *Journal of Chemical Physics*, 103(19):8577–8593, 1995.
- [74] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.
- [75] G. S. Fanourgakis, E. Apra, W. A. de Jong, and S. S. Xantheas. High-level ab initio calculations for the four low-lying families of minima of (h<sub>2</sub>o)<sub>20</sub>. ii. spectroscopic signatures of the dodecahedron, fused cubes, face-sharing pentagonal prisms, and edge-sharing pentagonal prisms hydrogen bonding networks. *Journal of Chemical Physics*, 122:134304, 2005.
- [76] G. S. Fanourgakis, E. Apra, and S. S. Xantheas. High-level ab initio calculations for the four low-lying families of minima of (h<sub>2</sub>o)<sub>20</sub>. i. estimates of mp2/cbs binding energies and comparison with empirical potentials. *Journal of Chemical Physics*, 121:2655–2663, 2004.
- [77] G. S. Fanourgakis and S. S. Xantheas. Development of transferable interaction potentials for water. v. extension of the flexible, polarizable, thole-type model potential (ttm3-f, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *Journal of Chemical Physics*, 128(7), 2008.

- [78] Anton K. Faradjian and Ron Elber. Computing time scales from reaction coordinates by milestoning. *The Journal of Chemical Physics*, 120(23):10880–10889, 2004.
- [79] J. C. Faver, M. L. Benson, X. A. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. Merz. Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes. *Journal of Chemical Theory and Computation*, 7(3):790–797, 2011.
- [80] G.; Schlegel H. B.; Scuseria G.; Robb M.; Cheeseman J.; Scalmani G.; Barone V.; Mennucci B.; Petersson G. Inc Wallingford CT 2009 200. Frisch, M.; Trucks. Gaussian 09.
- [81] Kratz Eric G., Walker Alice R., Lagardre Louis, Lipparini Filippo, Piquemal JeanPhilip, and Andrs Cisneros G. Lichem: A qm/mm program for simulations with multipolar and polarizable force fields. *Journal of Computational Chemistry*, 37(11):1019–1029, 2016.
- [82] N. Gailar and E. K. Plyler. Rotation-vibration spectra of deuterated water vapor. *Journal of Chemical Physics*, 24(6):1139–1165, 1956.
- [83] M. Gerstein and M. Levitt. Simulating water and the molecules of life. *Scientific American*, 279(5):100–105, 1998.
- [84] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: A

- critical review. *Biophysical Journal*, 72(3):1047–1069, 1997.
- [85] M. K. Gilson and H. X. Zhou. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*, 36:21–42, 2007.
- [86] Boon Chong Goh, Jodi A. Hadden, Rafael C. Bernardi, Abhishek Singharoy, Ryan McGreevy, Till Rudack, C. Keith Cassidy, and Klaus Schulten. Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annual Review of Biophysics*, 45(1):253–278, 2016.
- [87] Holger Gohlke and Gerhard Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie (International ed. in English)*, 41(15):2644–76, 2002.
- [88] P. A. Golubkov and P. Y. Ren. Generalized coarse-grained model based on gay-berne and point multipole potentials. *Abstracts of Papers of the American Chemical Society*, 232:274–274, 2006.
- [89] P. A. Golubkov and P. Y. Ren. Generalized coarse-grained model based on point multipole and gay-berne potentials. *Journal of Chemical Physics*, 125(6), 2006.
- [90] P. A. Golubkov, J. C. Wu, and P. Y. Ren. A transferable coarse-grained model for hydrogen-bonding liquids. *Physical Chemistry Chemical Physics*, 10(15):2050–2057, 2008.

- [91] D. Gonzalez, M. Elias, and E. Chabriere. The ding family of phosphate binding proteins in inflammatory diseases. *Oxidative Stress and Inflammation in Non-Communicable Diseases - Molecular Mechanisms and Perspectives in Therapeutics*, 824:27–32, 2014.
- [92] M. A. Gonzalez and J. L. F. Abascal. The shear viscosity of rigid water models. *Journal of Chemical Physics*, 132(9), 2010.
- [93] F. Grater, S. M. Schwarzl, A. Dejaegere, S. Fischer, and J. C. Smith. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *Journal of Physical Chemistry B*, 109(20):10474–10483, 2005.
- [94] N. Gresh, G. A. Cisneros, T. A. Darden, and J. P. Piquemal. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. a bottom-up strategy. *Journal of Chemical Theory and Computation*, 3(6):1960–1986, 2007.
- [95] Nohad Gresh, G. Andrs Cisneros, Thomas A. Darden, and Jean-Philip Piquemal. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligandmacromolecule complexes. a bottom-up strategy. *Journal of Chemical Theory and Computation*, 3(6):1960–1986, 2007.
- [96] A. Grossfield, P. Y. Ren, and J. W. Ponder. Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society*, 125(50):15671–15682, 2003.



- [97] Alan Grossfield, Pengyu Ren, and Jay W. Ponder. Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society*, 125(50):15671–15682, 2003.
- [98] M. Gruber, P. Greisen, C. M. Junker, and C. Helix-Nielsen. Phosphorus binding sites in proteins: Structural preorganization and coordination. *Journal of Physical Chemistry B*, 118(5):1207–1215, 2014.
- [99] Mathias F. Gruber, Elizabeth Wood, Sigurd Truelsen, Thomas ster-gaard, and Claus Hlix-Nielsen. Computational design of biomimetic phosphate scavengers. *Environmental Science Technology*, 49(16):9469–9478, 2015.
- [100] MATLAB Users Guide. The mathworks. *Inc., Natick, MA*, 5, 1998.
- [101] B. Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids*, 101(1-3):219–260, 2002.
- [102] James C. Gumbart, Benot Roux, and Christophe Chipot. Standard binding free energies from computer simulations: What is the best strategy? *Journal of Chemical Theory and Computation*, 9(1):794–802, 2013.
- [103] Hatice Gkcan, Eric Kratz, Thomas A. Darden, Jean-Philip Piquemal, and G. Andrs Cisneros. Qm/mm simulations with the gaussian electrostatic model: A density-based polarizable potential. *The Journal of Physical Chemistry Letters*, 9(11):3062–3067, 2018.

- [104] T. A. Halgren. Representation of vanderwaals (vdw) interactions in molecular mechanics force-fields - potential form, combination rules, and vdw parameters. *Journal of the American Chemical Society*, 114(20):7827–7843, 1992.
- [105] T. A. Halgren. Merck molecular force field .2. mmff94 van der waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, 17(5-6):520–552, 1996.
- [106] D. Hamelberg and J. A. McCammon. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *Journal of the American Chemical Society*, 126(24):7683–7689, 2004.
- [107] E. Harder, W. Damm, J. Maple, C. J. Wu, M. Reboul, J. Y. Xiang, L. L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner. Opls3: A force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation*, 12(1):281–296, 2016.
- [108] Matthew Harger, Daniel Li, Zhi Wang, Kevin Dalby, Louis Lagardre, Jean-Philip Piquemal, Jay Ponder, and Pengyu Ren. Tinker-openmm: Absolute and relative alchemical free energies using amoeba on gpus. *Journal of Computational Chemistry*, 38(23):2047–2055, 2017.

- [109] A. E. Hargrove, S. Nieto, T. Z. Zhang, J. L. Sessler, and E. V. Anslyn. Artificial receptors for the recognition of phosphorylated molecules. *Chemical Reviews*, 111(11):6603–6782, 2011.
- [110] Amanda E. Hargrove, Sonia Nieto, Tianzhi Zhang, Jonathan L. Sessler, and Eric V. Anslyn. Artificial receptors for the recognition of phosphorylated molecules. *Chemical Reviews*, 111(11):6603–6782, 2011.
- [111] P. H. Hemmerich and A. H. von Mikecz. Defining the subcellular interface of nanoparticles by live-cell imaging. *Plos One*, 8(4), 2013.
- [112] Niel M. Henriksen, Andrew T. Fenley, and Michael K. Gilson. Computational calorimetry: High-precision calculation of hostguest binding thermodynamics. *Journal of Chemical Theory and Computation*, 11(9):4377–4394, 2015.
- [113] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.
- [114] A. K. Hirsch, F. R. Fischer, and F. Diederich. Phosphate recognition in structural biology. *Angew Chem Int Ed Engl*, 46(3):338–52, 2007.
- [115] A. K. H. Hirsch, F. R. Fischer, and F. Diederich. Phosphate recognition in structural biology. *Angewandte Chemie-International Edition*, 46(3):338–352, 2007.

- [116] P. Hocht, S. Boresch, W. Bitomsky, and O. Steinhauser. Rationalization of the dielectric properties of common three-site water models in terms of their force field parameters. *Journal of Chemical Physics*, 109(12):4927–4937, 1998.
- [117] E. G. Hohenstein and C. D. Sherrill. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *Journal of Chemical Physics*, 133(1), 2010.
- [118] H. W. Horn, W. Swope, J. Pitner, J. D. Madura, T. J. Dick, G. L. B. Hura, and T. Head-Gordon. Development of an improved four-site water model for bio-molecular simulations: Tip4p-ew. *Abstracts of Papers of the American Chemical Society*, 228:U530–U531, 2004.
- [119] H. W. Horn, W. C. Swope, J. W. Pitner, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. *Journal of Chemical Physics*, 120(20):9665–9678, 2004.
- [120] K. N. Houk, A. G. Leach, S. P. Kim, and X. Y. Zhang. Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. *Angewandte Chemie-International Edition*, 42(40):4872–4897, 2003.
- [121] Jing Huang, Andrew C. Simmonett, Frank C. Pickard, Alexander D. MacKerell, and Bernard R. Brooks. Mapping the drude polarizable force field onto a multipole and induced dipole model. *The Journal of Chemical Physics*, 147(16):161702, 2017.

- [122] E. C. Hulme and M. A. Trevethick. Ligand binding assays at equilibrium: validation and interpretation. *British Journal of Pharmacology*, 161(6):1219–1237, 2010.
- [123] T. Hunter. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 367(1602):2513–2516, 2012.
- [124] G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon. A high-quality x-ray scattering experiment on liquid water at ambient conditions. *Journal of Chemical Physics*, 113(20):9140–9148, 2000.
- [125] Brooke E. Husic and Vijay S. Pande. Markov state models: From an art to a science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018.
- [126] M. J. Hwang, T. P. Stockfish, and A. T. Hagler. Derivation of class-ii force-fields .2. derivation and characterization of a class-ii force-field, cff93, for the alkyl functional-group and alkane molecules. *Journal of the American Chemical Society*, 116(6):2515–2525, 1994.
- [127] T. Ichiye and M. L. Tan. Soft sticky dipole-quadrupole-octupole potential energy function for liquid water: An approximate moment expansion. *Journal of Chemical Physics*, 124(13), 2006.
- [128] S. Imoto, S. S. Xantheas, and S. Saito. Molecular origin of the difference in the hoh bend of the ir spectra between liquid water and ice. *Journal*

- of Chemical Physics*, 138(5), 2013.
- [129] J. Israelachvili and H. Wennerstrom. Role of hydration and water structure in biological and colloidal interactions. *Nature*, 379(6562):219–25, 1996.
- [130] S. Izadi, R. Anandakrishnan, and A. V. Onufriev. Building water models: A different approach. *Journal of Physical Chemistry Letters*, 5(21):3863–3871, 2014.
- [131] Sofie Jakobsen and Frank Jensen. Systematic improvement of potential-derived atomic multipoles and redundancy of the electrostatic parameter space. *Journal of Chemical Theory and Computation*, 10(12):5493–5504, 2014.
- [132] George A. Jeffrey. *An introduction to hydrogen bonding*. Topics in physical chemistry. Oxford University Press, New York, 1997.
- [133] Y. J. Jeon, S. Y. Kim, Y. H. Ko, S. Sakamoto, K. Yamaguchi, and K. Kim. Novel molecular drug carrier: encapsulation of oxaliplatin in cucurbit 7 uril and its effects on stability and reactivity of the drug. *Organic Biomolecular Chemistry*, 3(11):2122–2125, 2005.
- [134] B. Jeziorski, R. Moszynski, and K. Szalewicz. Perturbation-theory approach to intermolecular potential-energy surfaces of van-der-waals complexes. *Chemical Reviews*, 94(7):1887–1930, 1994.

- [135] D. Jiao, P. A. Golubkov, T. A. Darden, and P. Ren. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences of the United States of America*, 105(17):6290–6295, 2008.
- [136] D. Jiao, J. J. Zhang, R. E. Duke, G. H. Li, M. J. Schnieders, and P. Y. Ren. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *Journal of Computational Chemistry*, 30(11):1701–1711, 2009.
- [137] J. E. Jones. On the determination of molecular fields iii - from crystal measurements and kinetic theory data. *Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character*, 106(740):709–718, 1924.
- [138] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids .10. transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. *Journal of the American Chemical Society*, 103(2):335–340, 1981.
- [139] W. L. Jorgensen. The many roles of computation in drug discovery. *Science*, 303(5665):1813–1818, 2004.
- [140] W. L. Jorgensen. Efficient drug lead discovery and optimization. *Accounts of Chemical Research*, 42(6):724–733, 2009.

- [141] W. L. Jorgensen. Foundations of biomolecular modeling. *Cell*, 155(6):1199–1202, 2013.
- [142] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926–935, 1983.
- [143] W. L. Jorgensen, J. D. Madura, and C. J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society*, 106(22):6638–6646, 1984.
- [144] W. L. Jorgensen, 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*, 118(45):11225–11236, 1996.
- [145] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza. Benchmark database of accurate (mp2 and ccSD(T) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics*, 8(17):1985–1993, 2006.
- [146] S. C. L. Kamerlin, P. K. Sharma, R. B. Prasad, and A. Warshel. Why nature really chose phosphate. *Quarterly Reviews of Biophysics*, 46(1):1–132, 2013.
- [147] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins



- via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.
- [148] Zigui Kan, Qiang Zhu, Lijiang Yang, Zhixiong Huang, Biaobing Jin, and Jing Ma. Polarization effects on the cellulose dissolution in ionic liquids: Molecular dynamics simulations with polarization model and integrated tempering enhanced sampling method. *The Journal of Physical Chemistry B*, 121(17):4319–4332, 2017.
- [149] F. N. Keutsch and R. J. Saykally. Water clusters: Untangling the mysteries of the liquid, one molecule at a time. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10533–10540, 2001.
- [150] P. T. Kiss and A. Baranyai. A systematic development of a polarizable potential of water. *Journal of Chemical Physics*, 138(20), 2013.
- [151] Douglas B. Kitchen, Helene Decornez, John R. Furr, and Jurgen Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3(11):935–949, 2004.
- [152] Miriam Kohagen, Martin Lepk, and Pavel Jungwirth. Calcium binding to calmodulin by molecular dynamics with effective polarization. *The Journal of Physical Chemistry Letters*, 5(22):3964–3969, 2014.
- [153] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srin-

- vasan, D. A. Case, and T. E. Cheatham. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, 33(12):889–897, 2000.
- [154] Michal H. Kol and Pavel Hobza. Computer modeling of halogen bonds and other -hole interactions. *Chemical Reviews*, 116(9):5155–5187, 2016.
- [155] X. J. Kong and C. L. Brooks. lambda-dynamics: A new approach to free energy calculations. *Journal of Chemical Physics*, 105(6):2414–2423, 1996.
- [156] Christian Kramer, Alexander Spinn, and Klaus R. Liedl. Charge anisotropy: Where atomic multipoles matter most. *Journal of Chemical Theory and Computation*, 10(10):4488–4496, 2014.
- [157] B. D. Kubena, H. Luecke, H. Rosenberg, and F. A. Quioco. Crystallization and x-ray-diffraction studies of a phosphate-binding protein involved in active-transport in escherichia-coli. *Journal of Biological Chemistry*, 261(17):7995–7996, 1986.
- [158] R. Kumar, F. F. Wang, G. R. Jenness, and K. D. Jordan. A second generation distributed point polarizable water model (vol 132, 014309, 2010). *Journal of Chemical Physics*, 132(13), 2010.
- [159] Igor V. Kurnikov and Maria Kurnikova. Modeling electronic polarizability changes in the course of a magnesium ion water ligand exchange

- process. *The Journal of Physical Chemistry B*, 119(32):10275–10286, 2015.
- [160] G. Lamoureux, A. D. MacKerell, and B. Roux. A simple polarizable model of water based on classical drude oscillators. *Journal of Chemical Physics*, 119(10):5185–5197, 2003.
- [161] M. L. Laury, L. P. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder. Revised parameters for the amoeba polarizable atomic multipole water model. *J Phys Chem B*, 2015.
- [162] Marie L. Laury, Lee-Ping Wang, Vijay S. Pande, Teresa Head-Gordon, and Jay W. Ponder. Revised parameters for the amoeba polarizable atomic multipole water model. *The Journal of Physical Chemistry B*, 119(29):9423–9437, 2015.
- [163] M. Lawrenz, J. Wereszczynski, J. M. Ortiz-Sanchez, S. E. Nichols, and J. A. McCammon. Thermodynamic integration to predict host-guest binding affinities. *Journal of Computer-Aided Molecular Design*, 26(5):569–576, 2012.
- [164] P. S. Ledvina, A. L. Tsai, Z. M. Wang, E. Koehl, and F. A. Quiocho. Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: Equilibrium, kinetic, and crystallographic studies. *Protein Science*, 7(12):2550–2559, 1998.

- [165] J. W. Lee, S. Samal, N. Selvapalam, H. J. Kim, and K. Kim. Cucurbituril homologues and derivatives: New opportunities in supramolecular chemistry. *Accounts of Chemical Research*, 36(8):621–630, 2003.
- [166] M. S. Lee, F. R. Salsbury, and C. L. Brooks. Constant-ph molecular dynamics using continuous titration coordinates. *Proteins-Structure Function and Bioinformatics*, 56(4):738–752, 2004.
- [167] Justin A. Lemkul, Jing Huang, Benot Roux, and Alexander D. MacKerell. An empirical polarizable force field based on the classical drude oscillator model: Development history and recent applications. *Chemical Reviews*, 2016.
- [168] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- [169] Elisa Liberatore, Rocco Meli, and Ursula Rothlisberger. A versatile multiple time step scheme for efficient ab initio molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 14(6):2834–2842, 2018.
- [170] Dorothee Liebschner, Mikael Elias, Sbastien Moniot, Bertrand Fournier, Ken Scott, Christian Jelsch, Benoit Guillot, Claude Lecomte, and Eric Chabriere. Elucidation of the phosphate binding mode of ding proteins revealed by subangstrom x-ray crystallography. *Journal of the American Chemical Society*, 131(22):7879–7886, 2009.

- [171] S. Lifson, A. T. Hagler, and P. Dauber. Consistent force-field studies of inter-molecular forces in hydrogen-bonded crystals .1. carboxylic-acids, amides, and the c=o...h- hydrogen-bonds. *Journal of the American Chemical Society*, 101(18):5111–5121, 1979.
- [172] T. C. Lim. Scaling function between the exponential-6 and the generalized lennard-jones potential functions. *Journal of Mathematical Chemistry*, 33(3-4):279–285, 2003.
- [173] Filippo Lipparini, Louis Lagardre, Christophe Raynaud, Benjamin Stamm, Eric Cancs, Benedetta Mennucci, Michael Schnieders, Pengyu Ren, Yvon Maday, and Jean-Philip Piquemal. Polarizable molecular dynamics in a polarizable continuum solvent. *Journal of Chemical Theory and Computation*, 11(2):623–634, 2015.
- [174] Chengwen Liu, Rui Qi, Qiantao Wang, J-P Piquemal, and Pengyu Ren. Capturing many-body interactions with classical dipole induction models. *Journal of Chemical Theory and Computation*, 13(6):27512761, 2017.
- [175] Cui Liu, Yue Li, Bing-Yu Han, Li-Dong Gong, Li-Nan Lu, Zhong-Zhi Yang, and Dong-Xia Zhao. Development of the abeam polarization force field for base pairs with amino acid residue complexes. *Journal of Chemical Theory and Computation*, 13(5):2098–2111, 2017.
- [176] Y. Liu and T. Ichiye. Soft sticky dipole potential for liquid water: A new model. *Journal of Physical Chemistry*, 100(7):2723–2730, 1996.

- [177] Daniele Loco, Francesco Buda, Johan Lugtenburg, and Benedetta Mennucci. The dynamic origin of color tuning in proteins revealed by a carotenoid pigment. *The Journal of Physical Chemistry Letters*, 9(9):2404–2410, 2018.
- [178] Daniele Loco, Louis Lagardre, Stefano Caprasecca, Filippo Lipparini, Benedetta Mennucci, and Jean-Philip Piquemal. Hybrid qm/mm molecular dynamics with amoeba polarizable embedding. *Journal of Chemical Theory and Computation*, 13(9):4025–4033, 2017.
- [179] Daniele Loco, tienne Polack, Stefano Caprasecca, Louis Lagardre, Filippo Lipparini, Jean-Philip Piquemal, and Benedetta Mennucci. A qm/mm approach using the amoeba polarizable embedding: From ground state energies to electronic excitations. *Journal of Chemical Theory and Computation*, 12(8):3654–3661, 2016.
- [180] Hartmut Luecke and Florante A. Quiocho. High specificity of a phosphate-transport protein determined by hydrogen-bonds. *Nature*, 347(6291):402–406, 1990.
- [181] Chao Lv, Xubin Li, Dongsheng Wu, Lianqing Zheng, and Wei Yang. Predictive sampling of rare conformational events in aqueous solution: Designing a generalized orthogonal space tempering method. *Journal of Chemical Theory and Computation*, 12(1):41–52, 2016.
- [182] Baker Christopher M. Polarizable force fields for molecular dynamics

- simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(2):241–254, 2015.
- [183] M. W. Mahoney 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*, 112(20):8910–8922, 2000.
- [184] N. Manin, M. C. da Silva, I. Zdravkovic, O. Eliseeva, A. Dyshin, O. Yasar, D. R. Salahub, A. M. Kolker, M. G. Kiselev, and S. Y. Noskov. LiCl solvation in n-methyl-acetamide (nma) as a model for understanding li(+) binding to an amide plane. *Phys Chem Chem Phys*, 18(5):4191–200, 2016.
- [185] Kumar Manjeet, Simonson Thomas, Ohanessian Gilles, and Clavagura Carine. Structure and thermodynamics of mg:phosphate interactions in water: A simulation study. *ChemPhysChem*, 16(3):658–665, 2015.
- [186] Y. Mao, O. Demerdash, M. Head-Gordon, and T. Head-Gordon. Assessing ion-water interactions in the amoeba force field using energy decomposition analysis of electronic structure calculations. *J Chem Theory Comput*, 12(11):5422–5437, 2016.
- [187] Y. J. Mao and Y. W. Zhang. Thermal conductivity, shear viscosity and specific heat of rigid water models. *Chemical Physics Letters*, 542:37–41, 2012.

- [188] Daniel T. Margul and Mark E. Tuckerman. A stochastic, resonance-free multiple time-step algorithm for polarizable models that permits very large time steps. *Journal of Chemical Theory and Computation*, 12(5):2170–2180, 2016.
- [189] D. W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11:431–441, 1963.
- [190] Garland R. Marshall. Limiting assumptions in molecular modeling: electrostatics. *Journal of Computer-Aided Molecular Design*, 27(2):107–114, 2013.
- [191] M. S. Marshall, L. A. Burns, and C. D. Sherrill. Basis set convergence of the coupled-cluster correction,  $\Delta(\text{ccsd}(t))(\text{mp2})$ : Best practices for benchmarking non-covalent interactions and the attendant revision of the s22, nbc10, hbc6, and hsg databases. *Journal of Chemical Physics*, 135(19), 2011.
- [192] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. Explicit reversible integrators for extended systems dynamics. *Molecular Physics*, 87(5):1117–1157, 1996.
- [193] E. Masson, X. X. Ling, R. Joseph, L. Kyeremeh-Mensah, and X. Y. Lu. Cucurbituril chemistry: a tale of supramolecular success. *Rsc Advances*, 2(4):1213–1247, 2012.



- [194] S. L. Mayo, B. D. Olafson, and W. A. Goddard. Dreiding - a generic force-field for molecular simulations. *Journal of Physical Chemistry*, 94(26):8897–8909, 1990.
- [195] A. Y. Mehandzhiyski, E. Riccardi, T. S. van Erp, H. Koch, P. O. Astrand, T. T. Trinh, and B. A. Grimes. Density functional theory study on the interactions of metal ions with long chain deprotonated carboxylic acids. *J Phys Chem A*, 119(40):10195–203, 2015.
- [196] P. Mikulskis, D. Cioloboc, M. Andrejic, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Soderhjelm, and U. Ryde. Free-energy perturbation and quantum mechanical study of sampl4 octa-acid host-guest binding energies. *Journal of Computer-Aided Molecular Design*, 28(4):375–400, 2014.
- [197] Donghong Min, Lianqing Zheng, William Harris, Mengen Chen, Chao Lv, and Wei Yang. Practically efficient qm/mm alchemical free energy simulations: The orthogonal space random walk strategy. *Journal of Chemical Theory and Computation*, 6(8):2253–2266, 2010.
- [198] Niraj Modi, Roland Benz, Robert E. W. Hancock, and Ulrich Kleinekathfer. Modeling the ion selectivity of the phosphate specific channel oprp. *The Journal of Physical Chemistry Letters*, 3(23):3639–3645, 2012.
- [199] Niraj Modi, Ivn Brcena-Uribarri, Manjeet Bains, Roland Benz, Robert E. W. Hancock, and Ulrich Kleinekathfer. Tuning the affinity of anion

- binding sites in porin channels with negatively charged residues: Molecular details for oprp. *ACS Chemical Biology*, 10(2):441–451, 2015.
- [200] V. Molinero and E. B. Moore. Water modeled as an intermediate element between carbon and silicon. *Journal of Physical Chemistry B*, 113(13):4008–4016, 2009.
- [201] J. Mongan, D. A. Case, and J. A. McCammon. Constant ph molecular dynamics in generalized born implicit solvent. *Journal of Computational Chemistry*, 25(16):2038–2048, 2004.
- [202] J. I. Monroe and M. R. Shirts. Converging free energies of binding in cucurbit 7 uril and octa-acid host-guest systems from sampl4 using expanded ensemble simulations. *Journal of Computer-Aided Molecular Design*, 28(4):401–415, 2014.
- [203] W. T. M. Mooij, F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and B. P. van Eijck. Transferable ab initio intermolecular potentials. 1. derivation from methanol dimer and trimer calculations. *Journal of Physical Chemistry A*, 103(48):9872–9882, 1999.
- [204] J. J. More and D. C. Sorensen. Computing a trust region. *SIAM Journal on Scientific and Statistical Computing*, 4:553–572, 1983.
- [205] H. Morhenn, S. Busch, H. Meyer, D. Richter, W. Petry, and T. Unruh. Collective intermolecular motions dominate the picosecond dynamics of short polymer chains. *Phys Rev Lett*, 111(17):173003, 2013.

- [206] Joseph A. Morrone, Thomas E. Markland, Michele Ceriotti, and B. J. Berne. Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances. *The Journal of Chemical Physics*, 134(1):014103, 2011.
- [207] Jrmie Mortier, Christin Rakers, Marcel Bermudez, Manuela S. Murgueitio, Sereina Riniker, and Gerhard Wolber. The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discovery Today*, 20(6):686–702, 2015.
- [208] R. D. Mountain and D. Thirumalai. Ergodic measures for the simulation of dielectric-properties of water. *Computer Physics Communications*, 62(2-3):352–359, 1991.
- [209] X. J. Mu, Q. T. Wang, L. P. Wang, S. D. Fried, J. P. Piquemal, K. N. Dalby, and P. Y. Ren. Modeling organochlorine compounds and the sigma-hole effect using a polarizable multipole force field. *Journal of Physical Chemistry B*, 118(24):6456–6465, 2014.
- [210] Xiaojiao Mu, Qiantao Wang, Lee-Ping Wang, Stephen D. Fried, Jean-Philip Piquemal, Kevin N. Dalby, and Pengyu Ren. Modeling organochlorine compounds and the -hole effect using a polarizable multipole force field. *The Journal of Physical Chemistry B*, 118(24):6456–6465, 2014.
- [211] H. S. Muddana, C. D. Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe, and M. K. Gilson. Blind prediction of host-guest binding

- affinities: a new sampl3 challenge. *Journal of Computer-Aided Molecular Design*, 26(5):475–487, 2012.
- [212] H. S. Muddana, J. Yin, N. V. Sapra, A. T. Fenley, and M. K. Gilson. Blind prediction of sampl4 cucurbit[7]uril binding affinities with the mining minima method. *Journal of Computer-Aided Molecular Design*, 28(4):463–474, 2014.
- [213] Hari S. Muddana, Andrew T. Fenley, David L. Mobley, and Michael K. Gilson. The sampl4 host-guest blind prediction challenge: an overview. *Journal of Computer-Aided Molecular Design*, 28(4):305–317, 2014.
- [214] W. F. Murphy. Rotation-vibration raman-spectrum of water-vapor. *Journal of the Optical Society of America*, 67(10):1397–1397, 1977.
- [215] T. Negami, K. Shimizu, and T. Terada. Coarse-grained molecular dynamics simulations of protein-ligand binding. *Journal of Computational Chemistry*, 35(25):1835–1845, 2014.
- [216] G. Nemethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga. Energy parameters in polypeptides .10. improved geometrical parameters and nonbonded interactions for use in the ecepp/3 algorithm, with application to proline-containing peptides. *Journal of Physical Chemistry*, 96(15):6472–6484, 1992.

- [217] Ian J. Nessler, Jacob M. Litman, and Michael J. Schnieders. Toward polarizable amoeba thermodynamics at fixed charge efficiency using a dual force field approach: application to organic crystals. *Physical Chemistry Chemical Physics*, 18(44):30313–30322, 2016.
- [218] V. Ngo, M. C. da Silva, M. Kubillus, H. Li, B. Roux, M. Elstner, Q. Cui, D. R. Salahub, and S. Y. Noskov. Quantum effects in cation interactions with first and second coordination shell ligands in metalloproteins. *J Chem Theory Comput*, 11(10):4992–5001, 2015.
- [219] Sergei Yu Noskov, Simon Bernche, and Benot Roux. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature*, 431:830, 2004.
- [220] Demerdash Omar, Wang LeePing, and HeadGordon Teresa. Advanced models for water simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1355, 2018.
- [221] P. Paricaud, M. Predota, A. A. Chialvo, and P. T. Cummings. From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a gaussian charge polarizable model. *Journal of Chemical Physics*, 122(24), 2005.
- [222] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill. Levels of symmetry adapted perturbation theory (sapt). i. efficiency and performance for interaction energies. *Journal of Chemical Physics*, 140(9), 2014.

- [223] T. M. Parker and C. D. Sherrill. Assessment of empirical models versus high-accuracy ab initio methods for nucleobase stacking: Evaluating the importance of charge penetration. *Journal of Chemical Theory and Computation*, 11(9):4197–4204, 2015.
- [224] Robert M. Parrish, Lori A. Burns, Daniel G. A. Smith, Andrew C. Simmonett, A. Eugene DePrince, Edward G. Hohenstein, Uur Bozkaya, Alexander Yu Sokolov, Roberto Di Remigio, Ryan M. Richard, Jrme F. Gonthier, Andrew M. James, Harley R. McAlexander, Ashutosh Kumar, Masaaki Saitow, Xiao Wang, Benjamin P. Pritchard, Prakash Verma, Henry F. Schaefer, Konrad Patkowski, Rollin A. King, Edward F. Valeev, Francesco A. Evangelista, Justin M. Turney, T. Daniel Crawford, and C. David Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *Journal of Chemical Theory and Computation*, 13(7):3185–3197, 2017.
- [225] Sandeep Patel and Charles L. Brooks. Charmm fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *Journal of Computational Chemistry*, 25(1):1–16, 2004.
- [226] D. A. Pearlman and P. A. Kollman. The lag between the hamiltonian and the system configuration in free-energy perturbation calculations. *Journal of Chemical Physics*, 91(12):7831–7839, 1989.

- [227] Xiangda Peng, Yuebin Zhang, Huiying Chu, Yan Li, Dinglin Zhang, Liaoran Cao, and Guohui Li. Accurate evaluation of ion conductivity of the gramicidin a channel using a polarizable force field without any corrections. *Journal of Chemical Theory and Computation*, 12(6):2973–2982, 2016.
- [228] E. Persch, O. Dumele, and F. Diederich. Molecular recognition in chemical and biological systems. *Angewandte Chemie-International Edition*, 54(11):3290–3327, 2015.
- [229] G. D. J. Phillies. Self-consistency of hydrodynamic models for the zero-shear viscosity and the self-diffusion coefficient. *Macromolecules*, 35(19):7414–7418, 2002.
- [230] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Protein Simulations*, 66:27–+, 2003.
- [231] J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon. Current status of the amoeba polarizable force field. *Journal of Physical Chemistry B*, 114(8):2549–2564 PMID: PMC2918242, 2010.
- [232] Jay W. Ponder, Chuanjie Wu, Pengyu Ren, Vijay S. Pande, John D. Chodera, Michael J. Schnieders, Imran Haque, David L. Mobley, Daniel S. Lambrecht, Robert A. DiStasio, Martin Head-Gordon, Gary N. I. Clark,

- Margaret E. Johnson, and Teresa Head-Gordon. Current status of the amoeba polarizable force field. *The Journal of Physical Chemistry B*, 114(8):2549–2564, 2010.
- [233] M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen. Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the opl-aa force field. *Journal of Computational Chemistry*, 22(13):1340–1352, 2001.
- [234] Rui Qi, Zhifeng Jing, Chengwen Liu, Jean-Philip Piquemal, Kevin N. Dalby, and Pengyu Ren. Elucidating the phosphate binding mode of pbp: The critical effect of buffer solution. *The Journal of Physical Chemistry B*, 2018.
- [235] Rui Qi, Lee-Ping Wang, Qiantao Wang, Vijay S. Pande, and Pengyu Ren. United polarizable multipole water model for molecular mechanics simulation. *The Journal of Chemical Physics*, 143(1):014504, 2015.
- [236] Rui Qi, Qiantao Wang, and Pengyu Ren. General van der waals potential for common organic molecules. *Bioorganic Medicinal Chemistry*, 24(20):4911–4919, 2016.
- [237] F. A. Quiocho. Atomic basis of the exquisite specificity of phosphate and sulfate transport receptors. *Kidney International*, 49(4):943–946, 1996.



- [238] Joshua A Rackers, Qiantao Wang, Chengwen Liu, Jean-Philip Piquemal, Pengyu Ren, and Jay W Ponder. An optimized charge penetration model for use with the amoeba force field. *Physical Chemistry Chemical Physics*, 19:276–291, 2017.
- [239] Li Rao, Qiang Cui, and Xin Xu. Electronic properties and desolvation penalties of metal ions plus protein electrostatics dictate the metal binding affinity and selectivity in the copper efflux regulator. *Journal of the American Chemical Society*, 132(51):18092–18102, 2010.
- [240] N. N. Rao and A. Torriani. Molecular aspects of phosphate-transport in escherichia-coli. *Molecular Microbiology*, 4(7):1083–1090, 1990.
- [241] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.
- [242] Maria M. Reif, Philippe H. Hnenberger, and Chris Oostenbrink. New interaction parameters for charged amino acid side chains in the gromos force field. *Journal of Chemical Theory and Computation*, 8(10):3705–3723, 2012.
- [243] P. Y. Ren and J. W. Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry*, 23(16):1497–1506, 2002.

- [244] P. Y. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.
- [245] P. Y. Ren, C. J. Wu, and J. W. Ponder. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation*, 7(10):3143–3161, 2011.
- [246] Pengyu Ren, Jaehun Chun, Dennis G. Thomas, Michael J. Schnieders, Marcelo Marucho, Jiajing Zhang, and Nathan A. Baker. Biomolecular electrostatics and solvation: a computational perspective. *Quarterly Reviews of Biophysics*, 45(4):427–491, 2012.
- [247] Pengyu Ren and Jay W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.
- [248] Pengyu Ren, Chuanjie Wu, and Jay W. Ponder. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation*, 7(10):3143–3161, 2011.
- [249] J. Rezac, 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*, 7(8):2427–2438, 2011.
- [250] J. Rezac, K. E. Riley, and P. Hobza. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures

- (vol 7, pg 2427, 2011). *Journal of Chemical Theory and Computation*, 10(3):1359–1360, 2014.
- [251] S. W. Rick. A reoptimization of the five-site water potential (tip5p) for use with ewald sums. *Journal of Chemical Physics*, 120(13):6085–6093, 2004.
- [252] S. W. Rick and S. J. Stuart. Potentials and algorithms for incorporating polarizability in computer simulations. *Reviews in Computational Chemistry, Vol 18*, 18:89–146, 2002.
- [253] S. W. Rick and S. J. Stuart. *Potentials and algorithms for incorporating polarizability in computer simulations*, volume 18 of *Reviews in Computational Chemistry*, pages 89–146. Wiley-Vch, Inc, New York, 2002.
- [254] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen. Improved peptide and protein torsional energetics with the opls-aa force field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, 2015.
- [255] C. Sagui, 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. *J Chem Phys*, 120(1):73–87, 2004.
- [256] Dmitri V. Sakharov and Carmay Lim. Zn protein simulations including charge transfer and local polarization effects. *Journal of the American*

*Chemical Society*, 127(13):4921–4929, 2005.

- [257] Romelia Salomon-Ferrer, Andreas W. Gtz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, 2013.
- [258] Patel Sandeep and Brooks Charles L. Charmm fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *Journal of Computational Chemistry*, 25(1):1–16, 2004.
- [259] P. Satpati, C. Clavaguera, G. Ohanessian, and T. Simonson. Free energy simulations of a gtpase: Gtp and gdp binding to archaeal initiation factor 2. *Journal of Physical Chemistry B*, 115(20):6749–6763, 2011.
- [260] Priyadarshi Satpati, Carine Clavagura, Gilles Ohanessian, and Thomas Simonson. Free energy simulations of a gtpase: Gtp and gdp binding to archaeal initiation factor 2. *The Journal of Physical Chemistry B*, 115(20):6749–6763, 2011.
- [261] T. Schnabel, J. Vrabec, and H. Hasse. Unlike lennard-jones parameters for vapor-liquid equilibria. *Journal of Molecular Liquids*, 135(1-3):170–178, 2007.
- [262] M. J. Schnieders, N. A. Baker, P. Y. Ren, and J. W. Ponder. Polarizable atomic multipole solutes in a poisson-boltzmann continuum. *Journal of*

*Chemical Physics*, 126(12), 2007.

- [263] David Semrouni, William C. Isley, Carine Clavagura, Jean-Pierre Dognon, Christopher J. Cramer, and Laura Gagliardi. Ab initio extension of the amoeba polarizable force field to  $Fe^{2+}$ . *Journal of Chemical Theory and Computation*, 9(7):3062–3071, 2013.
- [264] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O’Neill, Jr. DiStasio, R. A., R. C. Lochan, T. Wang, G. J. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherill, A. C. Simmonett, J. E. Subotnik, 3rd Woodcock, H. L., W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, 3rd Schaefer, H. F., J. Kong, A. I. Krylov, P. M. Gill, and M. Head-Gordon. Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics*, 8:3172–3191, 2006.
- [265] Y. Shi, Z. Xia, J. J. Zhang, R. Best, C. J. Wu, J. W. Ponder, and P. Y.

- Ren. Polarizable atomic multipole-based amoeba force field for proteins. *Journal of Chemical Theory and Computation*, 9(9):4046–4063, 2013.
- [266] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren. Polarizable atomic multipole-based amoeba force field for proteins. *Journal of Chemical Theory and Computation*, 9(9):4046–4063, 2013.
- [267] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics*, 129(12), 2008.
- [268] L. B. Skinner, C. C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide q-range. *Journal of Chemical Physics*, 138(7), 2013.
- [269] B. J. Smith, D. J. Swanton, J. A. Pople, H. F. Schaefer, and L. Radom. Transition structures for the interchange of hydrogen atoms within the water dimer. *Journal of Chemical Physics*, 92:1240–1247, 1990.
- [270] A. K. Soper. The radial distribution functions of water and ice from 220 to 673 k and at pressures up to 400 mpa. *Chemical Physics*, 258(2-3):121–137, 2000.
- [271] A. K. Soper and M. G. Phillips. A new determination of the structure of water at 25-degrees-c. *Chemical Physics*, 107(1):47–60, 1986.

- [272] Oleg N. Starovoytov, Hedieh Torabifard, and G. Andrs Cisneros. Development of amoeba force field for 1,3-dimethylimidazolium based ionic liquids. *The Journal of Physical Chemistry B*, 118(25):7156–7166, 2014.
- [273] R. P. Steele, Jr. DiStasio, R. A., Y. Shao, J. Kong, and M. Head-Gordon. Dual-basis second-order moller-plesset perturbation theory: A reduced-cost reference for correlation calculations. *Journal of Chemical Physics*, 125:074108, 2006.
- [274] R. P. Steele, R. A. DiStasio, and M. Head-Gordon. Non-covalent interactions with dual-basis methods: Pairings for augmented basis sets. *Journal of Chemical Theory and Computation*, 5:1560–1572, 2009.
- [275] F. H. Stillinger and A. Rahman. Molecular-dynamics study of liquid water under high compression. *Journal of Chemical Physics*, 61(12):4973–4980, 1974.
- [276] A. J. Stone. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation*, 1(6):1128–1132, 2005.
- [277] M. Tafipolsky and B. Engels. Accurate intermolecular potentials with physically grounded electrostatics. *Journal of Chemical Theory and Computation*, 7(6):1791–1803, 2011.
- [278] J. A. Te and T. Ichiye. Temperature and pressure dependence of the optimized soft-sticky dipole-quadrupole-octupole water model. *Journal*

- of Chemical Physics*, 132(11), 2010.
- [279] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59(3):341–350, 1981.
- [280] Pratyush Tiwary, Vittorio Limongelli, Matteo Salvalaglio, and Michele Parrinello. Kinetics of proteinligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences*, 112(5):E386–E391, 2015.
- [281] Hedieh Torabifard and G. Andres Cisneros. Computational investigation of o2 diffusion through an intra-molecular tunnel in alkb; influence of polarization on o2 transport. *Chemical Science*, 8(9):6230–6238, 2017.
- [282] K. Toukan and A. Rahman. Molecular-dynamics study of atomic motions in water. *Physical Review B*, 31(5):2643–2648, 1985.
- [283] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular-dynamics. *Journal of Chemical Physics*, 97(3):1990–2001, 1992.
- [284] I. Tunon, M. T. C. MartinsCosta, C. Millot, and M. F. RuizLopez. A hybrid density functional classical molecular dynamics simulation of a water molecule in liquid water. *Journal of Molecular Modeling*, 1(4):196–201, 1995.
- [285] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L.



- Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford. Psi4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 2(4):556–565, 2012.
- [286] E. Tzoracoleftherakis, J. Maroulis, S. Katsanou, and J. Androulakis. Primary tamoxifen treatment for elderly women with operable breast cancer. *10th International Congress on Senology - Breast Diseases of the Senologic International Society*, pages 739–742, 1998.
- [287] J. A. Ubersax and Jr. Ferrell, J. E. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol*, 8(7):530–41, 2007.
- [288] Oliver T. Unke, Mike Devereux, and Markus Meuwly. Minimal distributed charges: Multipolar quality at the cost of point charge electrostatics. *The Journal of Chemical Physics*, 147(16):161712, 2017.
- [289] Piet Th van Duijnen and Marcel Swart. Molecular and atomic polarizabilities: thole’s model revisited. *J. Phys. Chem. A*, 102(14):2399–2407, 1998.
- [290] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. MacKerell. Charmm general force field (cgenff): A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.

- [291] C. Vega and J. L. F. Abascal. Simulating water with rigid non-polarizable models: a general perspective. *Physical Chemistry Chemical Physics*, 13(44):19663–19688, 2011.
- [292] H. Venthur, A. Mutis, J. J. Zhou, and A. Quiroz. Ligand binding and homology modelling of insect odorant-binding proteins. *Physiological Entomology*, 39(3):183–198, 2014.
- [293] Verhoeve.J and A. Dymanus. Magnetic properties and molecular quadrupole tensor of water molecule by beam-maser zeeman spectroscopy. *Journal of Chemical Physics*, 52(6):3222–, 1970.
- [294] N. K. Vyas, M. N. Vyas, and F. A. Quioco. Crystal structure of m tuberculosis abc phosphate transport receptor: specificity and charge compensation dominated by ion-dipole interactions. *Structure*, 11(7):765–74, 2003.
- [295] W. Wagner and A. Pruss. The iapws formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *Journal of Physical and Chemical Reference Data*, 31(2):387–535, 2002.
- [296] M. Waldman and A. T. Hagler. New combining rules for rare-gas van-der-waals parameters. *Journal of Computational Chemistry*, 14(9):1077–1084, 1993.

- [297] T. T. Waldron, M. A. Modestou, and K. P. Murphy. Anion binding to a protein-protein complex lacks dependence on net charge. *Protein Science*, 12(4):871–874, 2003.
- [298] S. Walker, R. Oun, F. J. McInnes, and N. J. Wheate. The potential of cucurbit n urils in drug delivery. *Israel Journal of Chemistry*, 51(5-6):616–624, 2011.
- [299] Tiffany R. Walsh and Marc R. Knecht. Biointerface structural effects on the properties and applications of bioinspired peptide-based nanomaterials. *Chemical Reviews*, 117(20):12641–12704, 2017.
- [300] Junmei Wang, Piotr Cieplak, Jie Li, Qin Cai, Meng-Juei Hsieh, Ray Luo, and Yong Duan. Development of polarizable models for molecular mechanical calculations. 4. van der waals parametrization. *The Journal of Physical Chemistry B*, 116(24):7088–7101, 2012.
- [301] L.-P. Wang, J. Chen, and T. van Voorhis. Systematic parametrization of polarizable force fields from quantum chemistry data. *Journal of Chemical Theory and Computation*, 9:452–460, 2013.
- [302] L. P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande. Systematic improvement of a classical molecular model of water. *Journal of Physical Chemistry B*, 117(34):9956–9972, 2013.

- [303] Lee-Ping Wang, Teresa Head-Gordon, Jay W. Ponder, Pengyu Ren, John D. Chodera, Peter K. Eastman, Todd J. Martinez, and Vijay S. Pande. Systematic improvement of a classical molecular model of water. *The Journal of Physical Chemistry B*, 117(34):9956–9972, 2013.
- [304] Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J. P. Piquemal, and P. Ren. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *J Chem Theory Comput*, 11(6):2609–2618, 2015.
- [305] Q. T. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J. P. Piquemal, and P. Y. Ren. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *Journal of Chemical Theory and Computation*, 11(6):2609–2618, 2015.
- [306] Qiantao Wang, Joshua A. Rackers, Chenfeng He, Rui Qi, Christophe Narth, Louis Lagardere, Nohad Gresh, Jay W. Ponder, Jean-Philip Piquemal, and Pengyu Ren. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *J. Chem. Theory Comput.*, 11(6):2609–2618, 2015.
- [307] Z Wang, A Choudhary, P S Ledvina, and F A Quioco. Fine tuning the specificity of the periplasmic phosphate transport receptor. site-directed mutagenesis, ligand binding, and crystallographic studies. *Journal of Biological Chemistry*, 269(40):25091–4, 1994.

- [308] Z. M. Wang, H. Luecke, N. H. Yao, and F. A. Quioco. A low energy short hydrogen bond in very high resolution structures of protein receptor phosphate complexes. *Nature Structural Biology*, 4(7):519–522, 1997.
- [309] Arieh Warshel and Arno Papazyan. Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogenbonds. *Proceedings of the National Academy of Sciences*, 93(24):13665–13670, 1996.
- [310] J. Wereszczynski and J. A. McCammon. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Quarterly Reviews of Biophysics*, 45(1):1–25, 2012.
- [311] F. H. Westheimer. Why nature chose phosphates. *Science*, 235(4793):1173–1178, 1987.
- [312] K. T. Wikfeldt, E. R. Batista, F. D. Vila, and H. Jonsson. A transferable h<sub>2</sub>o interaction potential based on a single center multipole expansion: Scme. *Physical Chemistry Chemical Physics*, 15(39):16542–16556, 2013.
- [313] D. E. Williams. Representation of the molecular electrostatic potential by atomic multipole and bond dipole models. *Journal of Computational Chemistry*, 9(7):745–763, 1988.
- [314] Hao Wu, Fabian Paul, Christoph Wehmeyer, and Frank No. Multiensemble markov models of molecular thermodynamics and kinetics.

*Proceedings of the National Academy of Sciences*, 2016.

- [315] J. C. Wu, G. Chattree, and P. Y. Ren. Automation of amoeba polarizable force field parameterization for small molecules. *Theoretical Chemistry Accounts*, 131(3), 2012.
- [316] Johnny C. Wu, Gaurav Chattree, and Pengyu Ren. Automation of amoeba polarizable force field parameterization for small molecules. *Theoretical Chemistry Accounts*, 131(3):1138, 2012.
- [317] P. Wu, L. G. Ma, X. L. Hou, M. Y. Wang, Y. R. Wu, F. Y. Liu, and X. W. Deng. Phosphate starvation triggers distinct alterations of genome expression in arabidopsis roots and leaves. *Plant Physiology*, 132(3):1260–1271, 2003.
- [318] Q. Wu and W. T. Yang. Empirical correction to density functional theory for van der waals interactions. *Journal of Chemical Physics*, 116(2):515–524, 2002.
- [319] Xiaojing Wu, Carine Clavaguera, Louis Lagardre, Jean-Philip Piquemal, and Aurlien de la Lande. Amoeba polarizable force field parameters of the heme cofactor in its ferrous and ferric forms. *Journal of Chemical Theory and Computation*, 14(5):2705–2720, 2018.
- [320] Matthew A. Wyczalkowski, Andreas Vitalis, and Rohit V. Pappu. New estimators for calculating solvation entropy and enthalpy and compara-

- tive assessments of their accuracy and precision. *The Journal of Physical Chemistry B*, 114(24):8166–8180, 2010.
- [321] S. S. Xantheas and E. Apra. The binding energies of the d2d and s4 water octamer isomers: High-level electronic structure and empirical potential results. *Journal of Chemical Physics*, 120:823–828, 2004.
- [322] S. S. Xantheas, C. J. Burnham, and R. J. Harrison. Development of transferable interaction models for water. ii. accurate energetics of the first few water clusters from first principles. *Journal of Chemical Physics*, 116(4):1493–1499, 2002.
- [323] Miaoren Xia, Zhifang Chai, and Dongqi Wang. Polarizable and non-polarizable force field representations of ferric cation and validations. *The Journal of Physical Chemistry B*, 121(23):5718–5729, 2017.
- [324] Jin Yu Xiang and Jay W. Ponder. A valence bond model for aqueous cu(ii) and zn(ii) ions in the amoeba polarizable force field. *Journal of Computational Chemistry*, 34(9):739–749, 2013.
- [325] W. Yang, Y. Q. Gao, Q. Cui, J. Ma, and M. Karplus. The missing link between thermodynamics and structure in f<sub>1</sub>/sub<sub>2</sub>-atpase. *Proceedings of the National Academy of Sciences*, 100(3):874–879, 2003.
- [326] Zhong-Zhi Yang, JianJiang Wang, and Dong-Xia Zhao. *Valence state parameters of all transition metal atoms in metalloproteins - Development of ABEEM fluctuating charge force field*, volume 35. 2014.

- [327] S. M. Yannone, S. Hartung, A. L. Menon, M. W. Adams, and J. A. Tainer. Metals in biology: defining metalloproteomes. *Curr Opin Biotechnol*, 23(1):89–95, 2012.
- [328] Nanhua Yao, Polly S. Ledvina, Abha Choudhary, and Florante A. Quioco. Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. *Biochemistry*, 35(7):2079–2085, 1996.
- [329] I. C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *Journal of Physical Chemistry B*, 108(40):15873–15879, 2004.
- [330] S. Yoo and S. S. Xantheas. Communication: The effect of dispersion corrections on the melting temperature of liquid water. *Journal of Chemical Physics*, 134(12), 2011.
- [331] W. B. Yu, P. E. M. Lopes, B. Roux, and A. D. MacKerell. Six-site polarizable model of water based on the classical drude oscillator. *Journal of Chemical Physics*, 138(3), 2013.
- [332] Qiao Zeng and WanZhen Liang. Analytic energy gradient of excited electronic state within tddft/mmpol framework: Benchmark tests and parallel implementation. *The Journal of Chemical Physics*, 143(13):134104, 2015.



- [333] X. Zeng, J. Li, H. Xie, and L. Liu. A novel dismantling process of waste printed circuit boards using water-soluble ionic liquid. *Chemosphere*, 93(7):1288–94, 2013.
- [334] C. S. Zhang, C. Lu, Q. T. Wang, J. W. Ponder, and P. Y. Ren. Polarizable multipole-based force field for dimethyl and trimethyl phosphate. *Journal of Chemical Theory and Computation*, 11(11):5326–5339, 2015.
- [335] Changsheng Zhang, Chao Lu, Zhifeng Jing, Chuanjie Wu, Jean-Philip Piquemal, Jay W. Ponder, and Pengyu Ren. Amoeba polarizable atomic multipole force field for nucleic acids. *Journal of Chemical Theory and Computation*, 14(4):2084–2108, 2018.
- [336] Lianqing Zheng, Menggen Chen, and Wei Yang. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51):20227–20232, 2008.
- [337] Lianqing Zheng, Menggen Chen, and Wei Yang. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *Journal of Chemical Physics*, 130(23), 2009.
- [338] Lianqing Zheng and Wei Yang. Practically efficient and robust free energy calculations: Double-integration orthogonal space tempering. *Journal of Chemical Theory and Computation*, 8(3):810–823, 2012.