



Evaluating and Improving the Description of London Dispersion Interactions in Molecular Mechanical Force Fields Using the Exchange-Hole Dipole Moment Model

by

© Mohamad Mohebifar

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirements for the
degree of Master of Science.

Department of Chemistry
Memorial University of Newfoundland

September 2018

St. John's, Newfoundland and Labrador, Canada

Abstract

Molecular simulations are used extensively to model processes in biophysics and biochemistry. These methods approximate the intramolecular and intermolecular interactions of the molecules in the system with a set of simplified mathematical expressions. London dispersion forces account for a significant portion of intermolecular interactions. These interactions play an important role in condensed matter physics and many biophysical phenomena. In this thesis, the eXchange-hole Dipole Moment model (XDM) of density functional theory was used to evaluate the dispersion coefficients in popular molecular mechanical models that are often used for simulations of water, organic molecules, and proteins. The dispersion coefficients derived from XDM calculations were compared to those extracted from molecular mechanical models with parameters from the GAFF, CGenFF, and OPLS force fields. For the generalized force fields, 88 organic molecules were evaluated. The Amber ff14sb, OPLS-AA, and CHARMM36 protein force fields were also evaluated using side chains models. Generally, the force field molecular C_6 dispersion coefficients overestimate the XDM C_6 dispersion coefficients by 50–60%. Despite this, these models predict the solvation energies of these molecules correctly. This trend was attributed to the neglect of higher order dispersion terms. In the empirical parameterization of these force fields, the interaction energy that should arise from these higher order terms will be spuriously added to the C_6 term. In the final chapter, a water model was developed with an improved non-bonded potential that describes repulsive forces more accurately using an exponential Buckingham-type term and includes C_6 and C_8 dispersion terms. High-performance GPU-CUDA and vectorized expressions for this potential were implemented in OpenMM. The model is able to predict the structural, physical, and transport properties of liquid water accurately.

Acknowledgements

I would like to thank my supervisor Dr. Christopher Rowley for the constant support and patient guidance throughout my time as a student in his group, and Evan Walters for the much needed contribution on the evaluation of protein force fields. I would also like to thank NSERC of Canada for funding through the Discovery Grants program (Application 418505-2012 and RGPIN-05795-2016), the School of Graduate Studies at Memorial University for a graduate fellowship, and Dr. Liqin Chen for a scholarship. Computational resources were provided by Compute Canada (RAPI: djc-615-ab). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

Statement of Contribution

For Chapter 2, I wrote the scripts for performing QM-based geometry optimizations, running Postg to determine coefficients using the XDM model and parsing the output files, extracting force field dispersion coefficients for the test molecules. Also, I made a python program that automates the grid scan of force field parameters. In conjunction with Drs. Rowley and Johnson, I prepared this work for publication in the Journal of Computational and Theoretical Chemistry.

For Chapter 3, I wrote the scripts to read CHARMM, Amber, and OpenMM parameter files as well as different topology file formats to extract force field dispersion coefficients for side chain residues and nucleobases. I provided advice and training on how to perform these calculations to an Honours student, Evan Walters, who performed the QM and TI/FEP calculations in this chapter. In conjunction with Dr. Rowley, Dr. Johnson, and Mr. Walters, I prepared this work for publication in the Journal of Physical Chemistry B.

For Chapter 4, I implemented the Buckingham potential code using C++ for both CPU (using AVX) and GPU (using CUDA), and ported it into OpenMM. I also performed water model parameterization using the Force Balance program. Also, I wrote scripts to post-process MD simulations to study temperature and pressure dependence of liquid properties.

Dr. Erin R. Johnson of Dalhousie University provided guidance on the XDM model, instructions of the use of the postg program, and assistance in preparing the published manuscripts.

In addition, I wrote and published PyQueue, an open-source python library to submit batch jobs on Compute Canada clusters.

Table of Contents

Title page	i
Abstract	ii
Acknowledgements	iii
Statement of Contribution	iv
Table of Contents	v
List of Tables	viii
List of Figures	x
List of Symbols	xii
List of Abbreviations	xiii
1 Introduction	1
1.1 Non-bonded Interactions	1
1.1.1 London Dispersion Interactions	2
1.1.2 Mathematical Description of London Dispersion Interactions	2
1.2 Molecular Mechanics	4
1.3 Molecular Dynamics	5
1.3.1 Periodic Boundary Conditions	6
1.4 Determination of Dispersion Coefficients	7
1.4.1 Molecular Mechanical Force Fields	7
1.4.2 Exchange-Hole Dipole Moment	7
1.5 The Significance of Dispersion Interactions in Biophysics	8

1.6	Outline	12
2	Evaluating the Dispersion Coefficients of General Force Fields for Organic Molecules Using XDM	13
2.1	Abstract	13
2.2	Introduction	15
2.3	Computational Methods	17
2.3.1	XDM Calculations	17
2.3.2	Molecular Dynamics Simulations	17
2.4	Results and Discussion	18
2.4.1	Molecular Dispersion Coefficients	18
2.4.2	Atomic Dispersion Coefficients	22
2.4.3	The Drude Force Field	31
2.4.4	Water Models	34
2.4.5	Force Field Development using XDM-Derived Dispersion Coefficients	35
2.5	Conclusions	38
3	Evaluating the Dispersion Coefficients of General Force Fields for Proteins Using XDM	41
3.1	Abstract	41
3.2	Introduction	42
3.3	Computational Methods	43
3.3.1	Side-Chain Models	43
3.3.2	XDM Calculations	44
3.3.3	Free Energy Simulations	44
3.3.4	Molecular Dynamics Calculations	45
3.3.5	Ab Initio Molecular Dynamics	46
3.4	Results and Discussion	47
3.4.1	Molecular Dispersion Coefficients of Amino Acid Side-Chains	47
3.4.2	Atomic Dispersion Coefficients	49
3.4.3	Side-Chain Hydration Energies	52
3.4.4	Nitrogen-Containing Residues	60
3.4.5	Improving the Description of Dispersion in Force Fields	60
3.4.6	Evaluation of Nucleic Acid Force Fields	63

3.5	Conclusions	65
4	Development of a Water Model with an Improved Non-bonded Potential	67
4.1	Abstract	67
4.2	Introduction	68
4.3	Computational Methods	72
4.4	Results and Discussion	73
4.5	Conclusions	76
5	Conclusions and Future Work	77
5.1	Conclusions	77
5.2	Future Work	78
	Bibliography	80
A	Conversion of Dispersion Coefficients	99
B	Tables of Side Chain Hydration Energies	101

List of Tables

2.1	Calculated liquid properties obtained with force fields for which the molecular C_6 coefficients differ from the XDM values by a wide margin.	21
2.2	Average values of the homoatomic C_6 dispersion coefficients, grouped by element, for the compounds in the Virtual Chemistry test set	23
2.3	Homoatomic dispersion coefficients for Drude atom types where the force-field C_6 is much larger than the corresponding XDM value. . . .	33
2.4	Molecular C_6 coefficients for various molecular-mechanical water models that include a Lennard-Jones 12-6 potential to represent dispersion.	35
2.5	Physical properties of 5 different liquids calculated using the original CGenFF force field and a reparameterized force field (FF-XDM). . . .	37
2.6	Parameterization space for Lennard-Jones parameters of dispersion-bound liquids	37
3.1	Dispersion component of side-chain hydration energies extrapolated from XDM and those calculated for each force field using REMD-TI . .	57
3.2	Sulfur-oxygen radii calculated from the methylthiol radial distribution functions, the force field Lennard-Jones parameters, and the dispersion component of the methylthiol solvation energy.	59
3.3	The dispersion component of the hydration energy of methane calculated using the free-energy simulations using the XDM C_6 , C_8 , and C_{10} coefficients	61
4.1	Optimal parameters for a 4-point water model with the Buckingham 6-8 potential.	73
4.2	Properties of liquid water predicted by the optimized B68 water model.	74

B.1	Hydration energies of side-chain models calculated using the CHARMM36 force field.	101
B.2	Hydration energies of side-chain models calculated using the Amber ff14sb force field.	102
B.3	Hydration energies of side-chain models calculated using the OPLS-AA force field	102

List of Figures

1.1	Examples of biophysical processes that involve changes in dispersion interactions.	9
1.2	Dispersion coefficients of water models.	11
2.1	Correlation between the molecular dispersion coefficients of the three force fields and XDM	19
2.2	Correlation between the homoatomic dispersion coefficients of the CGenFF, GAFF, and OPLS force fields and XDM.	24
2.3	Correlation between the homoatomic dispersion coefficients of the CGenFF, GAFF, and OPLS force fields and XDM for H, C, N, O, and F atoms .	25
2.4	The atomic XDM dispersion coefficients for electron-rich carbon atoms	26
2.5	Examples of XDM dispersion coefficients for methyl-group carbon atoms.	27
2.6	Examples of XDM dispersion coefficients for carbon atoms in aromatic systems.	28
2.7	Examples of XDM dispersion coefficients for an electron-poor amide nitrogen and an electron-rich amine nitrogen	28
2.8	Correlation between the homoatomic and molecular dispersion coefficients calculated using the Drude force field and XDM.	32
3.1	Structures of compounds used to model amino acid side chains.	44
3.2	The correlation between molecular C_6 coefficients of the protein side chain models.	48
3.3	The atomic C_6 coefficients of the amino acid side chain models.	50
3.4	Correlation between experimental and calculated hydration energies of the side-chain models for the neutral amino acids.	53

3.5	Correlation between the force field C_6 molecular dispersion coefficients and the dispersion component of the hydration energy calculated from the free-energy simulations and the correlation between the XDM C_6 molecular dispersion coefficients and the component of the hydration energy calculated using the free-energy simulations.	55
3.6	Components of the hydration energy of the Cys and Met side chain models.	58
3.7	S–O radial distribution functions for methyl thiol in an aqueous solution.	59
3.8	Molecular dispersion coefficients for nucleobases calculated using XDM and the CHARMM36 and Amber ff14sb force fields.	64
4.1	Radial distribution functions of TIP3P, TIP4P-FB, and B68 water models.	74
4.2	Temperature dependence of liquid water properties calculated using the TIP3P, TIP4P-FB, and B68 water models compared to the experimental values.	75

List of Symbols

k_B	Boltzmann's constant
R	gas constant
C_6	6th-order London dispersion coefficient
σ	Lennard-Jones radius
ϵ	Lennard-Jones well-depth
\mathcal{V}	potential energy
ρ	density
ϵ_0	dielectric constant
q	atomic charge
α	thermal expansion coefficient
κ	isothermal compressibility coefficient
C_p	heat capacity
ΔH_{vap}	enthalpy of vaporization
V	volume
D	diffusion coefficient
η	viscosity

List of Abbreviations

a.u.	atomic units
AIMD	Ab Initio Molecular Dynamics
AMBER	Assisted Model Building with Energy Refinement
B68	Buckingham-type potential with 6th and 8th order dispersion
CGenFF	CHARMM generalized force field
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CPU	Central processing unit
CUDA	Compute Unified Device Architecture
DFT	density functional theory
GAFF	Generalized AMBER force field
GPU	Graphical processing unit
GROMACS	GRONingen MACHine for Chemical Simulations
GROMOS	GRONingen MOlecular Simulation
LJ	Lennard-Jones potential
MD	Molecular Dynamics
MM	Molecular Mechanics
NAMD	Nanoscale Molecular Dynamics
OPLS	Optimized Potentials for Liquid Simulations
OPLS-AA	OPLS All-atom force field
PME	Particle Mesh Ewald
QM	Quantum Mechanics
RDF	radial distribution function
TIP3P	three-site transferable intermolecular potential
TIP3P-FB	three-site transferable intermolecular potential, ForceBalance parameterization
TIP4P-FB	four-site transferable intermolecular potential, ForceBalance parameterization
XDM	eXchange-hole Dipole Moment

Chapter 1

Introduction

Parts of this chapter have been published in the following articles:

- Walters, E., Mohebifar, M., Johnson, E.R., Rowley, C. N., Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-Hole Dipole Moment Model, *J. Phys. Chem. B*, **2018**, 122 (26), 6690–6701 doi: 10.1021/acs.jpcc.8b02814
- Mohebifar, M., Johnson, E. R., Rowley, C. N., Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model, *J. Chem. Theory Comput.*, **2017** 13 (12), 6146–6157 doi: 10.1021/acs.jctc.7b00522

1.1 Non-bonded Interactions

Non-bonded forces are a group of distance-dependent forces acting between atoms, either of the same or different molecules. These interactions all originate from complex electron–electron, proton–proton, and electron–proton Coulombic interactions,

so they can be effectively simplified into pairwise electrostatic, Pauli repulsion interactions, and London dispersion. Pairwise electrostatic interactions arise from unequally distributed charges in a molecule, which are typically approximated atom-centered partial charges. Pauli repulsion forces act at very short distances due to the Pauli exclusion principle. London dispersion is an attractive force between atoms and molecules that results from fluctuations in the electron density that create instantaneous electric moments. These electric moments induce electric moments in neighboring atoms, resulting in an attractive Coulombic interaction. This thesis explores how London dispersion interactions are currently represented in computer simulations and how this representation can be improved.

1.1.1 London Dispersion Interactions

Although the dispersion interaction between a pair of atoms is generally weak and short range, these interactions amount to a significant cohesive force in condensed phases. They are notably significant in determining the structure and function of biomolecules like proteins, nucleic acids, and lipids [1; 2]. Dispersion is also an important factor in the stability of crystal phases and interfacial complexes [3; 4; 5]. Atomistic simulations of these materials require a molecular-mechanical force field that provides a realistic description of dispersion interactions.

1.1.2 Mathematical Description of London Dispersion Interactions

The potential energy of London dispersion interactions can be well-approximated by

$$\mathcal{V}_{\text{disp},ij}(r_{ij}) = - \sum_{n=6,8,10\dots} \frac{C_{n,ij}}{r_{ij}^n}, \quad (1.1)$$

where r_{ij} is the distance between atomic pairs and the $C_{n,ij}$ are coefficients that depend on the identity of the interacting pair [6]. In general, terms of order greater than 10 are negligible due to their rapid decay as r increases [6], yielding

$$\mathcal{V}_{disp}(r_{ij}) = -\frac{C_{6,ij}}{r_{ij}^6} - \frac{C_{8,ij}}{r_{ij}^8} - \frac{C_{10,ij}}{r_{ij}^{10}} \quad (1.2)$$

where $C_{6,ij}$, $C_{8,ij}$, and $C_{10,ij}$ are the dispersion coefficients.

The Lennard-Jones potential also uses Eqn. 1.2 truncated at the C_6 term [7],

$$\mathcal{V}_{LJ,ij}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{6,ij}}{r_{ij}^6}. \quad (1.3)$$

The A/r^{12} term of this equation is intended to represent Pauli repulsion. Dispersion interactions are represented only by the C_6/r^6 term. More commonly, this potential is defined in terms of the atomic radius (σ) and well-depth (ϵ),

$$\mathcal{V}_{LJ,ij}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (1.4)$$

where $C_{6,ij} = 4\epsilon_{ij}\sigma_{ij}^6$.

In principle, parameters can be defined for each unique pair of atoms [8; 9; 10], but the most common practice is to define parameters for like-pairs and calculate the parameters for unlike pairs using the Lorentz–Berthelot combination rule [11],

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}. \quad (1.5)$$

A different expression for non-bonded interactions was proposed by Buckingham to describe the interactions of noble gases [12],

$$\mathcal{V}_{nb}(r) = A_{ij} \cdot \exp(-b_{ij} \cdot r_{ij}) - \frac{C_{6,ij}}{r_{ij}^6} - \frac{C_{8,ij}}{r_{ij}^8}. \quad (1.6)$$

Here, the repulsive interaction is described by the $A \cdot \exp(-b \cdot r)$ term, where the A and b parameters define the strength of the repulsion. Dispersion interactions are represented by the C_6/r^6 and C_8/r^8 terms. This potential has several advantages over the Lennard-Jones potential. Pauli repulsion originates from the overlap of atomic electron clouds at close range [13]. As the electron density of atoms follows exponential decay, Pauli repulsion interactions can also be described in such a way. The exponential term is a more realistic description of Pauli repulsion than the polynomial A/r^{12} term, which is advantageous in simulations where strong attractive interactions, high temperatures, or high pressures result in frequent repulsive contacts.

1.2 Molecular Mechanics

Molecular mechanics (MM) is a computational method to compute the potential energy surface of a molecular system using Newtonian physics. The functional form of the equation of energy along with the parameter set is referred to as a force field. These force fields can be classified into two different groups: all-atoms and united atom force fields. All-atom force fields treat each atom as a particle, whereas united-atom force fields consider a group of atoms as one interaction center. The reduced number of degrees of freedom and interactions in coarse grain models reduces the computational cost of the calculations. When all the degrees of freedom are used in a study, all-atom force fields should be used to maximize the accuracy of results. Force fields typically represent both intramolecular interactions (i.e., bond, angle, and dihedral energy contributions) and intermolecular interactions (i.e., electrostatic, London dispersion, and repulsive interactions) [14].

In these models, covalent bond stretching is usually represented by a spring-like harmonic oscillator potential. The bond stretches and the angle bends are represented using harmonic potentials. The electrostatic interactions are computed using the Coulomb term, and the van der Waals interactions are calculated with a Lennard-Jones potential [14].

Typical force fields describe the energy of systems by Eqn. 1.7.

$$\begin{aligned} \mathcal{V}_{\text{total}} = & \sum_{\text{bond}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedral}} K_\chi [1 + \cos(n\chi - \tau)] \\ & + \sum_{\text{non-bonded pairs}} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon r_{ij}} \right) \end{aligned} \quad (1.7)$$

In this equation, K_b and r_0 are the force constant and equilibrium bond length, respectively. K_θ and θ_0 are the equilibrium bond angle and bond angle constant, respectively. K_χ , τ , and $n\chi$ are also the amplitude, shift, and periodicity of the dihedral potentials. ϵ_{ij} is the Lennard-Jones potential well depth and σ_{ij} is the sum of atomic radii. Also, q_i and q_j are the point charges located at the center of atoms i and j .

1.3 Molecular Dynamics

Molecular dynamics (MD) is a powerful tool used to simulate the dynamics of a molecular system using Newtonian physics. Molecular dynamics is usually used for inherently disordered systems, such as liquids, in order to sample many possible configurations. In this technique, the equations of motion of particles are solved numerically using high-performance computers [15].

There are many numerical methods to integrate the equations of motion. Many

simple but effective integrators can be derived from a second-order Taylor series expansion of the positions of the particles at a small time step in the future, $t + \Delta t$, using the positions, velocities, and forces at the current time step, t [16; 17].

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2} \frac{\vec{F}_i(t)}{m_i} \Delta t^2 \quad (1.8)$$

Here, F_i and m_i are the force imposed on particle i and its mass, respectively. Given the positions \vec{x} and velocities \vec{v} at the current time t , we can calculate the positions and velocities for some later time $t + \Delta t$. The time step Δt must be a very small value (e.g., 1 fs) to minimize the numerical error.

Usually, molecular dynamics simulations are performed at a constant temperature. Integrators like the Verlet algorithm can be coupled to a thermostat like the Nosé–Hoover thermostat [18; 19; 20]. To sample a constant pressure ensemble, the dynamics can also be coupled to a barostat, such as the Nosé–Hoover barostat [21].

1.3.1 Periodic Boundary Conditions

Molecular simulations are often used to predict the properties of bulk liquids; however, the number of particles associated with even a small droplet of a liquid would be prohibitively large to simulate. Instead, a periodic simulation cell is commonly used, which mimics the bulk properties of a liquid while restricting the number of particles to a limited number (e.g., 1,000–100,000 particles). By cutting off long-range non-bonded interactions and using lattice-summation methods for the long range component, the computational complexity of the pairwise interactions scales as $O(n \log(n))$ [22], which is tractable up to significant scales on modern computers.

1.4 Determination of Dispersion Coefficients

1.4.1 Molecular Mechanical Force Fields

The Lennard-Jones parameters used in molecular mechanical force fields are typically assigned empirically such that simulations using these parameters accurately predict the properties of representative bulk liquids (e.g., density, enthalpy of vaporization, dielectric constant, etc.) [10; 23; 24; 25; 26; 27; 28], although some quantum mechanics(QM)-based approaches have also been applied [29; 30; 31; 32; 33]. A drawback of parameterizing Lennard-Jones parameters empirically is that the parameterization space has two dimensions per atom type (i.e., ϵ and σ), so this empirical parameterization procedure may not result in a unique, physically-realistic set of parameters. For example, the Assisted Model Building with Energy Refinement ff14sb (Amber ff14sb) force field uses Lennard-Jones parameters for the thiol sulfur atom type of $\sigma = 3.64 \text{ \AA}$ and $\epsilon = 1.046 \text{ kJ/mol}$, yielding a dispersion coefficient of 149 a.u. [34]. For the same atom type, the Chemistry at Harvard Macromolecular Mechanics (CHARMM36) force field assigns parameters of $\sigma = 3.52 \text{ \AA}$ and $\epsilon = 1.59 \text{ kJ/mol}$, yielding a dispersion coefficient of 267 a.u. Further, because the underlying force field is approximate, the dispersion parameters can be assigned unphysical values in order to compensate for other issues. In the past, there was no general method to validate the parameters that come out of this process, so they are commonly treated as purely empirical, unrestrained parameters.

1.4.2 Exchange-Hole Dipole Moment

The exchange-hole dipole moment (XDM) model provides a method for calculating atomic and molecular dispersion coefficients from a simple density-functional theory (DFT) calculation [35; 36; 37]. XDM is based on second-order perturbation theory

[38; 39] and uses the multipole moments for a reference electron and its exchange hole as the source of the instantaneous moments that give rise to dispersion. It allows non-empirical calculation of atomic and molecular dispersion coefficients, to arbitrarily high order [40], based on properties of the DFT electron density. Because the atomic dispersion coefficients are dependent on the electron density via the exchange hole, they vary with local chemical environment [41] and implicitly include many-body electronic effects [42; 43] (or Type A and B non-additive effects in Dobson’s classification scheme [44]). In general, the computed molecular C_6 coefficients are accurate to within ca. 10% when compared to experimental reference data [45].

1.5 The Significance of Dispersion Interactions in Biophysics

Dispersion interactions are significant in many chemical and biophysical processes. The hydrophobic effect is most commonly discussed in terms of changes in electrostatic interactions and entropy, but changes in dispersion interactions can also play a significant, albeit secondary, role. Water is a polar, hydrogen-bonding liquid, but water molecules are only moderately polarizable, so the dispersion interaction between a molecule and an aqueous solution is relatively weak. Conversely, biomolecules like lipids, proteins, and nucleic acids are largely composed of more polarizable carbon atoms, which results in stronger dispersion interactions. As a result, there can be a large change in the strength of dispersion interactions when molecules transfer between different cellular environments or a biomolecule changes conformation such that its exposure to aqueous solution changes.

For example, a significant change in the strength of dispersion interactions can occur when a protein folds (Figure 1.1 A). The protein–solvent dispersion interactions

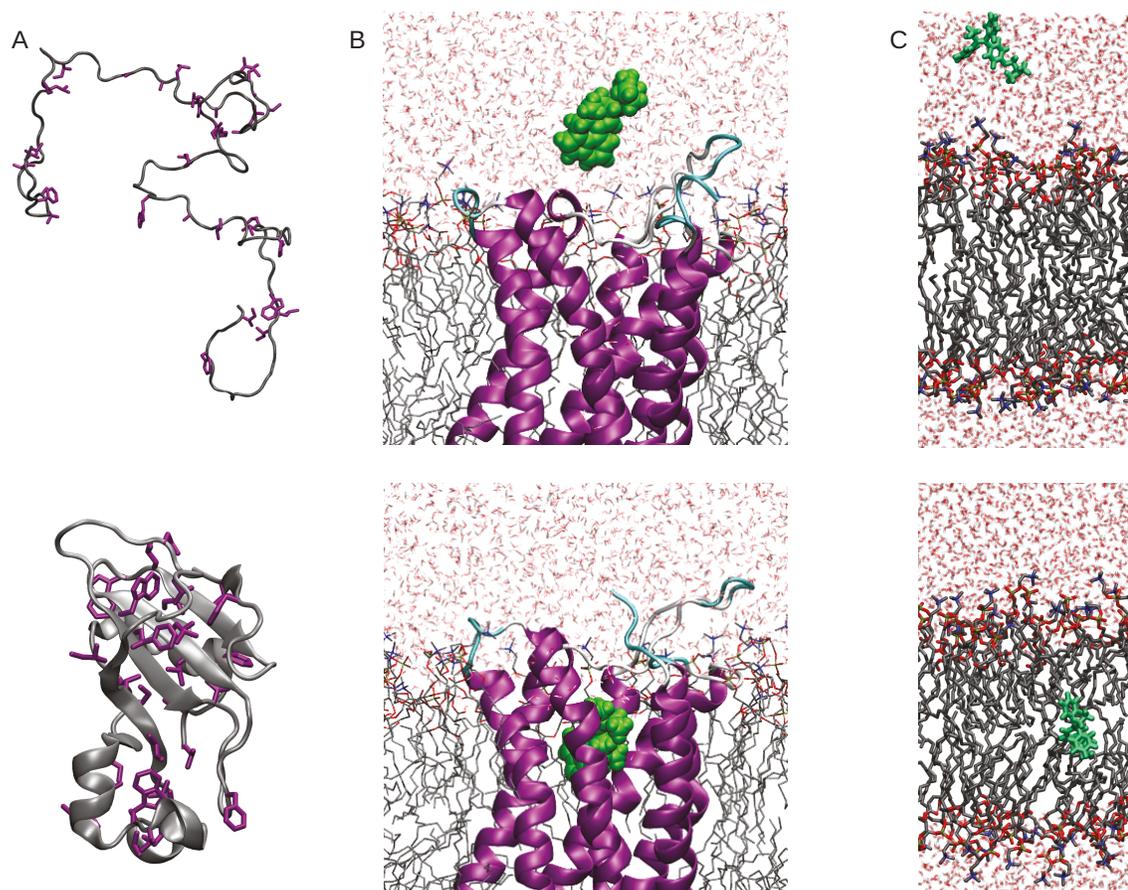


Figure 1.1: Examples of biophysical processes that involve changes in dispersion interactions. (a) Hydrocarbon side-chains (purple) of the protein barnase have dispersion interactions with water when unfolded, but form stronger dispersion interactions with each other in the folded state. (b) LSD (green) has stronger dispersion interactions with the amino acid residues of the binding site of a GPCR membrane protein (magenta) than with water molecules in solution. (c) Tamoxifen (green) has stronger dispersion interactions with the acyl chains of a lipid bilayer than with water molecules in solution.

decrease when a protein folds because a high proportion of side-chains are solvent-exposed in unfolded states, but many of these residues are buried in the interior of the protein in the folded states. Conversely, protein–protein dispersion interactions become much stronger when a protein folds, largely because aliphatic hydrophobic residues move into contact in the folded state. Early analysis of protein folding energetics by Honig and coworkers [46] using a continuum model for the solvent concluded that changes in dispersion interactions were not significant in protein folding, although more recent models have shown that more complete descriptions of dispersion interactions are needed to achieve quantitative accuracy [1; 47; 48; 49; 50].

Another example of a biophysical process that involves a large change in dispersion interactions is the permeation of molecules through lipid bilayers (Figure 1.1 B). Hydrophobic solutes will generally have stronger dispersion interactions in the membrane interior than they have with the aqueous solutions above and below the bilayer [51; 52]. For example, in a simulation of tamoxifen permeating through a POPC bilayer, the total dispersion interactions of tamoxifen become stronger when the molecule moves from solution to the center of the bilayer.

A final example where changes in dispersion interactions are significant is the binding of hydrophobic ligands to proteins (Figure 1.1 C). In the proteinaceous environment, the ligand will have stronger interactions with the polarizable carbon atoms of the protein than with the aqueous solution from which they originate. For example, the Lennard-Jones interactions of lysergic acid diethylamide (LSD) become stronger when this drug moves from solution into the binding site of a GPCR membrane protein. This is in keeping with the analysis of Roux and coworkers, who have shown that increases in dispersion interactions are often the largest component of the Gibbs energy of protein–ligand binding [53; 54].

The strength of the dispersion interactions in these simulations is sensitive to the

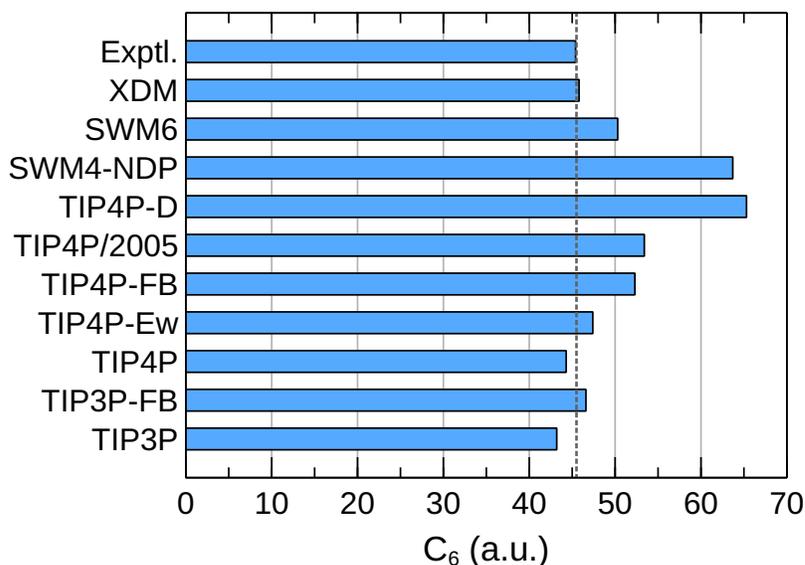


Figure 1.2: Dispersion coefficients of water models. The reference XDM value is indicated by the dotted line. Values are taken from Ref. 55. The experimental value is taken from Ref. 56.

parameters used in the force field. Figure 1.2 shows the broad range of C_6 coefficients used in popular water models that represent dispersion interactions through a single Lennard-Jones term centered on the oxygen atom. For example, the C_6 coefficient of the three-site transferable intermolecular potential (TIP3P) [57] model is 43.2 a.u., while the C_6 coefficient of the TIP4P/2005 [58] model is 53.4 a.u. The large variations in C_6 affect the strength of the dispersion interactions between water and the rest of the system. Stronger drug-water interactions reduce the degree to which dispersion facilitates the transfer of the drug from solution into the bilayer interior. The wide variation of dispersion coefficients in water models that have similar physical properties suggests the potential for inconsistency in these simulations.

Each of these processes involves the hydrophobic effect, where there are competing changes in dispersion interactions, electrostatic interactions, and entropy. For simulations to describe these processes rigorously, each component, including dispersion, must be represented accurately. In the most widely-used force fields, the only term

that accounts for dispersion interactions is the C_6/r^6 term of the Lennard-Jones potential, so this term and its C_6 parameters must describe the dispersion interactions in these systems realistically.

1.6 Outline

The research present in this thesis focuses on London dispersion interactions in molecular mechanical force fields for use in molecular simulations. Chapters 2 and 3 present the comparison of dispersion coefficients in molecular mechanical force fields with those calculated using the QM-based XDM model. Motivated by the results of Chapters 2 and 3, in Chapter 4, a new water force field is developed using a Buckingham-type non-bonded potential that includes higher-order dispersion interactions.

Chapter 2

Evaluating the Dispersion Coefficients of General Force Fields for Organic Molecules Using XDM

This chapter is adapted with permission of the publisher from the article:

Mohebifar, M., Johnson, E. R., Rowley, C. N., Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model, *J. Chem. Theory Comput.*, **2017**, *13* (12), 6146–6157 doi: 10.1021/acs.jctc.7b00522

2.1 Abstract

London dispersion interactions play an integral role in materials science and biophysics. Force fields for atomistic molecular simulations typically represent dispersion interactions by the 12-6 Lennard-Jones potential, using empirically-determined parameters. These parameters are generally underdetermined and there is no straightforward way to test if they are physically realistic. Alternatively, the exchange-hole

dipole moment (XDM) model from density-functional theory predicts atomic and molecular London dispersion coefficients from first principles, providing an innovative strategy to validate the dispersion terms of molecular-mechanical force fields. In this work, the XDM model was used to obtain the London dispersion coefficients of 88 organic molecules relevant to biochemistry and pharmaceutical chemistry, and the values are compared with those derived from the Lennard-Jones parameters of the CHARMM generalized force field (CGenFF), Generalized AMBER force field (GAFF), Optimized Potentials for Liquid Simulations (OPLS), and Drude polarizable force fields. The molecular dispersion coefficients for the CGenFF, GAFF, and OPLS models are systematically higher than the XDM-calculated values by a factor of roughly 1.5, likely due to neglect of higher-order dispersion terms and premature truncation of the dispersion-energy summation. The XDM dispersion coefficients span a large range for some molecular-mechanical atom types, suggesting an unrecognized source of error in force-field models, which assume that atoms of the same type have the same dispersion interactions. Agreement with the XDM dispersion coefficients is even poorer for the Drude polarizable force field. Popular water models were also examined and TIP3P was found to have dispersion coefficients similar to the experimental and XDM references, although other models employ anomalously high values. Finally, XDM-derived dispersion coefficients were used to parameterize molecular-mechanical force fields for five liquids – benzene, toluene, cyclohexane, *n*-pentane, and *n*-hexane – which resulted in improved accuracy in the computed enthalpies of vaporization despite only having to evaluate a much smaller section of the parameter space.

2.2 Introduction

Molecular dynamics simulations of organic molecules often employ a generalized molecular-mechanical force field. These force fields define parameters for the standard types of chemical bonds and functional groups present in organic molecules, making it possible to generate a force field automatically for an arbitrary molecule. CGenFF [59], GAFF [60], and OPLS [61] are popular generalized force fields. More recently, models that are capable of describing induced polarization, such as Drude polarizable force fields, have also been developed [62].

The CGenFF, GAFF, OPLS, and Drude force fields all account for inter-atomic London dispersion interactions via the Lennard-Jones 12-6 potential in terms of ϵ and σ . (Eqn. 1.4)

To reduce the number of parameters needed to define the force field, each atom in the system is assigned an atom type. All atoms of the same type are assumed to have the same Lennard-Jones parameters. The type of an atom is generally specified by its element, hybridization, and bonding partners. The number and definition of atom types vary widely between different force fields. For instance, for the molecules studied in this paper, the CGenFF, GAFF, and OPLS force fields have 67, 32, and 153 atom types, respectively.

The σ and ϵ parameters for molecular-mechanical force fields are typically assigned by performing simulations of bulk liquids using various parameter sets. For each set of parameters, properties like the density and enthalpy of vaporization of the neat liquid are calculated. The parameters that yield the most accurate properties are used as the standard Lennard-Jones potential for that atom type. While this practice has been effective, there are several associated drawbacks. This procedure assumes the Lennard-Jones parameters are transferable to atoms of the same type in other

molecules. Additionally, fitting parameters to properties of bulk liquids becomes more difficult for polyatomic molecules because the Lennard-Jones parameters for multiple atom types must be fit simultaneously. Both ϵ and σ are treated as free parameters of empirical force fields, along with hundreds of other parameters. This creates the possibility that the parameterization procedure will generate values for ϵ and σ that are not true representations of Pauli repulsion and 6th-order dispersion, but rather capture a broad range of intermolecular interactions in these terms.

The need for greater accuracy in molecular simulations has spurred efforts to validate force field parameters. The Virtual Chemistry database provides structures and topology files for simulations of molecular liquids with the CGenFF, GAFF, and OPLS force fields. This provides an extensive test set to evaluate the accuracy of the force-field parameters. Simulations of molecular liquids in this test set have shown that the computational predictions can be significantly in error for some properties, although it is not always apparent which parameter(s) require adjustment. The ability of XDM to compute atomic dispersion coefficients from first principles provides a novel way of determining if the C_6 dispersion coefficients of a force field are physically realistic. This can be used to highlight where the Lennard-Jones term of the force field could be serving to approximate other associative interactions rather than true London dispersion interactions.

This work presents the calculation of C_6 coefficients using the XDM model on 88 molecules from the Virtual Chemistry force-field test set. The calculated coefficients are compared to those derived from the Lennard-Jones parameters for the CGenFF, GAFF, OPLS, and Drude force fields. Revised force fields for liquid benzene, toluene, and cyclohexane are derived based on the XDM dispersion coefficients.

2.3 Computational Methods

2.3.1 XDM Calculations

XDM dispersion coefficients were calculated for a set of 88 molecules relevant to biochemistry and pharmaceutical chemistry. Gas-phase structures from the Virtual Chemistry database were taken as the initial geometries. These structures were energy minimized with the PBE0 functional [63] and the def2-SVP basis set [64] using ORCA 3.0 [65]. Further geometry optimization was then performed with PBE0/aug-cc-pVTZ using Gaussian 09 [66]. Single-point energy calculations with PBE0/aug-cc-pVTZ were carried out to generate the wavefunction files needed to determine the XDM dispersion coefficients. This method has been shown to be reliable for predicting the molecular electrostatic properties of small molecules [67]. The XDM dispersion coefficients were calculated from the PBE0/aug-cc-pVTZ wavefunction files using the postg program [36; 68; 69]. A Python script that automates the parsing of dispersion coefficients from the output of postg is available through GitHub [70]. Force field parameters for molecules in the Virtual Chemistry test set were extracted from the published itp files [71]. The equations for conversion of these parameters to a C_6 coefficient in atomic units are given in the appendix A.

2.3.2 Molecular Dynamics Simulations

Simulations to parameterize molecular-mechanical force fields were performed using GRONingen MAchine for Chemical Simulations (GROMACS) 5.1.4 [72]. The simulations were performed under periodic boundary conditions with unit cells containing 1000 molecules. Where possible, initial coordinates were taken from the Virtual Chemistry database. In the remaining cases, initial coordinates were generated using the GROMACS insert-molecules module. A Parrinello–Rahman barostat [73; 73]

and Nosé–Hoover thermostat [19; 74] were used in order to sample the isothermal-isobaric ensembles. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) with a grid spacing of 1 Å. Lennard-Jones interactions were calculated using the lattice-sum method [75]. Properties were calculated from a 1 ns simulation to equilibrate the system followed by a 1 ns simulation to sample the properties.

2.4 Results and Discussion

2.4.1 Molecular Dispersion Coefficients

The molecular C_6 dispersion coefficients for the molecules in the test set were calculated using XDM and the force-field parameters. The correlations between the molecular XDM dispersion coefficients and the molecular dispersion coefficients for the CGenFF, GAFF, and OPLS force fields are plotted in Figure 2.1. There is a systematic trend for the force fields to overestimate the molecular dispersion coefficients, with regression coefficients of 1.53, 1.59, and 1.55 for the CGenFF, GAFF, and OPLS force fields, respectively. This suggests that the dispersion interactions in molecular liquids will be overestimated by these molecular-mechanical force fields.

The overestimation of dispersion coefficients in these force fields may be the result of the neglect of some components of the intermolecular interactions. Because the C_6 coefficients are parameterized to empirical liquid properties, the C_6 coefficients will be assigned spuriously large values to compensate for these neglected intermolecular interactions. For instance, the generalized force-field models use fixed atomic charges to represent electrostatic interactions. This neglects electrostatic interactions due to induced polarization, so the C_6 coefficients of these force fields may have been overestimated to compensate for the underestimation of electrostatic interactions. The development of polarizable molecular-mechanical models is one route to address

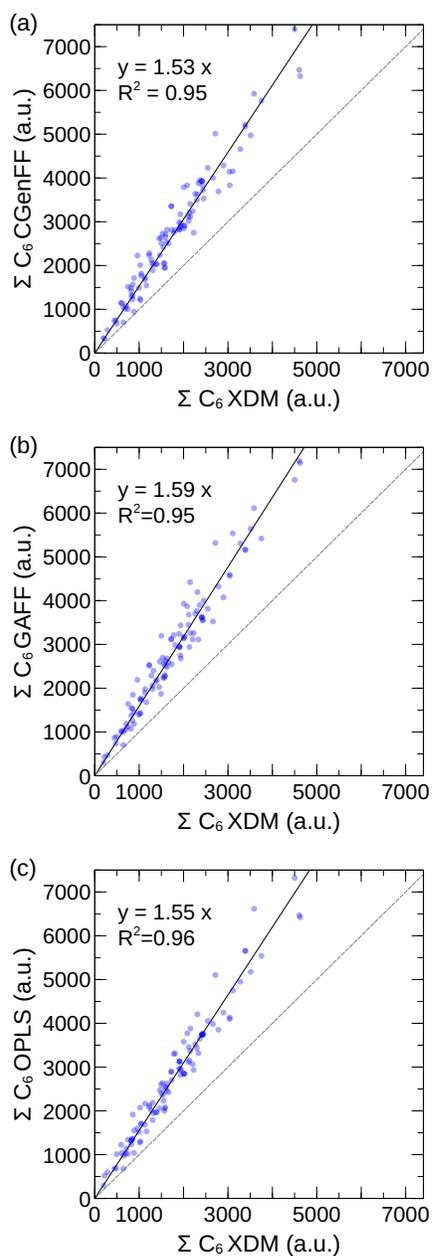


Figure 2.1: Correlation between the molecular dispersion coefficients of the (a) CGenFF, (b) GAFF, and (c) OPLS force fields and XDM. Each point represents a single molecule of the Virtual Chemistry test set.

these issues.

Stone attributed the tendency for force field C_6 coefficients to be larger than

physically-motivated values to the neglect of 8th- and 10th-order dispersion interactions in the Lennard-Jones interaction potential [31]. This is consistent with the results of quantum chemical calculations, which showed that the C_8 dispersion term accounts for ca. 30% of the dispersion energy in both molecular dimers [76] and in molecular crystals [77]. A moderate increase in the magnitude of the C_6 dispersion terms can compensate for the neglect of higher-order terms, because these terms are much shorter-range than the 6th-order terms. Non-bonded potentials that include higher-order dispersion coefficients have been proposed in the past [78], but have not been widely adopted. The rigorous inclusion of higher-order dispersion terms in molecular-mechanical force fields may result in more accurate calculations of dispersion interactions and C_6 coefficients that are in better agreement with the experimental and XDM values. XDM provides a first-principles method of obtaining these coefficients, which greatly simplifies the parameterization of these additional terms.

Although many-body effects are sometimes invoked as a neglected source of dispersion energy, analysis using methods like XDM have generally found that pairwise interactions account for the bulk of the dispersion interaction, while non-additive many-atom dispersion is slightly repulsive in general and only accounts for a very small fraction of the total dispersion energy [40; 79; 80].

Finally, the widespread practice of applying a switching function to terminate the Lennard-Jones interaction at a moderate distance (e.g., 12 Å) has also caused the force field C_6 dispersion coefficients to be exaggerated. This truncation is also used when the force field is parameterized, so the parameterization procedure of some force fields tends to assign a spuriously large C_6 coefficient to the parameterized atoms to compensate for the neglected long-range dispersion interactions (although a correction is made for this approximation in some force fields). Fisher et al. found that the

Table 2.1: Calculated liquid properties obtained with force fields for which the molecular C_6 coefficients differ from the XDM values by a wide margin. The data are taken from Ref. 83. Several molecules that contain sulfur or bromine atoms have enthalpies of vaporization that are significantly different from the experimental values. Dispersion coefficients, densities, and enthalpies of vaporization are given in atomic units, kg/m^3 , and kJ/mol respectively.

Force field	Molecule	$C_{6,\text{FF}}$	$C_{6,\text{XDM}}$	ρ expt.	ρ calc.	ΔH_{vap} expt.	ΔH_{vap} calc.
CGenFF	dibromomethane	2227.7	961.6	2496.8	2435.3	37.67	43.53
GAFF	dibromomethane	2191.2	961.6	2496.8	1962.8	37.67	28.83
GAFF	1-bromobutane	3436.0	1933.3	1275.1	1176.4	36.60	35.98
CGenFF	1,2-ethanedithiol	2630.8	1493.1	1113	1167	41.85	48.16
OPLS	1,2-ethanedithiol	2600.4	1493.1	1113	1157	41.85	46.68

enthalpies of vaporization of liquids in the Virtual Chemistry test set were systematically overestimated when the long-range dispersion interactions were calculated [81], which is consistent with our conclusion that the dispersion coefficients for these force fields are larger than they should be on physical grounds. For homogeneous systems, there are methods of correcting for the neglected long-range dispersion interactions without explicitly calculating them using a lattice-summation method [82], although this has not been used universally in force field parameterization. The redevelopment of force fields to include long-range dispersion interactions would likely result in smaller C_6 dispersion parameters.

If the force field significantly underestimates or overestimates the molecular C_6 coefficient relative to the XDM value, the simulated physical properties of the liquid are often greatly in error. For example, the GAFF force field overestimates the total dispersion coefficient of compounds containing bromine, such as dibromomethane and 1-bromobutane, by a large margin. As shown in Table 2.1, the density and enthalpy of vaporization of these compounds are significantly in error when anomalously-large C_6 parameters are used to simulate these liquids. In agreement with our analysis, Adluri et al. showed that the GAFF and CGenFF Lennard-Jones parameters for bromine

are not optimal [27]. The bromine dispersion coefficients for the reparameterized force field were closer to the XDM values (206.7 a.u. and 163.1 a.u. for the revised CGenFF and GAFF models, respectively). The CGenFF and OPLS force fields also predict anomalously high molecular dispersion coefficients for 1,2-ethanedithiol. These models overestimate the density and enthalpy of vaporization (Table 2.1).

2.4.2 Atomic Dispersion Coefficients

A comparison of the force-field and XDM homoatomic C_6 dispersion coefficients allows the validity of force-field atom typing to be assessed and reveals whether the systematic overestimation of the dispersion coefficients can be traced to particular elements. The average of the homoatomic dispersion coefficients for each element, and the accompanying standard deviation, were calculated for the full test set and are reported in Table 2.2. The force-field dispersion coefficients for each chemically-unique atom in the test set are plotted against the equivalent XDM dispersion coefficients in Figure 2.2. A more narrow distribution that is restricted to H, C, N, and O atoms is presented in Figure 2.3.

In general, the XDM dispersion coefficients span a modest range for a given element, with coefficients of variation that range from 0.01 for fluorine to 0.11 for oxygen and nitrogen. This suggests that, for organic compounds, variations in the chemical environment of an atom do not drastically affect the strength of its dispersion interactions (although this is not the case for metals [42] and for changing oxidation states [41]). Thus, physically-realistic force fields should only have moderate deviations from the average dispersion coefficient for each element. The comparisons of the force-field and XDM homoatomic dispersion coefficients are discussed for each element throughout this section.

Table 2.2: Average values of the homoatomic C_6 dispersion coefficients in atomic units, grouped by element, for the compounds in the Virtual Chemistry test set. The breadth of the distribution of dispersion coefficients for a given element is indicated by its standard deviation. For the XDM values, this represents the spread of dispersion coefficients for an element due to the variety of chemical environments in the test set. For the MM models, this reflects that various atom types with different Lennard-Jones parameters are used to represent the same element.

Element	XDM	CGenFF	GAFF	OPLS
H	2.5 ± 0.2	1.7 ± 0.7	1.2 ± 0.5	1.7 ± 0.7
C	22.0 ± 1.3	40.4 ± 9.1	44.4 ± 5.2	39.4 ± 9.8
N	15.9 ± 1.8	44.3 ± 20.9	58.2 ± 0.0	59.6 ± 6.1
O	12.6 ± 1.4	34.6 ± 16.7	43.7 ± 6.2	38.3 ± 7.2
F	8.2 ± 0.1	9.9 ± 0.8	16.3 ± 0.0	9.9 ± 0.8
S(II)	91.8 ± 2.6	212.7 ± 54.4	149.0 ± 0.0	205.9 ± 55.5
S(IV)	58	208	145	149
Cl	70.8 ± 1.5	154.0 ± 3.8	135.0 ± 0.0	135.0 ± 0
Br	134.2 ± 2.2	238.0 ± 0.0	356.0 ± 0.0	238.0 ± 0

Hydrogen

The force-field dispersion coefficients for hydrogen atoms are systematically underestimated relative to the XDM values. As shown in Table 2.2, the average XDM C_6 coefficient for hydrogen atoms is 2.5 a.u., which is higher than the averages of 1.7 a.u., 1.2 a.u., and 1.7 a.u. for the CGenFF, GAFF, and OPLS force fields, respectively. As hydrogen atoms have relatively weak dispersive and repulsive interactions in comparison to their parent atoms, Lennard-Jones parameters of hydrogen atoms are sometimes assigned somewhat arbitrary parameters or are even neglected in some cases.

XDM calculates atom-in-molecule dispersion coefficients by using the Hirshfeld [84] method to partition the molecular electron density into atomic regions, although the choice of partitioning scheme is somewhat arbitrary. Partitioning of the electron density between hydrogen atoms and their parent atoms is particularly sensitive to

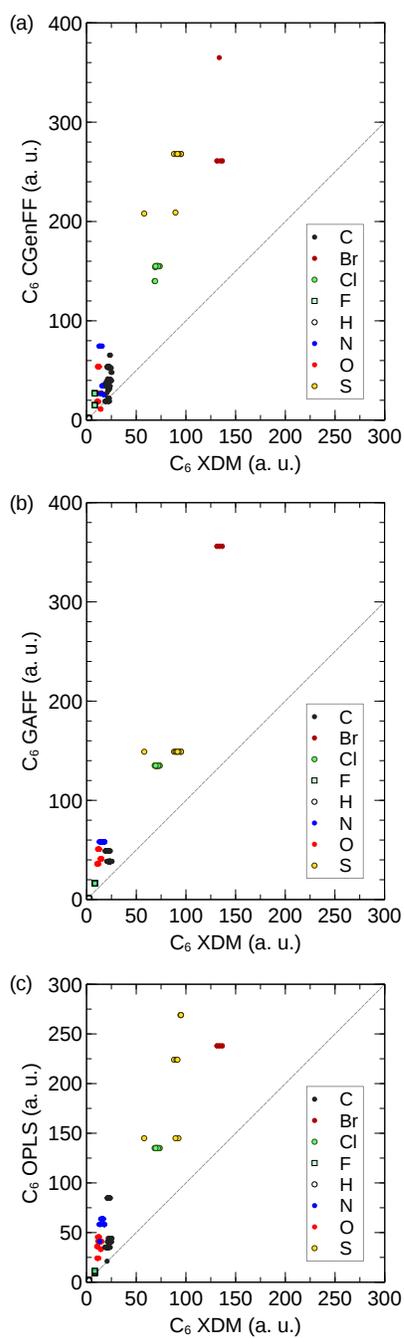


Figure 2.2: Correlation between the homoatomic dispersion coefficients of the (a) CGenFF, (b) GAFF, and (c) OPLS force fields and XDM. Each point represents a unique atom from the Virtual Chemistry set.

the method used. The choice of the Hirshfeld method causes the hydrogen atoms to account for a greater portion of the molecular dispersion interactions than do the

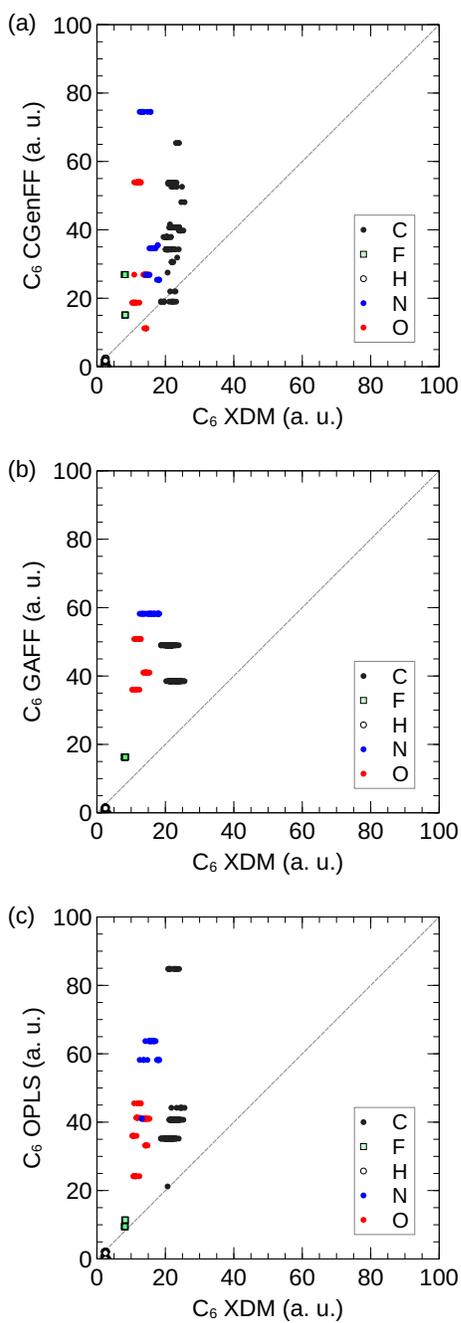


Figure 2.3: Correlation between the homoatomic dispersion coefficients of the (a) CGenFF, (b) GAFF, and (c) OPLS force fields and XDM, restricted to $C_{6,\text{XDM}} = [0, 100 \text{ a.u.}]$, enclosing H, C, N, O, and F atoms.

force fields. Newer methods, like the Iterative Hirshfeld Method [85], could yield hydrogen C_6 dispersion coefficients closer to the force-field values. The net dispersion

interaction for an atom and its bound hydrogens is more consistent between different partitioning schemes. These grouped dispersion coefficients also show a systematic overestimation of dispersion coefficients by the force-field models.

Carbon

Carbon atoms have sizable dispersion coefficients due to their high polarizability ($\langle C_{6,\text{XDM}} \rangle = 22.0 \pm 1.3$ a.u.). Bonding partners have the largest effect on the dispersion coefficient of carbon atoms. As shown in Figure 2.4, the XDM dispersion coefficient of the electron-poor tertiary carbon of *t*-butylamine is particularly low, with a value of 18.7 a.u. At the other extreme, the β -carbon of 1,1-dichloroethene has a C_6 coefficient of 25.6 a.u.

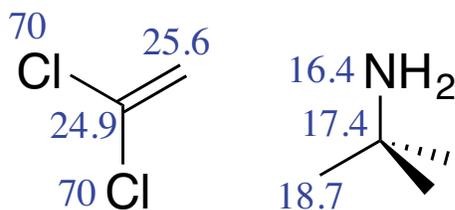


Figure 2.4: The atomic XDM dispersion coefficients (blue, a.u.) for electron-rich carbon atoms, such as those in 1,1-dichloroethene, and those of electron-poor carbon atoms like those of *t*-butylamine.

The force fields all overestimate the majority of the carbon dispersion coefficients, with the average values roughly a factor of two higher than the XDM values (Table 2.2). Moreover, the force fields give a much larger range of dispersion coefficients for carbon than obtained with XDM. The CGenFF force field shows a particularly broad range; the lowest C_6 value is 19.0 a.u. for methyl-group carbons (atom type CG311) and the highest is 65.4 a.u. for carbonyl carbons (atom type CG2O1). Of the three force fields, GAFF has the most narrow range of C_6 coefficients for carbon, but still gives a standard deviation of $\sigma = 5.2$ a.u. compared to $\sigma = 1.3$ a.u. with XDM.

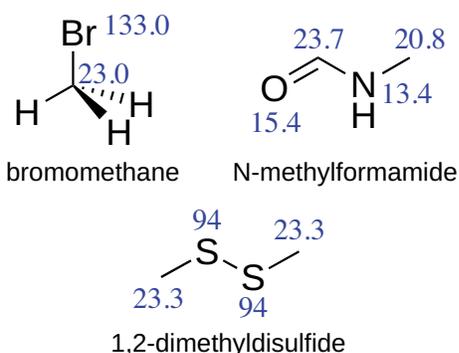


Figure 2.5: Examples of XDM dispersion coefficients (a.u.) for methyl-group carbon atoms.

Conversely, within some force-field carbon atom types, the XDM dispersion coefficients show significant variation. For instance, the CGenFF CG331 atom type represents all methyl groups. This type includes the electron-deficient methyl group of N-methylformamide, which has an XDM dispersion coefficient of 20.8 a.u. (Figure 2.5). At the other extreme, the XDM-computed dispersion coefficient range from 22 a.u. to 23 a.u. for carbons bound to bromine or sulfur, as in dimethyl sulfide or bromoethane. The XDM dispersion coefficients also span a significant range for aromatic carbons (Figure 2.6), such as those represented by the CGenFF CG2R61 atom type. The computed XDM coefficients for this atom type range between 21.5–25.5 a.u. The dispersion coefficients of carbon atoms in electron-rich heteroaromatics, like furan and pyrrole, range between 24–25 a.u., which is incrementally higher than the value of 23.6 a.u. calculated for benzene. There is also a significant variation in the C_6 coefficients of aromatic carbons due to substituent effects. For example, the *ipso* carbon of anisole has a C_6 coefficient of 21.9 a.u., but the *para*-carbon has a dispersion coefficient of 23.8 a.u. As force fields use the same dispersion coefficients for all atoms of the same type, the calculated strength of interatomic dispersion interactions could be in error by up to 10% due to the use of atom types.

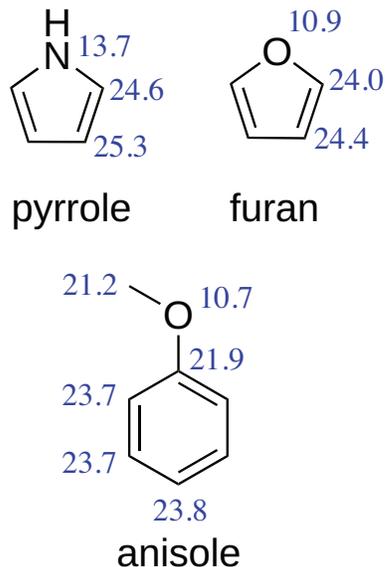


Figure 2.6: Examples of XDM dispersion coefficients (a.u.) for carbon atoms in aromatic systems.

Nitrogen

XDM predicts nitrogen atoms to have dispersion interactions of moderate strength, with an average C_6 coefficient of 15.9 a.u. Nitrogen atoms have one of the broadest distributions of C_6 coefficients for this test set; electron-poor amide nitrogens have particularly low dispersion coefficients (e.g. 13.4 a.u. in N-methylformamide), while electron-rich alkyl nitrogens have high dispersion coefficients (e.g. 17.2 a.u. in ethane-1,2-diamine). The XDM dispersion coefficients for these molecules are presented in Figure 2.7.

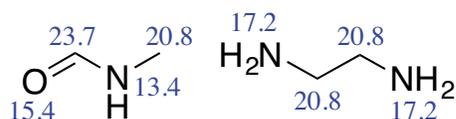


Figure 2.7: Examples of XDM dispersion coefficients (a.u.) for an electron-poor amide nitrogen (N-methylformamide) and an electron-rich amine nitrogen (ethane-1,2-diamine).

The dispersion coefficients for nitrogen atoms are systematically overestimated by the GAFF and OPLS force fields, which give average C_6 values of 58.2 a.u. and 59.6 a.u., respectively. The CGenFF coefficients for amines are somewhat closer to the XDM values (12–34 a.u. for CGenFF vs. 14–17 a.u. for XDM), but the coefficients for amide nitrogens (i.e., NG2S0, NG2S1, NG2S1, NG2S2, and NG2S3 atom types) are assigned an anomalously large value (74.5 a.u.). The XDM coefficients for these atoms are actually lower than the average for nitrogen atoms, ranging from 12–15 a.u. The CGenFF amide-nitrogen parameters are shared with the protein backbone of the CHARMM36 force field [86]. The Lennard-Jones parameters for these amide nitrogens were adjusted to provide more accurate backbone hydrogen bonding, but this appears to have caused the C_6 dispersion coefficient to be anomalously high. Although this modification of the Lennard-Jones parameters may describe the energetics of short-range interactions more accurately, the long-range dispersion interactions will be overestimated as a result.

Oxygen

The XDM C_6 coefficients for oxygen are generally smaller than those of carbon and nitrogen ($\langle C_{6,O} \rangle = 12.6$ a.u.) and span a fairly narrow range of values ($\sigma_O = 1.4$ a.u.). Electron-rich carbonyl oxygens have coefficients in the 13–15 a.u. range, while alcohols have smaller coefficients, between 12–12.5 a.u. The oxygen C_6 parameters from the three force fields are considerably larger than the XDM values. CGenFF has the smallest oxygen dispersion coefficients, with an average of 34.6 a.u. Carbonyl and ether oxygens are only moderately overestimated ($C_6 < 30$ a.u.), but the coefficients for alcohols are 4–5 times larger than the XDM values (e.g., 53.9 a.u. for OG311). The average C_6 coefficients of the GAFF and OPLS force fields are systematically higher than the XDM values across most oxygen atom types, by factors of 3.5 and

3.0, respectively. For these force fields, the strengths of dispersion interactions for oxygen atoms are similar to those for carbon atoms, despite oxygen’s much lower polarizability.

Sulfur

For sulfur, results for the II and IV oxidation states will be considered separately due to the large effect of oxidation state on C_6 dispersion coefficients [41]. The average XDM dispersion coefficient for S(II) atoms is 91.8 a.u. and the standard deviation is 2.6 a.u. The coefficient averages for the CGenFF and OPLS models are overestimated by more than a factor of two and the range of values are extremely large, with standard deviations of 54.5 and 55.5 a.u., respectively. For the CGenFF force field, this is because of the difference between the disulfide and sulfide atom types; the SG301 atom type (C–S–S–C) has a dispersion coefficient of 209 a.u., while the SG311 atom type (SH, –S–) has a dispersion coefficient of 268 a.u.

Sulfolane is the only compound in the test set where the sulfur atom is in the IV oxidation state. The XDM C_6 coefficient for this atom (57.8 a.u.) is significantly lower than for the S(II) atoms. The CGenFF and OPLS force fields assign the S(IV) atom type (SG302) a dispersion coefficient of 208 a.u., which is lower than for S(II), although it is still overestimated relative to XDM. The GAFF force field assigns S(IV) the same Lennard-Jones parameters as S(II). Given that the sulfur dispersion coefficients are very sensitive to the oxidation state, several distinct sets of Lennard-Jones parameters should be determined for this element.

Halogens

Fluorine has a small XDM dispersion coefficient ($\langle C_{6,F} \rangle = 8.2$ a.u.), consistent with its low atomic polarizability. The standard deviation of these values is small ($\sigma =$

0.1 a.u.), suggesting that a single set of Lennard-Jones parameters are appropriate for fluorine atoms in organofluorines. The fluorine CGenFF and OPLS force-field dispersion coefficients are in reasonable agreement with the XDM values, although the dispersion coefficients for GAFF are roughly double the XDM value.

Chlorine has a large average XDM dispersion coefficient ($\langle C_6 \rangle = 70.8$ a.u.), but also has a relatively small standard deviation ($\sigma = 1.0$ a.u.). The force-field dispersion coefficients for chlorine are much larger than the XDM values. The GAFF and OPLS force field share Lennard-Jones parameters for chlorine, so their C_6 coefficients are also the same ($C_6 = 135$ a.u.), while the CGenFF force field assigns even larger values ($\langle C_{6,Cl} \rangle = 154$ a.u.).

Bromine has the largest XDM dispersion coefficients in the test set ($\langle C_6 \rangle = 134$ a.u.). The bromine dispersion coefficients are overestimated by all three force fields. In particular, the GAFF force field grossly overestimates the strength of the bromine C_6 , with an average value of 356 a.u. The performance of GAFF for the physical properties of organobromine liquids is notably poor, suggesting that the GAFF Lennard-Jones parameters are not physically realistic. Adluri et al. showed that the GAFF model for bromomethane could be improved by reparameterizing the Lennard-Jones terms for Br [27].

2.4.3 The Drude Force Field

The Drude force field incorporates the effect of induced polarization by adding charged “Drude” particles that are harmonically tethered to their parent atoms [87]. This model uses a Lennard-Jones potential to represent dispersion interactions, although the parameters for these models generally need to be refit so that they are appropriate for molecules interacting through different Coulombic-interaction terms that are present in polarizable force fields. To evaluate the dispersion parameters of the Drude

model, the XDM dispersion coefficients were calculated for 73 molecules from the July 2015 revision of the Drude force field. The correlation between the Drude and XDM C_6 coefficients is presented in Figure 2.8.

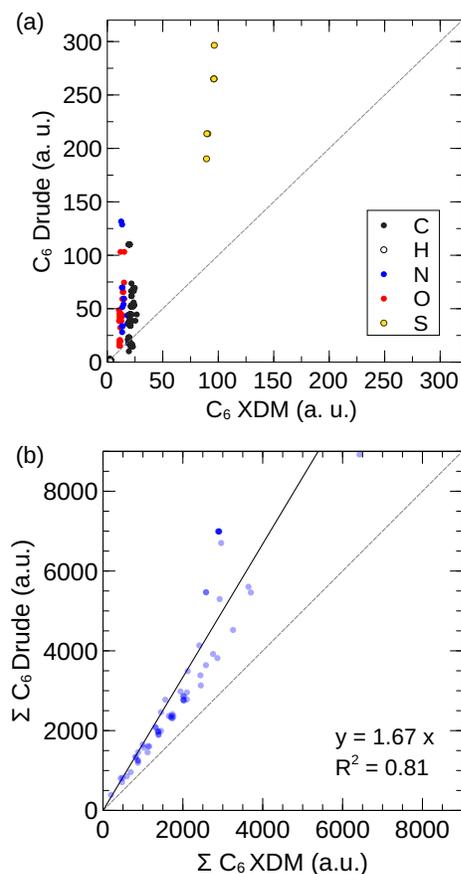


Figure 2.8: (a) Correlation between the homoatomic dispersion coefficients calculated using the Drude force field and XDM. Each point represents a unique atom from the test set of molecules available in the current version of the Drude force field. (b) Correlation between the Drude and XDM molecular C_6 dispersion coefficients. Each point represents a single molecule of the test set.

The correlation between the XDM and Drude molecular C_6 coefficients is poorer than for the non-polarizable models; the regression coefficient is 1.67 and the coefficient of determination is 0.81 (Figure 2.8 (b)). The distribution of atomic dispersion coefficients for the Drude force field is extremely wide, with values ranging from 20 a.u.

Table 2.3: Homoatomic dispersion coefficients (in atomic units) for Drude atom types where the force-field C_6 is much larger than the corresponding XDM value.

Atom type	Environment	XDM	Drude
CD31FA	C1 carbon of carbohydrates	19.8	110.1
ND2R5D	GUA/ADE 5-member ring	13.4	128.7
OD2C2B	carboxylate O, anionic phosphate, lipids	12.0	103.2
ND2A3	amide, tertiary DMA	12.5	131.7
SD31A	alkyl thiol sulfur	96.4	298.5

to more than 100 a.u. for some C, N, and O atom types (Figure 2.8 (a)). The dispersion coefficients for sulfur, spanning between 200 and 300 a.u., are also overestimated by a large margin. As with the non-polarizable force fields, the systematic overestimation of the dispersion coefficients is likely due to neglect of higher-order C_8 terms and premature truncation of the dispersion-energy summation.

The poor correlation between Drude and XDM dispersion coefficients is exacerbated by five atom types where the force-field dispersion coefficient is several times larger than the corresponding XDM value (Table 2.3). Atypically large Lennard-Jones ϵ parameters have been assigned to these atoms. For example, atom type ND2R5D, which represents the nitrogen at the 9 position of purines, has an ϵ parameter of -0.23 kcal/mol, which is a factor of 2 larger than is typical for nitrogen atom types. The procedure that determined these parameters included unconventional target data, such as QM interaction energies and lattice energies [88; 89; 90]. These terms are less sensitive to the Lennard-Jones parameters than traditional parameterization target data, like the density or enthalpy of vaporization, so it is possible that the parameters were over-fit. Imposing constraints on the Lennard-Jones parameters to ensure that the C_6 coefficients are in the typical range for each element would be a simple way of avoiding parameters that cause the long-range dispersion interactions to become unrealistically large.

2.4.4 Water Models

One of the most common applications of generalized force fields is to simulate organic solutes in an aqueous solution. A wide range of water models are available, including models with charges at the three atomic centers (e.g., TIP3P), models with an off-center charge (e.g., TIP4P), and polarizable water models (e.g., SWM4-NDP). The generalized force fields evaluated here are typically used with the TIP3P water model. The TIP3P model overestimates the dielectric constant and diffusivity of water [91], so there has been interest in adopting water models that describe the properties of water more accurately. To describe solvation properly, the dispersion coefficients of the water model must be balanced with those of the solute force field. The C_6 dispersion coefficients for 11 popular water models are presented in Table 2.4.

A molecular C_6 coefficient for H_2O of 45.4 a.u. has been estimated by Zeiss and Meath using photoabsorption and high energy inelastic scattering experiments [56]. XDM is in close agreement with this value, with a predicted C_6 coefficient of 45.8 a.u. The TIP3P, TIP3P-FB, SPC/E, OPC3, TIP4P, and TIP4P-Ew water models have C_6 coefficients that are close to the experimental value, ranging from 43 to 49 a.u. The TIP4P-FB, TIP4P/2005, and SWM6-NDP models modestly overestimate the dispersion coefficient, ranging from 50–55 a.u. The SWM4-NDP and TIP4P-D models overestimate the C_6 coefficient by a large margin, with values of 63.7 and 65.3 a.u., respectively.

Our analysis of the molecular C_6 coefficients of the generalized force fields in Section 2.4.1 indicated that the dispersion interactions are systematically overestimated in comparison to the predictions by XDM. In contrast, many of the water models used in combination with these force fields have dispersion coefficients that are comparable to the water dispersion coefficient calculated by XDM or determined experimentally. This suggests that the coefficients for the dispersion interaction between TIP3P-model

Table 2.4: Molecular C_6 coefficients for various molecular-mechanical water models that include a Lennard-Jones 12-6 potential to represent dispersion. The XDM value was calculated for a single gas-phase water molecule using PBE0/aug-cc-pVTZ.

Water model	C_6 (a.u.)	Ref.
TIP3P	43.2	57
TIP3P-FB	46.6	93
SPCE	45.4	94
TIP4P	44.3	57
TIP4P-Ew	47.4	95
TIP4P-FB	52.3	93
TIP4P/2005	53.4	58
TIP4P-D	65.3	96
OPC3	48.5	97
SWM4-NDP	63.7	98
SWM6	50.3	99
XDM	45.8	this work
exptl.	45.4	56

water and a solute described using one of these generalized force fields could be unbalanced. Best et al. found that water–protein interactions were predicted to be too strong and had to be attenuated to describe intrinsically disordered proteins correctly [92].

2.4.5 Force Field Development using XDM-Derived Dispersion Coefficients

To test if XDM-derived C_6 coefficients can be used to parameterize a molecular-mechanical force field, we developed new Lennard-Jones parameters for benzene, toluene, cyclohexane, n-pentane, and n-hexane using XDM data. The internal energy terms and atomic charges of the CGenFF force field models of these molecules were used without modification, but the atomic Lennard-Jones parameters were selected such that the molecular C_6 dispersion coefficient is near the XDM value.

To perform the fitting for benzene, toluene, and cyclohexane, a 4-dimensional grid of ε_H , σ_H , ε_C , and σ_C was considered for each unique C–H pair of CGenFF atom types (see Table 2.6). The Lennard-Jones (LJ) parameters for the sp^2 carbons and bonded hydrogens for benzene were transferrable to toluene and, as such, only the additional parameters for the methyl-group atoms were fit to the toluene reference data. Parameter sets yielding a molecular dispersion coefficient that deviated from the XDM value by more than a given threshold were discarded (see Table 2.6). Molecular dynamics simulations were performed using the remaining parameter sets to calculate the density and enthalpy of vaporization of the liquid. Dispersion interactions were calculated using the LJ-PME method so that long-range dispersion interactions were included [81]. The enthalpy of vaporization was calculated from the average potential energy of the liquid simulation,

$$\Delta H_{vap} = RT + \langle \mathcal{V}_{pot} \rangle_{gas} - \langle \mathcal{V}_{pot} \rangle_{liquid}. \quad (2.1)$$

The final, optimum parameters were those that yielded the lowest deviation from experiment based on the target function,

$$\delta(\sigma, \varepsilon) = \left(\frac{\rho_{exptl.} - \rho_{calc}}{\rho_{exptl.}} \right)^2 + \left(\frac{\Delta H_{vap,exptl.} - \Delta H_{vap,calc}}{\Delta H_{vap,exptl.}} \right)^2, \quad (2.2)$$

The parameters for the linear alkanes were determined by the same XDM-constrained parameterization procedure, but only the top 5 parameter pairs from the search for the analogous sp^3 carbon atom types from cyclohexane and toluene were considered. This procedure successfully identified effective parameters from only 25 MD simulations of trial parameters.

Both the densities and enthalpies of vaporization predicted using the new parameters are in good agreement with the experimental values and are in better agreement

Table 2.5: Physical properties of 5 different liquids calculated using the original CGenFF force field and a reparameterized force field (FF-XDM). Units are kg/m^3 and kcal/mol for density (ρ) and enthalpy of vaporization (ΔH_{vap}), respectively.

Compound	FF-XDM		CGenFF		Exptl.	
	ρ	ΔH_{vap}	ρ	ΔH_{vap}	ρ	ΔH_{vap}
benzene	869	33.6	861	28.5	876	33.9
toluene	872	38.3	864	46.7	865	38.1
cyclohexane	766	31.8	783	35.3	778	32.0
n-pentane	621	20.6	628	11.9	626	26.2
n-hexane	667	34.05	666	27.91	655	31.51

Table 2.6: Parameterization space for Lennard-Jones parameters of dispersion-bound liquids. The range of Lennard-Jones σ and ϵ parameters in the search for each atom type is given. 10 parameters in that range were evaluated, yielding 10000 potential parameter combinations. The percentage of parameter combinations eliminated by the XDM criteria and the allowable deviation of the force field molecular C_6 from the XDM value are also given.

Compound	Atom type	σ range (Å)	ϵ range (kcal/mol)	Parameter space reduction (%)	XDM deviation threshold
benzene	CG2R61	[3.380, 3.720]	[0.214, 0.372]	94.86%	5%
	HGR61	[2.250, 2.590]	[0.047, 0.204]		
toluene	CG331	[3.553, 3.753]	[0.279, 0.374]	98.69%	1.5%
	HGA3	[2.288, 2.488]	[0.053, 0.148]		
cyclohexane	CG321	[3.481, 3.681]	[0.187, 0.282]	98.11%	1%
	HGA2	[2.288, 2.488]	[0.099, 0.194]		

than the original CGenFF parameters in some cases (Table 2.5). This is particularly true for the enthalpies of vaporization. Between 95–98% of potential parameters combinations were excluded by the XDM criteria, making this approach much more efficient than a traditional grid search of the Lennard-Jones parameter space. This suggests that XDM molecular dispersion coefficients can provide bounds for Lennard-Jones parameters that limit the parameter space to physically-realistic values.

2.5 Conclusions

The XDM method was used to calculate the C_6 dispersion coefficients of 88 molecules from the Virtual Chemistry test set. These density-functional-theory-derived dispersion coefficients were compared to dispersion coefficients defined through the Lennard-Jones potential parameters of the CGenFF, GAFF, and OPLS molecular-mechanical force fields.

All three force fields systematically overestimate the molecular dispersion coefficients relative to XDM. The empirical parameterization process likely led to anomalously high C_6 dispersion coefficients due to the neglect of long-range dispersion interactions, higher-order dispersion terms, and induced electronic polarization. Next-generation force fields that account for these terms may return more realistic C_6 dispersion interactions. Improved models for five organic liquids were successfully developed, demonstrating that it is possible to define accurate force fields with physically-realistic C_6 dispersion coefficients while using the standard form of the force field. This procedure can also dramatically reduce the cost of the parameterization by reducing the number of putative parameters to the small subset that are consistent with the XDM-derived dispersion coefficients.

Molecular-mechanical force fields use the same dispersion interaction parameters for all atoms of the same type. In some cases, the XDM dispersion coefficients spanned a significant range of values for atoms of the same type, indicating that the Lennard-Jones parameters are not optimal for some atoms. This is particularly true for the GAFF and OPLS force fields, which have fewer variants of Lennard-Jones parameters than the CGenFF force field. One example where molecular-mechanical atom typing breaks down are methyl carbon atoms in electron-rich versus electron-poor environments, which XDM predicts to have significantly different dispersion coefficients. The

introduction of new atom types motivated by groupings of C_6 values may improve the accuracy of force-field models.

The Drude polarizable force field displayed an even poorer correspondence with the XDM C_6 coefficients than the three non-polarizable models. Several atom types in the Drude force field have dispersion coefficients that are many times greater than the XDM values. It is possible that the increased number of terms in this force field, due to the incorporation of the polarizability, cause the parameters to be underdetermined. The parameter-fitting process could benefit from additional constraints, such as physically-reasonable molecular dispersion coefficients derived from XDM.

Some standard water models, such as TIP3P, have dispersion coefficients that are very similar to the XDM and experimental values. Other models, such as TIP4P-D and SWM4-NDP, overestimate the magnitude of the dispersion coefficient by up to 50%. Thus, the C_6 coefficient may be a worthwhile term to consider in the evaluation of water models in order to ensure the dispersion interactions are physically realistic.

The use of quantum-chemical methods like XDM in the parameterization of dispersion interactions could provide new opportunities to develop more realistic force fields, as illustrated here in the cases of benzene, toluene, cyclohexane, n-pentane, and n-hexane. An immediate application of XDM will be to validate Lennard-Jones parameters of force fields to ensure that the molecular and atomic C_6 dispersion coefficients do not deviate from the XDM values by a large margin. XDM also provides an effective means to calculate the C_8 and C_{10} dispersion coefficients, which are currently neglected from conventional molecular-mechanical force fields. Parameterizing these terms had been impractical because the model is already underdetermined, but XDM could provide reasonable *ab initio* values. This suggests a general strategy for parameterizing the non-bonded parameters for force fields, where XDM is used to assign the atomic dispersion coefficients. The repulsive component could be derived

from a QM potential energy surface, ab initio molecular dynamics (AIMD) simulation, topological analysis of the electron density, or empirical fitting from molecular dynamics simulations of condensed states.

Chapter 3

Evaluating the Dispersion Coefficients of General Force Fields for Proteins Using XDM

This chapter is adapted with permission of the publisher from the article:

Walters, E., Mohebifar, M., Johnson, E.R., Rowley, C. N., Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-Hole Dipole Moment Model, *J. Phys. Chem. B*, **2018**, 122 (26), 6690–6701 doi: 10.1021/acs.jpcc.8b02814

3.1 Abstract

London dispersion is one of the fundamental intermolecular interactions involved in protein folding and dynamics. The popular CHARMM36, Amber ff14sb, and OPLS All-atom (OPLS-AA) force fields represent these interactions through the C_6/r^6 term of the Lennard-Jones potential. The C_6 parameters are assigned empirically, so these

parameters are not necessarily a realistic representation of the true dispersion interactions. In this work, dispersion coefficients of all three force fields were compared to corresponding values from quantum-chemical calculations using the exchange-hole dipole moment (XDM) model. The force field values were found to be roughly 50% larger than the XDM values for protein backbone and side-chain models. The CHARMM36 and Amber OL15 force fields for nucleic acids were also found to exhibit this trend. To explore how these elevated dispersion coefficients affect predicted properties, the hydration energies of the side-chain models were calculated using the staged REMD-TI method of Deng and Roux for the CHARMM36, Amber ff14sb, and OPLS-AA force fields. Despite having large C_6 dispersion coefficients, these force fields predict side-chain hydration energies that are in generally good agreement with the experimental values, including for hydrocarbon residues where the dispersion component is the dominant attractive solute–solvent interaction. This suggests that these force fields predict the correct total strength of dispersion interactions, despite C_6 coefficients that are considerably larger than XDM predicts. An analytical expression for the water–methane dispersion energy using XDM dispersion coefficients shows that that higher-order dispersion terms (i.e., C_8 and C_{10}) account for roughly 37.5% of the hydration energy of methane. This suggests that the C_6 dispersion coefficients used in contemporary force fields are elevated to account for the neglected higher-order terms. Force fields that include higher-order dispersion interactions could resolve this issue.

3.2 Introduction

XDM calculations provide a straightforward way to evaluate force-field dispersion coefficients using quantum chemistry. Mohebifar et al. used XDM to assess the

dispersion coefficients of the GAFF, CGenFF, and OPLS-AA force fields for a set of small organic molecules [55]. This report showed that the molecular C_6 coefficients for GAFF, CGenFF, and OPLS-AA force fields are systematically 50% higher than the XDM values. This trend is not universal; the C_6 coefficient of the TIP3P water model is close to the XDM value, suggesting that the strength of dispersion interactions may be unbalanced in simulations of molecules in solution.

In this work, we extend this analysis to evaluate the dispersion coefficients of molecular-mechanical force fields for proteins (Amber ff14sb [34], CHARMM36 [100], and OPLS-AA [101]) using XDM. The Amber OL15 and CHARMM36 nucleic acid force fields are also evaluated. We note that many earlier versions of these force fields use the same Lennard-Jones parameters. The relationship between the molecular dispersion coefficient and the computed hydration energy is investigated. Finally, we explore the possible origins and potential resolutions to issues with the treatment of dispersion in molecular-mechanical force fields.

3.3 Computational Methods

3.3.1 Side-Chain Models

Molecular models for the amino acid side chains were defined by deleting the amine and carboxylic acid moieties and replacing the α -carbon with a hydrogen atom. The hydration energy of N-methylacetamide (NMA), a model for the protein backbone, was also considered. The structures of these species are presented in Figure 3.1.

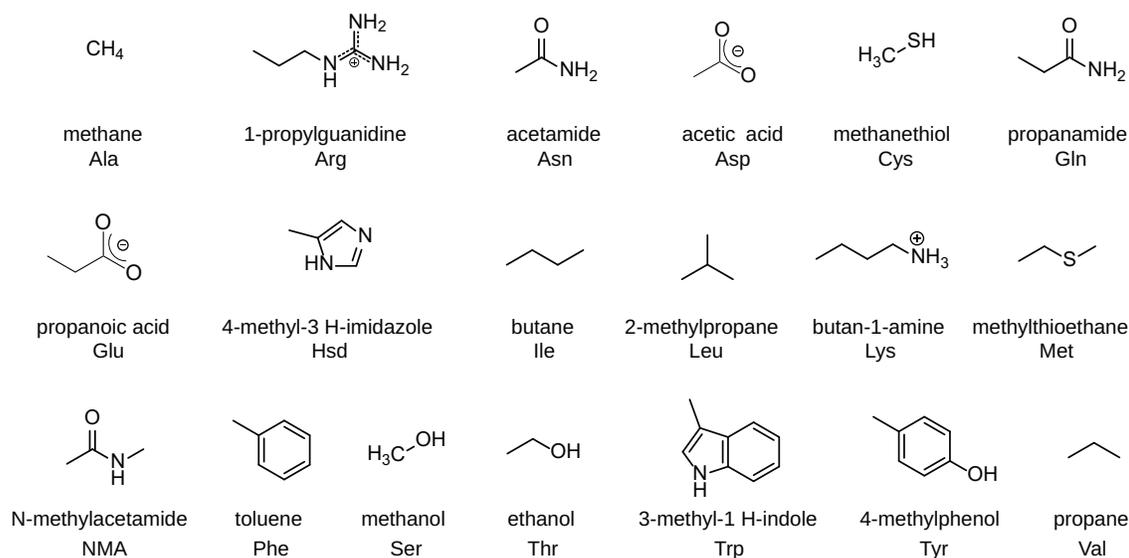


Figure 3.1: Structures of compounds used to model amino acid side chains.

3.3.2 XDM Calculations

To calculate the XDM dispersion coefficients, the geometries of all side-chain and nucleobase structures were optimized using DFT calculations with Gaussian 09 [66]. The PBE0 exchange-correlation functional [63] and the aug-cc-pVTZ basis set [102] were used for all calculations. This functional and basis set combination is generally reliable for the calculation of molecular dipole moments and polarizabilities [67]. The XDM dispersion coefficients of each side chain and nucleobase were calculated from the Gaussian wavefunction file using the postg code [70; 69].

3.3.3 Free Energy Simulations

The hydration energies of the side-chain models were calculated with CHARMM 41b1 using the staged Weeks, Chandler, and Andersen protocol of Deng and Roux [103]. The electrostatic component of the Gibbs energy of solvation was calculated from

thermodynamic integration (TI) simulations with 11 values of λ spaced evenly between 0 and 1, which corresponds to the solute–solvent electrostatic interaction. The dispersion component of the hydration energy is then calculated by turning off the solute–solvent dispersion interactions. Lastly, the repulsive component was calculated from a 9-stage free energy perturbation (FEP) calculation. A correction for the truncation of long-range Lennard-Jones interactions was obtained by calculating the average of the difference of the solute–solvent Lennard-Jones interactions for a 12 Å cutoff vs. a 40 Å cutoff, for a 1 ns simulation of the solute in a 72 Å \times 72 Å \times 72 Å cubic cell of water. Replica exchange was used to sample the configurational space of the states in the TI/FEP process more effectively by allowing exchanges between neighboring replicas every 1000 time steps, following the implementation of Jiang et al [104]. Uncertainties were estimated by dividing the sampled data into three equal sets and taking the standard deviation of the energies calculated from these values. Water molecules were represented using the TIP3P model [57]. A water–vacuum interfacial potential of -520 mV was included in the calculation of the Gibbs energy of hydration of charged residues [105].

3.3.4 Molecular Dynamics Calculations

To provide quantitative examples of dispersion interactions, several molecular dynamics simulations were conducted using Nanoscale Molecular Dynamics (NAMD) 2.12 [106]. In these simulations, the protein and lipid components were described using the CHARMM36 force field, while the Tamoxifen and LSD molecules were described using the CGenFF force field [59]. CHARMM-GUI was used to construct a pure POPC bilayer containing 48 lipids [107]. The coordinates for the folded state of the protein barnase were taken from the crystallographic structure (PDB ID: 1BRS) [108].

Unfolded states were generated by 10 ns simulations at 600 K, followed by a 1 ns re-equilibration at 298 K with the protein backbone atoms restrained to their unfolded coordinates. The simulations of the LSD-GPCR complex were generated from the crystallographic structure of 5-HT_{2B} reported by Wacker et al (PDB ID: 4IB4) [109]. In all cases, the average Lennard-Jones interactions were calculated using the NAM-Denergy plugin of VMD from three 2 ns simulations, where coordinates were saved every 20 ps.

3.3.5 Ab Initio Molecular Dynamics

The AIMD simulation of aqueous methylthiol was performed using CP2K 2.5.1 [110]. The simulation cell contained methylthiol and 162 water molecules in a cubic cell with an edge length of 16.7 Å. The revPBE exchange-correlation functional [111] with the D3 dispersion correction [112] was used, as it has been reported to be effective in describing the structural properties of liquid water [113]. The electron density of the system was represented using a TZVP basis set with a multigrid cutoff of 300 Ry and Goedecker-Teter-Hutter pseudopotentials [114]. The methylthiol radial distribution function was calculated from a 100 ps simulation that followed a 20 ps equilibration. Temperature was regulated using Langevin dynamics with a friction coefficient of 5 ps^{-1} . Bonds containing hydrogen were constrained to a distance of 0.96 Å using the SHAKE algorithm [115].

3.4 Results and Discussion

3.4.1 Molecular Dispersion Coefficients of Amino Acid Side-Chains

One model for understanding the folding and membrane partitioning of proteins is the hydration energies of side-chain models. Molecular models for protein side chains are devised by truncating the side chain with a hydrogen atom that replaces the alpha carbon, as shown in Figure 3.1. This provides a rough estimate of the thermodynamic cost for the side chain to be exposed to the solvent. Together with N-methyl acetamide (NMA), a model for the amide moiety of the protein backbone, these molecules serve as test systems to evaluate the non-bonded force-field parameters.

The atomic and molecular dispersion coefficients of the molecules in Figure 3.1 were calculated using the CHARMM36 [100], OPLS-AA [101], and Amber ff14sb [34] force fields using the relation,

$$C_6 = 4\epsilon\sigma^6 \quad (3.1)$$

where ϵ and σ are the atomic Lennard-Jones parameters defined for a given atom in this force field. The molecular dispersion coefficient ($C_{6,\text{mol}}$) is calculated from the sum of all atomic pairs ($C_{6,ij}$)

$$C_{6,\text{mol}} = \sum_{ij} C_{6,ij}. \quad (3.2)$$

The molecular dispersion coefficients of NMA and the side-chain models calculated using XDM are correlated to the force-field dispersion coefficients in Figure 3.2. All three force fields have molecular C_6 dispersion coefficients for the side chain models

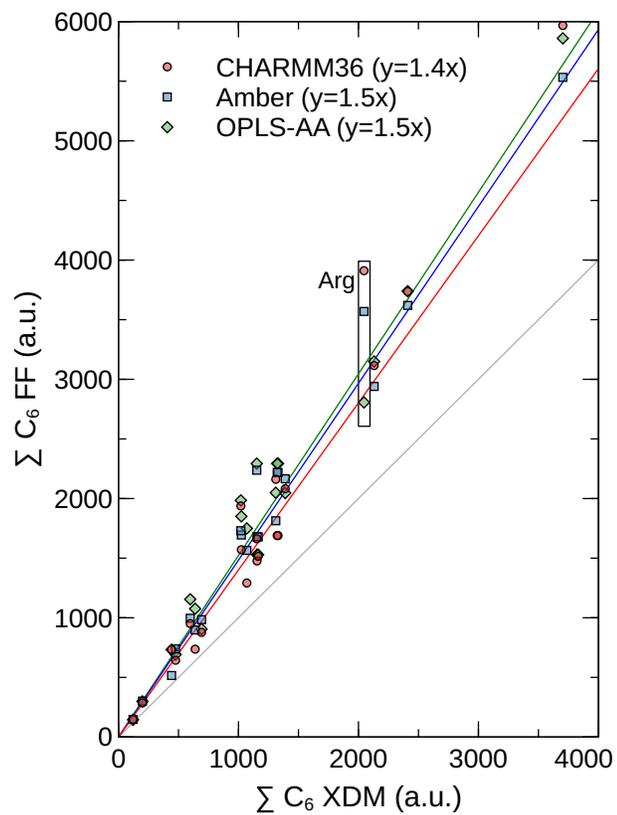


Figure 3.2: The correlation between molecular C_6 coefficients of the protein side chain models. The dispersion coefficients for arginine are highlighted. The gray line indicates where a 1:1 correlation would lie. The points corresponding to the arginine side chain are highlighted because of their broad range.

that are systematically 40–50% higher than the QM values. In Chapter 2, the molecular C_6 coefficients for a set of organic molecules from the Generalized Amber Force Field (GAFF), OPLS-AA, and CGenFF force fields were also found to be roughly 50% larger than the XDM values.

3.4.2 Atomic Dispersion Coefficients

To identify why the force-field molecular C_6 coefficients are so much larger than the XDM values, we compare the individual atomic dispersion coefficients. The correlation between the force field and XDM atomic dispersion coefficients is presented in Figure 3.3. One immediate observation is that the variation in atomic dispersion coefficients for a given element is much larger for the force fields than XDM predicts. The XDM analysis suggests that the range of atomic dispersion coefficients in molecular systems should be modest.¹ Specific comparisons of atomic dispersion coefficients are presented in the following sections.

Carbon

As carbon is the most common non-hydrogen element in the side-chain models, its parameters affect the molecular dispersion coefficients of the side chain-models most significantly. Broadly, the force-field dispersion coefficients for carbon are systematically larger and are distributed over a larger range than the XDM values; the XDM values range from 19 a.u. to 24 a.u.,² but the CHARMM36 values range from 19 a.u. to 65 a.u. This trend is also present in the Amber ff14sb and OPLS-AA force fields, where the ranges for C-atom dispersion coefficients are 38–39 a.u. and 35–84 a.u.,

¹Iterative Hirshfeld approaches have been proposed as an alternative method of performing electron density partitioning, which might lead to a larger variation in atomic dispersion coefficients [85].

²The carbon atom of the carboxylic acid in Asp and Glu have C_6 coefficients of 32 a.u. due to the electron-richness of the anion.

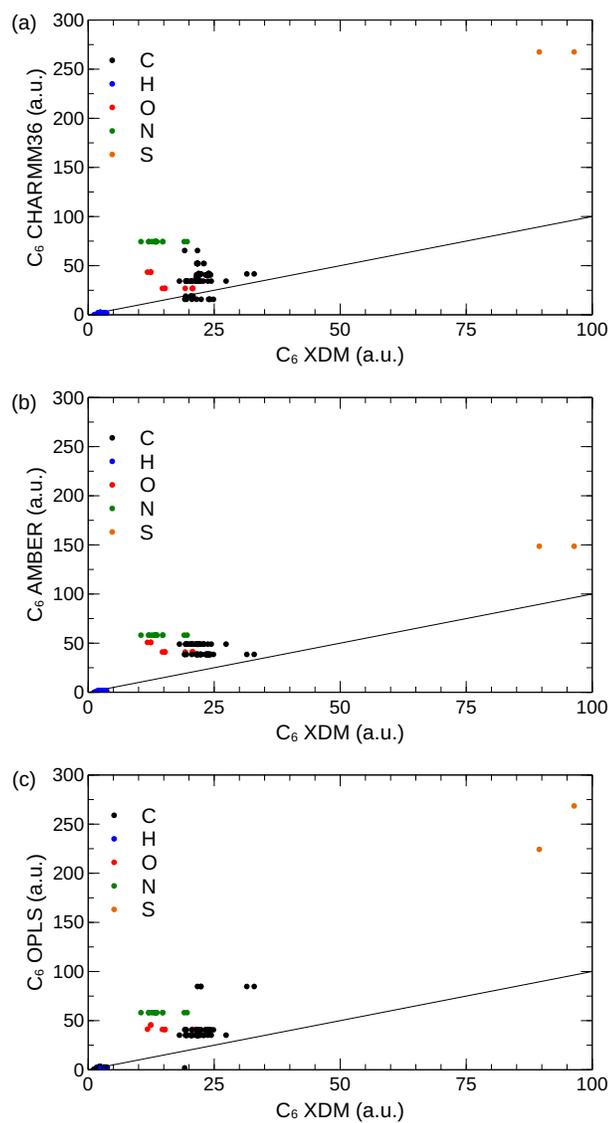


Figure 3.3: The atomic C_6 coefficients of the amino acid side chain models. The gray diagonal line indicates where the force field would match the XDM value.

respectively. This overestimation is partially countered by the small dispersion coefficients assigned to hydrogen atoms in the force field; the force field C_6 coefficients of hydrogen atoms are generally less than 1 a.u., but XDM predicts these values to lie in the 2–3 a.u. range. Nevertheless, the carbon atoms in all residues tend to have dispersion coefficients that are significantly larger than the XDM values.

Hydrogen-Bonding Atom Types

Atoms that participate in hydrogen-bonding interactions tend to have dispersion coefficients that are much larger than the XDM value. This is pronounced in the CHARMM36 force field, where the dispersion coefficients of all N atoms are assigned values of 74.5 a.u., while the XDM values range between 10.5 a.u. to 19.6 a.u. Similarly, the oxygen atoms of amides have values of 26.9 a.u., while the XDM value is only 15.1 a.u. Hydroxyl oxygens have dispersion coefficients of 43.4 a.u., but the XDM value is only 12.4 a.u. The force field atomic dispersion coefficients of these oxygen and nitrogen atoms are even larger than those of the carbon atoms, although physically, these atoms are significantly less polarizable than carbon atoms, so they should have weaker dispersion interactions. This occurs in all three force fields, but to different degrees.

One possible explanation for this trend is the empirical procedure used to parameterize these coefficients, where the Lennard-Jones parameters were adjusted to predict the liquid properties. In some cases, the hydrogen atom is assigned no Lennard-Jones interaction term at all, so the dispersion interactions stemming from the hydrogen atoms must be effectively absorbed into the dispersion-interaction term of their parent atom [33]. Further, because these models neglect induced polarization and charge transfer, the ϵ and σ terms are heavily adjusted to compensate for the missing associative interactions. This is an unattractive solution, because it could have unintended

consequences, such as making these groups spuriously soluble in high-dispersion environments (e.g., the interior of lipid bilayers). To resolve this issue, it would likely be necessary to introduce additional terms to describe the unique electrostatic and repulsive components of hydrogen bonding rigorously. Including induced polarization alone does not appear to resolve this issue because the parameters in the Drude polarizable force field also show this trend [55]. A recent reparameterization of the non-bonded parameters of the GRONingen MOlecular Simulation (GROMOS) parameter set for small organic molecules (2016H66) yielded atomic C_6 dispersion coefficients for nitrogen atoms in the 40–42 a.u. range, indicating that effective force fields can employ significantly smaller dispersion coefficients for nitrogen atoms [116].

Sulfur

The parameters for sulfur atom types found in cysteine and methionine vary significantly between the force fields. The dispersion coefficient for sulfur atom types in the CHARMM36 force field is 267.5 a.u., while the XDM values are 96.4 a.u. and 89.5 a.u., respectively. Similarly, the OPLS-AA force field assigns values of 268.5 a.u. and 224.3 a.u. to C_6 coefficients of the Cys and Met sulfur atoms, respectively. In contrast, the dispersion coefficient assigned to these atom types in the Amber ff14sb force field is only 149 a.u.

3.4.3 Side-Chain Hydration Energies

Although it is apparent that the force fields have dispersion coefficients that are higher than the XDM values, the effect of this difference on the calculated properties is less apparent. To explore this, we calculated the absolute Gibbs energy of hydration of the side-chain models in an explicit solvent using the Weeks–Chandler–Andersen (WCA) decomposition scheme of Deng and Roux [103]. This allows the hydration

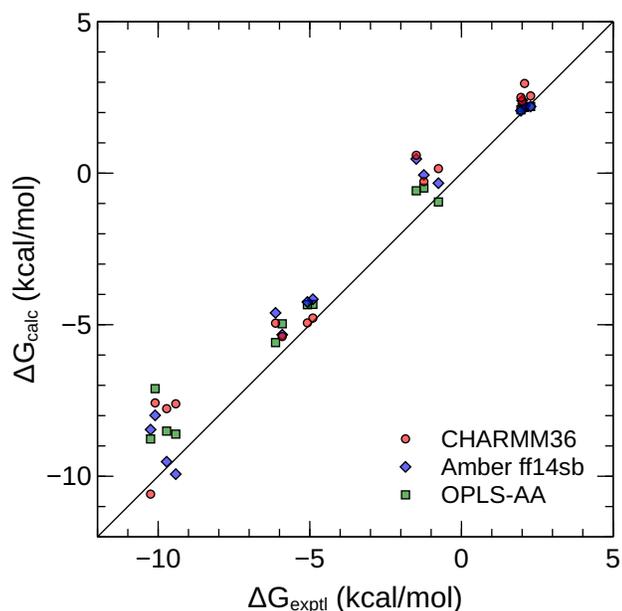


Figure 3.4: Correlation between experimental and calculated hydration energies of the side-chain models for the neutral amino acids.

energy to be divided into electrostatic, dispersion, and repulsive components.³ The dispersion component is of particular interest here, as this term results from the interactions between the solute and the solvent through the dispersion component of the Lennard-Jones interactions of the force fields. The correlation between the calculated and experimental hydration energies is presented in Figure 3.4. A table of the hydration energies and their decomposition into electrostatic, dispersion, and repulsive components is included in the Appendix in Tables B.1, B.2, and B.3.

In general, the agreement between the side-chain hydration energies and the experimental values is reasonably good for all three force fields, which is consistent with previous reports [103; 117; 118; 119; 120; 121]. All three force fields exhibit a small bias to underestimate the solubility of the side-chains and the performance of the

³This is an inherently path-dependent procedure, so other decomposition schemes will assign different values to the terms of the hydration energy.

CHARMM36 force field is incrementally better on the whole. The dispersion component of the hydration energy is generally larger for the CHARMM36 force field than for the Amber ff14sb or OPLS-AA models. Significantly, all three force fields predict reasonably-accurate hydration energies for the aliphatic side-chains (e.g., Ala, Val, Ile, Leu), where the hydration energy is almost exclusively dependent on the balance between the dispersion and repulsive components. This suggests that, although the force field C_6 dispersion coefficients are larger than the XDM values, the total solute-water dispersion interactions are probably of approximately the correct strength. The origin and potential solution of this apparent paradox is discussed in Section 3.4.5.

Side Chain Hydration Energy Prediction

The solvation energy of a hard sphere of radius R that experiences C_6 -dispersion interactions with a surrounding solvent can be estimated analytically to be⁴

$$\Delta G_{\text{disp}} = -4\pi\rho_w \frac{C_{6,\text{w-mol}}}{3R^3}. \quad (3.3)$$

The $C_{6,\text{mol}}$ terms are for like pairs of side chains, so the dispersion coefficient for the interaction between a water molecule and the side chain would be $C_{6,\text{w-mol}} = \sqrt{C_{6,\text{mol}}C_{6,\text{w}}}$. ρ_w is the number density of liquid water.

This suggests a linear relationship between the square root of the molecular C_6 coefficient and the calculated dispersion component of the hydration energy,

$$\Delta G_{\text{disp}} \approx a\sqrt{C_{6,\text{mol}}} + b. \quad (3.4)$$

This linear relationship is apparent in Figure 3.5(a), which correlates the dispersion component of the hydration energy calculated using free-energy simulations with the

⁴There are also more complex numerical methods for non-spherical structures. See Ref. 122.

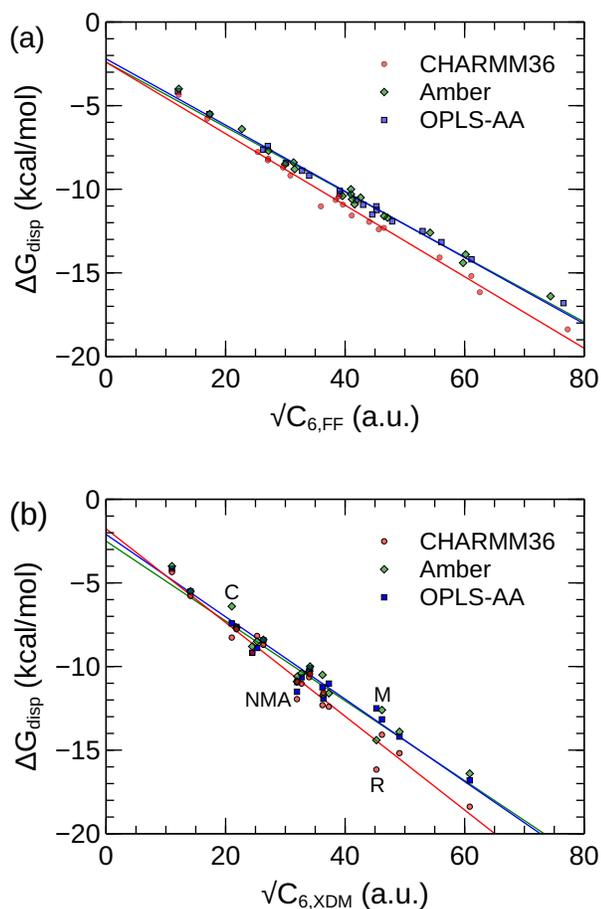


Figure 3.5: (a) Correlation between the force field C_6 molecular dispersion coefficients and the dispersion component of the hydration energy calculated from the free-energy simulations. A strong correlation is observed for all three force fields. (b) Correlation between the XDM C_6 molecular dispersion coefficients and the component of the hydration energy calculated using the free-energy simulations. This correlation illustrates where the force-field dispersion parameters yield a dispersion solvation energy that is inconsistent with the XDM-predicted dispersion coefficient.

molecular C_6 dispersion coefficient of the force fields. This relation is strongly linear for all three force fields, with coefficients of determination of about 0.99.

Section 3.4.1 showed that there is a roughly linear correlation between the XDM and force-field molecular dispersion coefficients (i.e. $\Delta C_{6,FF} \propto C_{6,XDM}$) for the side-chain models. Combining this with Eqn. 3.4, which relates the molecular dispersion coefficient and dispersion component of the hydration energy, the XDM dispersion coefficients can be used to highlight where the calculated dispersion hydration energy is inconsistent with the XDM dispersion coefficient. The correlations between the XDM molecular dispersion coefficient and the calculated dispersion component of the hydration energy are presented in Figure 3.5 (b). Points lying far from the line of regression for each force field indicate that the XDM molecular dispersion coefficient would predict a significantly different dispersion component of the Gibbs energy of hydration for that residue. The CHARMM36 models for Arg and NMA are significant outliers, as are the Amber ff14sb models for Cys and Met.

The hydration energies of the aliphatic residues (Ala, Leu, Ile, and Val) are generally in good agreement with the experimental values. As these residues have minimal electrostatic interactions with the solvent, the attractive component of the hydration energy is almost entirely due to dispersion interactions. Noting that the force field molecular dispersion coefficient is strongly correlated to the hydration energy and this molecular dispersion coefficient is proportional to the XDM dispersion coefficient, Eqn. 3.4 can be used to estimate dispersion hydration free energies directly from the XDM dispersion coefficients. Linear regression was used to fit the dispersion component of the CHARMM36 hydration energies of the aliphatic side chains using the XDM molecular dispersion coefficients ($a = -0.279$, $b = -1.44$, $R^2 = 0.9973$).

This linear relation was used to predict the dispersion contribution to the hydration energy for the remaining side-chain models. This analysis is intended to highlight

Table 3.1: Dispersion component of side-chain hydration energies extrapolated from XDM and those calculated for each force field using REMD-TI. Energies are in kcal/mol.

residue	XDM	CHARMM36	Amber ff14-SB	OPLS-AA
Ile	-10.94	-11.00	-10.30	-10.28
Leu	-10.98	-10.80	-10.00	-10.09
Ala	-4.52	-4.46	-4.00	-4.15
Val	-8.80	-8.98	-8.40	-8.43
Asn	-8.28	-9.47	-8.80	-9.18
Tyr	-15.16	-15.75	-13.90	-14.19
Thr	-7.54	-8.01	-7.70	-7.63
Cys	-7.32	-8.52	-6.40	-7.42
Ser	-5.39	-5.94	-5.50	-5.51
Met	-11.56	-12.75	-10.50	-11.24
Trp	-18.45	-19.10	-16.40	-16.81
Hid	-11.61	-11.95	-11.70	-11.92
Gln	-10.39	-11.24	-10.60	-10.93
Phe	-14.34	-14.60	-12.60	-13.16
Asp	-8.51	-8.42	-8.50	-8.89
Lys	-11.86	-12.83	-11.60	-11.02
Arg	-14.09	-16.75	-14.40	-12.50
Glu	-10.58	-11.36	-10.40	-10.64
NMA	-10.37	-12.36	-10.90	-11.51

where the force field dispersion component of the hydration energy is inconsistent with the first-principles, XDM prediction. The predicted values are compared with those calculated using free energy simulation in Table 3.1.

Generally, the dispersion components of the hydration energies calculated using the free energy simulations are within 1 kcal/mol of the XDM-predicted values. There are some systematic deviations, such as the relatively small dispersion component for the aromatic residues (Phe, Trp, and Tyr) for the Amber ff14sb force field. For the CHARMM36 force field, the dispersion component of the hydration energies of sulfur-containing and nitrogen-containing residues show the greatest deviations, which are discussed in the following sections.

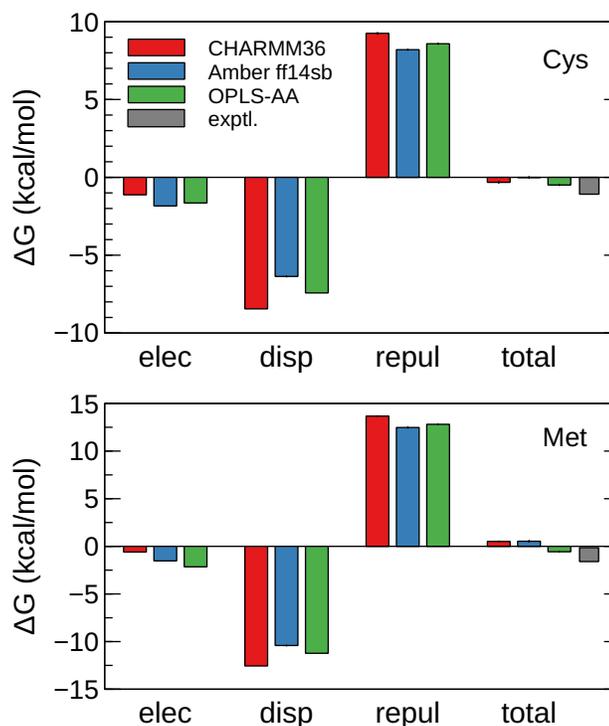


Figure 3.6: Components of the hydration energy of the Cys and Met side chain models. The CHARMM36 force field has the strongest dispersion interactions, but this is countered by a larger repulsive component.

Sulfur Containing Residues

The sulfur-containing side chains, Cys and Met, show a large variation in hydration energies (Figure 3.6). The difference in the dispersion component of the hydration energy is particularly significant; for the Cys side chain, this component ranges from -8.5 kcal/mol for the CHARMM36 force field to -6.4 kcal/mol for the Amber ff14sb force field. The OPLS-AA force field value of -7.4 kcal/mol is closest to the XDM-predicted value of -7.3 kcal/mol. This variation in dispersion energy stems from the atomic dispersion coefficients of the sulfur atoms (see Section 3.4.2); the Amber ff14sb sulfur C_6 coefficients are less than half the OPLS-AA and CHARMM36 values. These disparate values stem from the variation in the Lennard-Jones parameters for sulfur

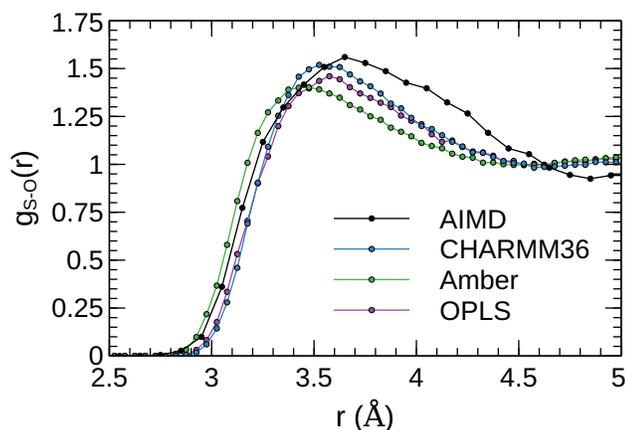


Figure 3.7: S–O radial distribution functions for methyl thiol (model for the Cys side chain) in an aqueous solution. The Amber ff14sb model underestimates the radius predicted by the AIMD simulation while the OPLS-AA and CHARMM36 models overestimate this difference.

Table 3.2: Sulfur–oxygen radii (r_{S-O}) calculated from the methylthiol radial distribution functions, the force field Lennard-Jones parameters (i.e., σ and ε), and the dispersion component of the methylthiol solvation energy (ΔG_{disp}).

Method	r_{S-O} (Å)	σ (Å)	ε (kcal/mol)	ΔG_{disp} (kcal/mol)
CHARMM36	3.25	3.56	-0.45	-8.27
Amber ff14sb	3.17	3.56	-0.25	-6.40
OPLS-AA	3.26	3.6	-0.425	-7.42
AIMD	3.21			

atoms in the force fields (Table 3.2).

The hydration energy of the Cys side chain is underestimated by all three models, although there is a significant variation in the components. The XDM-predicted dispersion component of the hydration energy is approximately -8 kcal/mol, most similar to value predicted by the CHARMM36 force field. The repulsive energy predicted by the CHARMM36 and OPLS-AA models are higher than those of the Amber ff14sb model. Analysis of the S–O(H_2) radial distribution function of aqueous methylthiol (Figure 3.7) shows that the CHARMM36 and OPLS-AA models yield an atomic radius (i.e., $g(r_{S-O}) = 1$) for sulfur that is approximately 0.05 Å larger than the *ab initio* value, suggesting that the repulsive component is too large. The same trend is

repeated for the methionine side chain. This analysis suggests that the parameters for the sulfur-containing residues could be improved in all three force fields.

3.4.4 Nitrogen-Containing Residues

For the CHARMM36 force field, residues containing one or more nitrogen atoms (e.g., Lys, Arg, NMA, Asn, and Gln) tend to have larger water–side-chain dispersion interactions than predicted from the XDM results. For each of these molecules, the XDM-predicted dispersion component of the hydration energy is 1–3 kcal/mol smaller than the value calculated using the CHARMM36 free-energy simulations. The spuriously large contribution of nitrogen atoms to the non-polar component of the hydration energy has also been noted by Mobley et al. [123]. This is consistent with the large atomic dispersion coefficients assigned to nitrogen atoms in the CHARMM36 force field (see Section 3.4.2), which results in a spurious increase in the strength of water–side-chain dispersion interactions.

Arginine is an important example because the dispersion interactions between arginine and the interior of lipid bilayers are significant in the gating of ion channels [124; 125] and arginine residues are often critical to binding nucleic acids [126]. As no experimental hydration energy of the arginine side chain is available, the XDM prediction is particularly informative. The dispersion component of the hydration energy calculated using the CHARMM36 force field is significantly larger than the XDM-predicted value (-16.75 kcal/mol vs -14.09 kcal/mol, respectively). The XDM analysis predicts that the Amber ff14-SB parameters are more consistent (-14.40 kcal/mol).

3.4.5 Improving the Description of Dispersion in Force Fields

The neglect of higher-order dispersion terms in the Lennard-Jones equation may explain why the force fields have high C_6 dispersion coefficients but yield reasonably

Table 3.3: The dispersion component of the hydration energy of methane calculated using the free-energy simulations (CHARMM36 alanine side-chain parameters) and those calculated using Eqn. 3.5 using the XDM C_6 , C_8 , and C_{10} coefficients.

method	n	ΔG
simulation		-4.4
analytical	6	-2.5
	8	-1.0
	10	-0.5
	total	-4.0

accurate hydration energies. Quantum chemical calculations have found that C_8/r^8 and C_{10}/r^{10} terms account for approximately 30% of the dispersion interactions in crystals [77]. When these terms are neglected, the empirical parameterization process used to define Lennard-Jones parameters will increase the strength of the C_6/r^6 interactions to compensate.

To test this possibility, we calculated the XDM C_6 , C_8 , and C_{10} dispersion coefficients for methane and water. These data can be used to estimate the dispersion component of the hydration energy using the expression,

$$\Delta G_{\text{disp}} = -4\pi\rho_w \left[\frac{C_{6,w-\text{mol}}}{3R^3} + \frac{C_{8,w-\text{mol}}}{5R^5} + \frac{C_{10,w-\text{mol}}}{7R^7} \right], \quad (3.5)$$

where ρ_w is the number density of water, $C_{n,w-\text{mol}}$ are the water-solute dispersion coefficients calculated using XDM, and R is the solute-water radius (taken from the arithmetic mean of methane and water Lennard-Jones radii, σ). The coefficients are calculated from the XDM-calculated molecular dispersion coefficients of the individual molecules, combined using the relation $C_{n,w-\text{mol}} = \sqrt{C_{n,w}C_{n,\text{mol}}}$.

Based on this analytical expression, the C_6 dispersion interaction accounts for the largest portion of the hydration energy, but the C_8 and C_{10} terms contribute 25% and 12% of the energy, respectively (Table 3.3). The total interaction calculated from all the dispersion terms is -4.0 kcal/mol, which is close to the dispersion component of

the Gibbs energy of hydration of methane (e.g., Ala) calculated with the CHARMM36 force field. Previous analysis by Floris et al. concluded that C_8 and C_{10} interactions account for 12% of the dispersion component of the hydration energy of methane [127]. This supports the hypothesis that the dispersion interactions stemming from higher-order terms are included effectively in the C_6 coefficients of the Lennard-Jones potential.

There are several significant drawbacks to the practice of including higher-order dispersion interactions through large “effective” C_6 coefficients. The distance dependence of this dispersion interaction will be spuriously strong at long range, but weak at short range. More generally, non-physical descriptions of dispersion interactions limit the accuracy of calculations where solutes move between environments, such as in membrane permeation and protein–ligand binding.

Replacing the Lennard-Jones potential with a non-bonded potential that explicitly includes higher-order dispersion terms would resolve this issue in a more rigorous way. Non-bonded potentials that include higher-order dispersion terms have been used by chemical physicists for decades and there was discussion of including them in early biomolecular force fields [78]. For example, in 1938, Buckingham devised an equation of state for gaseous helium, neon, and argon using a potential that included C_8 dispersion [12]:

$$E_{nb}(r_{ij}) = A \cdot \exp(-b \cdot r) - \frac{C_{6,ij}}{r_{ij}^6} - \frac{C_{8,ij}}{r_{ij}^8}. \quad (3.6)$$

The primary challenge associated with the inclusion of higher order dispersion terms is that two to three dispersion coefficients would have to be defined for each atom type, although methods like XDM could provide reasonable estimates.

Separately, many of the anomalously large atomic dispersion coefficients are those

in hydrogen-bonding groups (e.g., amides and hydroxyls). Adjustments to the Lennard-Jones parameters of these terms to capture the strength of hydrogen bonding despite the neglect of induced polarization and charge transfer may be the origin of this effect. The adoption of force fields that describe induced polarization explicitly [128; 129], or include explicit terms to represent hydrogen bonds [130; 131; 132; 133], may resolve this issue.

3.4.6 Evaluation of Nucleic Acid Force Fields

To test if trends in force-field dispersion coefficients identified for proteins also exist in nucleic-acid force fields, the dispersion coefficients of the nucleobases from the CHARMM36 [134] and Amber OL15 [135] force fields were compared to the XDM values. These models overestimate the molecular dispersion coefficients to an even larger degree, with dispersion coefficients that are more than 200% larger than the XDM values. A major cause of these large force field dispersion coefficients is the large atomic dispersion coefficient for the nucleobase nitrogen atoms; the CHARMM and Amber nucleic-acid force fields have dispersion coefficients of 74 a.u. and 58 a.u., respectively, for the nucleobase nitrogens, while XDM C_6 coefficients for these atoms range from 13–19 a.u. The propensity of the force fields to attribute stronger dispersion interactions should favor states where the nucleobases are in close contact with each other (e.g., in the base-paired state vs a solvent-exposed state), although, in reality, base-pairing is predominantly an electrostatic interaction [136]. In the absence of experimental nucleobase hydration energies, we cannot perform the same analysis on the nucleic acid force fields as we did on the protein force fields. The relationships between the force field dispersion coefficients and the XDM dispersion coefficients are very similar to those of the protein force fields, so nucleic acid force fields are expected to demonstrate similar trends in the strength of dispersion interactions as the

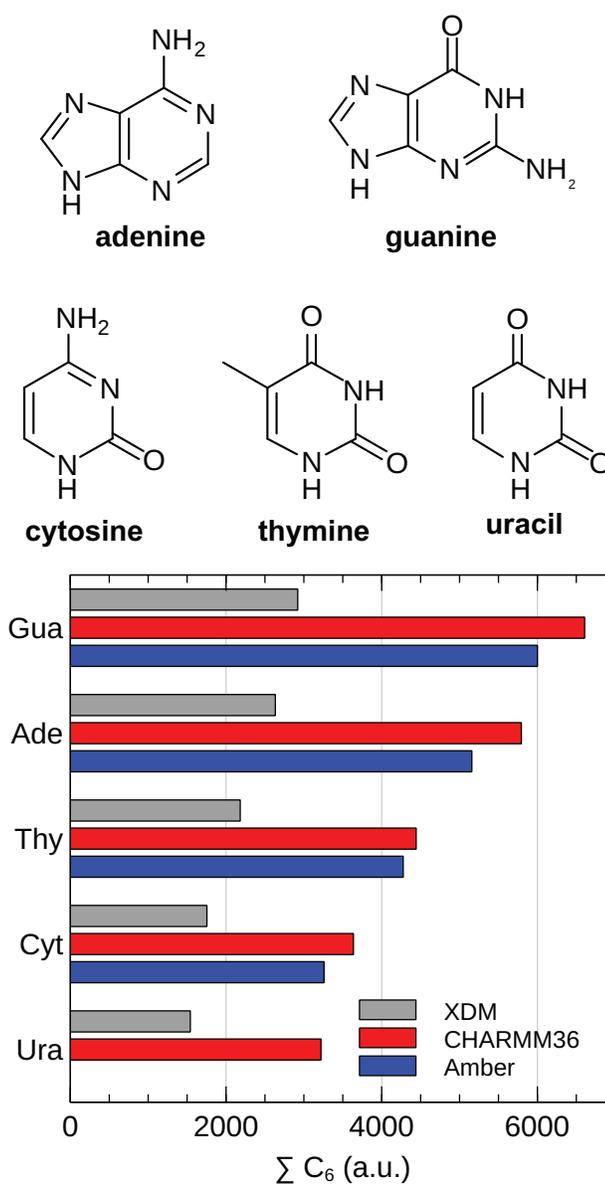


Figure 3.8: Molecular dispersion coefficients for nucleobases calculated using XDM and the CHARMM36 and Amber ff14sb force fields.

nitrogenous amino acids.

3.5 Conclusions

The XDM model was used to evaluate the dispersion coefficients in the CHARMM36, Amber ff14sb, and OPLS-AA force fields for proteins. In keeping with prior reports, we find that the C_6 dispersion coefficients of these force fields are systematically larger than the QM-based XDM model. The Amber OL15 and CHARMM nucleic acid force fields also showed this trend. This trend results from large atomic dispersion coefficients, particularly for hydrogen-bonding atom types. Interestingly, not all water models exhibit this trend; for example, the C_6 dispersion coefficient of the popular TIP3P water model is close to the XDM value.

Despite these large dispersion coefficients, hydration energies of the amino acid side chains calculated using the force fields are generally in good agreement with the experimental values. This is true even for aliphatic residues, where dispersion is the dominant attractive intermolecular interaction between the solvent and the solution. This suggests that the force-field parameters capture the correct strength of water–side-chain dispersion interactions, even though the C_6 dispersion coefficients are systematically larger than the XDM values. Internally consistent dispersion components of the solvation energy can also be estimated by this analysis, and the CHARMM force field was found to define anomalously strong dispersion interactions for the Cys, Met, Arg, and amide residues.

Analysis of the methane–water solvation energy using an analytical expression shows that higher order terms (i.e., C_8 and C_{10}) account for a significant fraction of the dispersion energy. This suggests that the C_6 coefficients in contemporary force fields are “effective” dispersion coefficients that capture neglected associative interactions,

like higher-order dispersion, in addition to dispersion strictly due to instantaneous-dipole-induced-dipole interactions. Although this is clearly effective for describing some simulated properties of proteins, this practice may limit the accuracy of molecular simulations, particularly for processes where there is a net change in the total strength of dispersion interactions. Explicitly including higher-order dispersion terms in molecular-mechanical force fields is a potential solution to this problem.

Chapter 4

Development of a Water Model with an Improved Non-bonded Potential

4.1 Abstract

A molecular mechanical model for liquid water is developed that replaces the Lennard-Jones potential with a revised, physically-motivated potential. The model has three-atomic sites and a virtual site located on the \angle HOH bisector (i.e., a TIP4P-type model). Dispersion interactions are represented by both C_6/r^6 and C_8/r^8 terms. This higher order C_8 dispersion term has been neglected by most force fields. This improved potential is implemented in OpenMM. Using this implementation, the ForceBalance code was used to define parameters that optimally reproduce the experimental physical properties of liquid water. The resulting model is in good agreement with the experimental density, dielectric constant, enthalpy of vaporization, thermal compressibility, isothermal expansion coefficient, diffusion coefficient, and radial distribution

function. This improved non-bonded potential could provide the basis to develop improved force fields that treat repulsion and dispersion interactions more rigorously.

4.2 Introduction

Molecular mechanical force fields underlie materials and biomolecular simulations. These models must effectively capture the significant intermolecular interactions present in a system using computationally-efficient functions. Although these interactions all originate from complex electron–electron, proton–proton, and electron–proton Coulombic interactions, they can be effectively simplified into pairwise electrostatic, London dispersion, and Pauli repulsion interactions.

In most popular molecular mechanical models, electrostatic interactions are described by Coulombic interactions between a set point charges (q) at atomic centers or positions defined with respect to those atomic centers. Some more elaborate models extend this to include the effects of induced polarization and charge transfer, but force field developers have been able to identify static charges that capture these complex electrostatic interactions in an effective way.

Dispersion interactions are a ubiquitous, attractive intermolecular force arising through interaction between instantaneous electric moments in neighbouring atoms. To a reasonable approximation, this interaction can be approximated as a pairwise sum,

$$\mathcal{V}_{disp}(r) = -\frac{C_6}{r^6} - \frac{C_8}{r^8} - \frac{C_{10}}{r^{10}} \quad (4.1)$$

Here, C_n is a coefficient that depends on the pair of interacting atoms. The C_6 term is the strongest and longest-range term, although quantum chemical analysis

and equations of state have found that C_8 and C_{10} terms can yield significant intermolecular interactions. Terms higher than C_{10} are typically insignificant.

Pauli repulsion originates from the overlap of electron density clouds of atoms at close range. As the electron density of atoms follows an exponential dependence, the interaction potential can be described accurately as an exponential decay (i.e., $A \cdot \exp(-br)$), but repulsive polynomial terms are often used instead (e.g., A/r^{12}). In most popular molecular mechanical models, the dispersion and repulsive terms are represented using the Lennard-Jones 12-6 potential [7],

$$\mathcal{V}_{nb}(r) = \frac{A_{LJ}}{r^{12}} - \frac{C_6}{r^6}. \quad (4.2)$$

In this potential, the A/r^{12} term is intended to represent Pauli repulsion, while the $-C_6/r^6$ term represents London dispersion interactions. Higher order dispersion terms are neglected.

These models require the definition of atomic charges and the A and C_6 Lennard-Jones parameters. Generally, this is performed by fitting the parameters such that the predicted properties reproduce the physical properties of the liquid to a reasonable degree of accuracy. As these intermolecular potentials and simulation algorithms are inexact, the parameters may take on “effective” values that result in correct predictions of the targeted physical properties, but may be inconsistent with the molecular origin of these properties.

The correct strength of dispersion interactions has been a contentious subject in biomolecular simulation [47; 48; 49; 50]. In Chapter 2 and 3, force field dispersion parameters were compared to *ab initio* values calculated using the eXchange-hole Dipole Moment (XDM) model. This analysis showed that molecular C_6 coefficients vary widely between force fields, but they were systematically higher than the *ab initio* values. This trend was attributed to the neglect of higher order dispersion terms from

the Lennard-Jones potential.

There are several alternatives to the Lennard-Jones potential that could resolve some of these issues. Buckingham proposed an intermolecular function to describe the interactions of noble gases [12],

$$\mathcal{V}_{nb}(r) = A \cdot \exp(-b \cdot r) - \frac{C_6}{r^6} - \frac{C_8}{r^8}. \quad (4.3)$$

Here, the repulsive interaction is described by the $A \cdot \exp(-b \cdot r)$ term, where the A and b parameters define the strength of the repulsion. Dispersion interactions are represented by the C_6/r^6 and C_8/r^8 terms. This potential has several advantages over the Lennard-Jones potential. The exponential term is a more realistic description of Pauli repulsion than the polynomial A/r^{12} term, which is advantageous in simulations where strong attractive interactions, high temperatures, or high pressures result in frequent repulsive contacts. Explicit inclusion of the C_8/r^8 allow the dispersion interaction to be more-realistically described as a combination of the longer-range C_6/r^6 term with a shorter range C_8/r^8 term, instead of the current practice where all dispersion interactions are effectively included in the C_6/r^6 term.

One issue with the basic form of the Buckingham potential is that the potential becomes infinitely negative as $r \rightarrow 0$. This is a consequence of the exponential term being finite at $r = 0$, but the dispersion terms becoming infinitely negative. As a result, Tang and Toennis proposed that the dispersion terms be damped by an incomplete gamma function [137],

$$f_{damp,n}(r) = 1 - \exp(-\zeta r) \sum_{k=0}^n \frac{(\zeta r)^k}{k!} \quad (4.4)$$

$$\mathcal{V}_{nb}(r) = A \cdot \exp(-b \cdot r) - f_{damp,6}(r) \frac{C_6}{r^6} - f_{damp,8}(r) \frac{C_8}{r^8}. \quad (4.5)$$

where ζ is a parameter that corresponds to the strength of the damping.

This non-bonded potential is much more amenable for calculating the long range component of the C_6 interaction energy using lattice summation methods. As the C_8/r^8 term is shorter range than the C_6/r^6 term, a non-bonded cutoff can be applied to this term without neglecting a large component of the dispersion energy.

Although this type of intermolecular potential has found use in chemical physics, condensed-matter molecular simulations still overwhelmingly depend on the Lennard-Jones potential to describe these interactions. The foremost barrier to adopting these non-bonded potentials is that it would be necessary to define a complete set of parameters to describe the interaction of each pair of atoms. Determination of optimal parameters has generally been a slow process, where a large set of parameter combinations must be tested. In 2014, Leeping, Martinez, and Pande released the ForceBalance code [93], which allows optimal parameters for force fields to be determined using a gradient-directed optimization of the parameters.

The ForceBalance code was successfully used to develop new parameters for the TIP3P and TIP4P water models that best described the physical properties of water. A target function was defined based on the enthalpy of vaporization, density, isothermal compressibility, heat capacity, dielectric constant, the thermal expansion coefficient of liquid water at standard conditions (298.15 K, 101.325 kPa).

$$\begin{aligned} \delta(A, b, C_6, C_8, q_{LP}, l_{LP}) = & b_1 \cdot (\rho_{ref} - \rho_{calc})^2 + \\ & b_2 \cdot (\Delta H_{vap,ref} - \Delta H_{vap,calc})^2 + \\ & b_3 \cdot (\varepsilon_{0,ref} - \varepsilon_{0,calc})^2 \end{aligned} \quad (4.6)$$

The models derived from this optimization, termed TIP3P-FB and TIP4P-FB, provided a significantly improved description of the physical, transport, and structure properties of water. This code provides a viable path to determining appropriate

parameters for simulations using the potential defined in Eqn. 4.5. In this work, we present the development of a new model for liquid water using ForceBalance. This water model is intended to serve as the cornerstone of a new force field based on this improved potential.

4.3 Computational Methods

The simulation cell was a cubic cell containing 215 water molecules (18.64 Å). Electrostatic interactions were calculated using the Particle Mesh Ewald method. The C_6 term of the dispersion interaction was calculated using a lattice-summation method to capture the long-range component.

All simulations were performed using a modified version of OpenMM 7.2 [138]. Both vectorized and Compute Unified Device Architecture (CUDA)-platform Graphical Processing Unit (GPU)-accelerated variants were implemented to enable high-performance on modern computing architectures.

The ForceBalance method allows the force field optimization process to be performed efficiently. Parameters can be determined efficiently using a gradient-directed optimization of the target function.

The enthalpy of vaporization was calculated from the average potential energy of the simulation,

$$\Delta H_{vap} = RT - \langle \mathcal{V} \rangle_{liq} / N_{mol} \quad (4.7)$$

where N_{mol} is the number of molecules in the simulation.

The thermal expansion coefficient (α_P) was calculated from,

$$\alpha_P = \frac{[\langle V \cdot \mathcal{V} \rangle - \langle U \cdot \mathcal{V} \rangle + P (\langle V^2 \rangle - \langle V \rangle^2)]}{k_B T^2 \langle V \rangle} \quad (4.8)$$

Table 4.1: Optimal parameters for a 4-point water model with the Buckingham 6-8 potential.

parameter	value
A (kJ mol ⁻¹)	1.25×10^6
b (nm ⁻¹)	41.15
C_6 (kJ mol ⁻¹ nm ⁶)	2.57×10^{-3}
C_8 (kJ mol ⁻¹ nm ⁸)	3.12×10^{-5}
q_O (e)	-1.03

where V is the volume of the system.

The dielectric constant was calculated using the relation

$$\varepsilon_0 = 1 + \frac{4\pi(\langle M^2 \rangle - \langle M \rangle^2)}{3\langle V \rangle k_B T} \quad (4.9)$$

where M is the net dipole moment of the simulation cell and V is the volume of the cell.

The diffusion coefficient was calculated from an NVE trajectory of the system using the Einstein relation with the correction for finite size effects by Yeh and Hummer [139],

$$D = \frac{1}{6t} \langle |r_i(t) - r_i(0)|^2 \rangle + 2.837297 \frac{k_B T}{6\pi\eta L}. \quad (4.10)$$

4.4 Results and Discussion

This method was used to assign the parameters for the Buckingham-type potential with 6th and 8th order dispersion (B68) water model. The optimization proceeded for 34 iterations until converging to an optimal set of parameters. These parameters are presented in Table 4.1.

This model is in generally good agreement for the targeted physical properties (Table 4.2). The dielectric constants are considerably improved over the TIP3P water

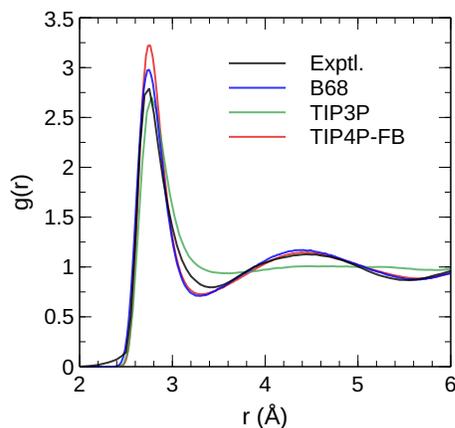


Figure 4.1: Radial distribution functions of TIP3P, TIP4P-FB, and B68 water models.

Table 4.2: Properties of liquid water (298 K and 101.325 kPa) predicted by the optimized B68 water model. Density, dielectric constant, enthalpy of vaporization, thermal expansion coefficient, isothermal compressibility, and heat capacity predicted by TIP3P, TIP4P-FB, and B68 as well as the experimental values are shown.

Property	TIP3P	TIP4P-FB	B68	Exptl.
ρ (kg m ⁻³)	983.627	995.779	997.100	997.045
ϵ_0	95.809	77.288	81.040	78.409
ΔH_{vap} (kJ · mol ⁻¹)	37.108	45.210	44.293	43.989
α (10 ⁻⁴ K ⁻¹)	9.049	2.442	2.609	2.572
κ (10 ⁻⁶ bar ⁻⁶)	57.820	45.191	43.042	45.247
C_p (10 ⁻⁶ cal · mol ⁻¹ · K ⁻¹)	16.629	18.961	17.182	18.002

model. The O–O radial distribution function of this model is presented in Figure 4.1, along with those for Lennard-Jones based TIP3P and TIP4P-FB models and an experimental profile determined by X-ray scattering experiments. The first peak of the B68 radial distribution function is less steep than the TIP3P and TIP4P-FB models, which is in better agreement with the experimental data. This is likely due to the softer, exponential repulsive potential in the B68 model. The regions corresponding to the first minimum and second coordination sphere ($r = 2.5 - 5.0$ Å) are similar to the experimental data and the TIP4P-FB model.

The temperature dependence of the dielectric constant, thermal expansion coefficient, compressibility, and density of liquid water calculated using these models is

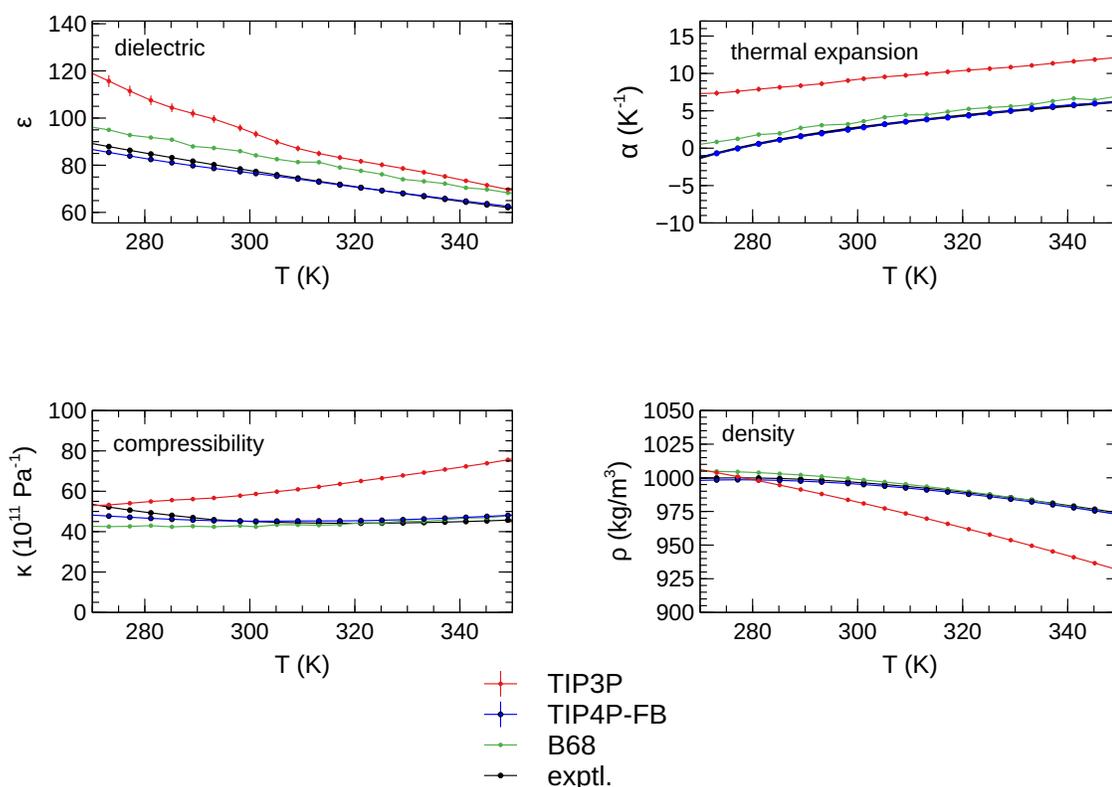


Figure 4.2: Temperature dependence of liquid water properties over the temperature range of 270 K to 350 K calculated using the TIP3P, TIP4P-FB, and B68 water models compared to the experimental values. Data for TIP3P and TIP4P-FB are taken from Ref. [93].

presented in Figure 4.2. Notably, the temperature dependence of the TIP3P water model is poor for all four properties. The B68 model tends to overestimate the dielectric constant over the full temperature range and underestimates the compressibility of water at low temperatures, but otherwise its performance is comparable to the TIP4P-FB water model for most properties.

To assess the additional computational cost of the more complex B68 potential in comparison to traditional Lennard-Jones models, benchmark simulations were performed. Vectorized and Graphical Processing Unit (GPU) accelerated versions of the B68 potential have been implemented in OpenMM 7.2. The benchmark simulations

were performed in NVE conditions using a $64 \text{ \AA} \times 64 \text{ \AA} \times 64 \text{ \AA}$ box containing 8673 molecules. For the GPU simulations, two NVIDIA Tesla P100 GPUs were used, and the CPU simulations were done using 12 Intel[®] Xeon[®] E5-2667 CPUs in multi-thread mode. For comparison, simulations were also performed using a simulation cell of the same dimensions with the TIP4P-FB water model, which employs a conventional Lennard-Jones non-bonded potential. The increased cost of the B68 non-bonded potential is modest, with the speed of the simulations decreasing by 21%. The B86 simulations are 12.8 times faster when performed on the GPU, although the Lennard-Jones code runs 14 times faster on the GPU. This can be attributed to the calculations of exponential terms in the repulsive term and damping functions in the B68 potential, which are computationally-slower operations than the purely polynomial terms in the Lennard-Jones potential.

4.5 Conclusions

Using the ForceBalance code, a molecular mechanical model for liquid water was developed. This model differs from established water models because it replaces the Lennard-Jones non-bonded potential with a more sophisticated potential that describes interatomic Pauli repulsion. The physical properties and radial distribution function are in excellent agreement with experimental data. The GPU-accelerated and vectorized implementations of this potential were incorporated into OpenMM 7.2. Benchmark simulations show this potential is only modestly more computationally-intensive than conventional Lennard-Jones-based potentials.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In Chapters 2 and 3, the XDM model from DFT was used to evaluate the London dispersion coefficients in selected popular general force fields (CGenFF, GAFF, OPLS, and Drude) and popular protein force fields (CHARMM36, OPLS-AA, and AMBERff14sb). The force field C_6 coefficients were determined through the Lennard-Jones parameters. All force fields systematically overestimate atomic and molecular dispersion coefficients relative to XDM.

To evaluate whether these inflated parameters affect the properties calculated from simulations, the hydration energies of protein side chain models were calculated. These hydration energies are in a good agreement with the experimental data, which suggests that despite the overestimation in force field C_6 coefficients, the total water-side-chain dispersion energies are of approximately the correct strength.

QM calculations have shown that the higher order dispersion interactions (i.e., C_8 and C_{10}) accounts for a large portion of the total dispersion energy. The Lennard-Jones potential used in the studied force fields neglect the higher order dispersion

terms. Also, the Lennard-Jones parameters of these force fields were empirically fit such that the models predict the physical properties of bulk liquids accurately. This suggests that the parameterization process has effectively increased the C_6 parameters to compensate for neglected higher order dispersion interactions. This approach limits the accuracy of simulations because higher-order dispersion terms have different distance dependencies, so the dispersion interactions of these models should be shorter-range than the current models that place all dispersion interactions in the C_6 term.

In Chapter 4, a new water model with an improved representation for London dispersion forces, including higher order dispersion terms, was developed. A Buckingham-type non-bonded potential was implemented into OpenMM, replacing the Lennard-Jones potential. The model was then optimized using the ForceBalance code. The optimized model was able to predict the physical properties and radial distribution function of water accurately.

5.2 Future Work

The improved treatment of dispersion has immediate applications to study aspects of chemistry such as high pressure phases of liquids [140] and host-guest systems like methane hydrates [141]. The development of molecular mechanical force fields with improved descriptions of London dispersion forces can be extended to a vast range of molecules and atom types, including organic molecules, proteins, and biomolecules. In the long term, a complete set of parameters for the B68 potential would allow biophysical phenomena like membrane permeation, protein folding, and protein-ligand binding to be simulated with greater accuracy. This could ultimately allow improved predictions of drug activity and protein structure. More accurate force fields also

require improved descriptions of electrostatic interactions like induced polarization and charge transfer. The combination of the B68 non-bonded potential with a charge-on-a-spring [25] or polarizable atomic multipole [142] model would provide a model that describes repulsion, higher-order dispersion, and included polarization rigorously.

To be adopted by the broader simulation community, the code for the new non-bonded potentials, that has already been implemented OpenMM, will have to be transferred into other molecular dynamics software packages, such as GROMACS, NAMD, etc. Also, further performance optimization is required in the current implementation of the Buckingham potential. This will require significant modifications to the code for calculations of non-bonded interactions, but it will also require extensive modifications to other components of these codes, such as the parameter file format specification.

Bibliography

- [1] T. Lazaridis, G. Archontis, and M. Karplus. Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. volume 47 of *Advances in Protein Chemistry*, pages 231–306. Academic Press, 1995.
- [2] A. D. MacKerell and M. Karplus. Importance of attractive van der Waals contribution in empirical energy function models for the heat of vaporization of polar liquids. *J. Phys. Chem.*, 95(26):10559–10560, 1991.
- [3] J. Moellmann and S. Grimme. Importance of London dispersion effects for the packing of molecular crystals: A case study for intramolecular stacking in a bis-thiophene derivative. *Phys. Chem. Chem. Phys.*, 12:8500–8504, 2010.
- [4] F. M. Fowkes. Additivity of intermolecular forces at interfaces. I. determination of the contribution to surface and interfacial tensions of dispersion forces in various liquids. *J. Phys. Chem.*, 67(12):2538–2541, 1963.
- [5] R. A. DiStasio, O. A. von Lilienfeld, and A. Tkatchenko. Collective many-body van der Waals interactions in molecular systems. *Proc. Natl. Acad. Sci. U.S.A.*, 109(37):14791–14795, 2012.

- [6] A. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2nd edition, 2013.
- [7] J. E. Jones. On the determination of molecular fields. II. from the equation of state of a gas. *Proc. R. Soc. London, A*, 106(738):463–477, 1924.
- [8] W. Song, P. J. Rossky, and M. Maroncelli. Modeling alkane + perfluoroalkane interactions using all-atom potentials: Failure of the usual combining rules. *J. Chem. Phys.*, 119(17):9145–9162, 2003.
- [9] C. M. Baker, P. E. M. Lopes, X. Zhu, B. Roux, and A. D. MacKerell Jr. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. *J. Chem. Theory Comput.*, 6(4):1181–1198, 2010.
- [10] S. Riahi and C. N. Rowley. Solvation of hydrogen sulfide in liquid water and at the water–vapor interface using a polarizable force field. *J. Phys. Chem. B*, 118(5):1373–1380, 2014.
- [11] D. Berthelot. Sur le mélange des gaz. *Compt. Rendus*, 126:1703–1706, 1898.
- [12] R. A. Buckingham. The classical equation of state of gaseous helium, neon and argon. *Proc. R. Soc. Lond. A*, 168(933):264–283, 1938.
- [13] H.-J. Böhm and R. Ahlrichs. A study of short-range repulsions. *J. Chem. Phys.*, 77(4):2028–2034, 1982.
- [14] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts. Oxford University Press, 2010.
- [15] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. general method. *J. Chem. Phys.*, 31(2):459–466, 1959.

- [16] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76(1):637–649, 1982.
- [17] L. Verlet. Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98–103, 1967-07.
- [18] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52(2):255–268, 1984.
- [19] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985-03.
- [20] D. J. Evans and B. L. Holian. The Nosé–Hoover thermostat. *J. Chem. Phys.*, 83(8):4069–4074, 1985.
- [21] W. G. Hoover. Constant-pressure equations of motion. *Phys. Rev. A*, 34:2499–2500, 1986-09.
- [22] T. Darden, D. York, and L. Pedersen. Particle Mesh Ewald: An N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.
- [23] W. L. Jorgensen, J. D. Madura, and C. J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.*, 106(22):6638–6646, 1984.
- [24] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

- [25] G. Lamoureux, A. D. MacKerell Jr., and B. Roux. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.*, 119(10):5185–5197, 2003.
- [26] S. Riahi and C. N. Rowley. A Drude polarizable model for liquid hydrogen sulfide. *J. Phys. Chem. B*, 117(17):5222–5229, 2013.
- [27] A. N. S. Adluri, J. N. Murphy, T. Tozer, and C. N. Rowley. Polarizable force field with a sigma-hole for liquid and aqueous bromomethane. *J. Phys. Chem. B*, 119(42):13422–13432, 2015.
- [28] S. Riniker. Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview. *J. Chem. Inf. Model.*, 58:565–578, 2018.
- [29] I. J. Chen, D. Yin, and A. D. MacKerell. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. *J. Comput. Chem.*, 23(2):199–213, 2002.
- [30] I. Adamovic and M. S. Gordon. Dynamic polarizability, dispersion coefficient c_6 and dispersion energy in the effective fragment potential method. *Mol. Phys.*, 103(2-3):379–387, 2005.
- [31] A. Stone and A. Misquitta. Atom–atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.*, 26:193–222, 2007.
- [32] P. Bajaj, A. W. Götz, and F. Paesani. Toward chemical accuracy in the description of ion–water interactions through many-body representations. I. halide–water dimer potential energy surfaces. *J. Chem. Theory Comput.*, 12(6):2698–2705, 2016.

- [33] D. J. Cole, J. Z. Vilseck, J. Tirado-Rives, M. C. Payne, and W. L. Jorgensen. Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *J. Chem. Theory Comput.*, 12(5):2312–2323, 2016.
- [34] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, 11(8):3696–3713, 2015.
- [35] A. D. Becke and E. R. Johnson. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.*, 127(15):154108, 2007.
- [36] A. Otero-de-la Roza and E. R. Johnson. Non-covalent interactions and thermochemistry using XDM-corrected hybrid and range-separated hybrid density functionals. *J. Chem. Phys.*, 138:204109, 2013.
- [37] E. R. Johnson. The exchange-hole dipole moment dispersion model. In A. Otero-de-la Roza and G. A. DiLabio, editors, *Non-covalent Interactions in Quantum Chemistry and Physics*, chapter 5, pages 215–248. Elsevier, 2017.
- [38] L. Salem. The calculation of dispersion forces. *Mol. Phys.*, 3:441–452, 1960.
- [39] A. Dalgarno and W. D. Davison. The calculation of van der Waals interactions. *Adv. At. Mol. Phys.*, 2:1–32, 1966.
- [40] A. Otero-de-la Roza and E. R. Johnson. Many-body dispersion interactions from the exchange-hole dipole moment model. *J. Chem. Phys.*, 138:054103, 2013.
- [41] E. R. Johnson. Dependence of dispersion coefficients on atomic environment. *J. Chem. Phys.*, 135:234109, 2011.

- [42] M. S. Christian, A. Otero-de-la-Roza, and E. R. Johnson. Surface adsorption from the exchange-hole dipole moment dispersion model. *J. Chem. Theory Comput.*, 12:3305–3315, 2016.
- [43] T. Gould, E. R. Johnson, and S. A. Tawfik. Are dispersion corrections accurate outside equilibrium? A case study on benzene. *Beilstein J. Org. Chem.*, 14:1181–1191, 2018.
- [44] J. Dobson. Beyond pairwise additivity in London dispersion interactions. *Int. J. Quantum Chem.*, 114:1157–1161, 2014.
- [45] A. D. Becke and E. R. Johnson. A density-functional model of the dispersion interaction. *J. Chem. Phys.*, 123:154101, 2005.
- [46] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4):281–296, 1991.
- [47] R. M. Levy, L. Y. Zhang, E. Gallicchio, and A. K. Felts. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute–solvent interaction energy. *J. Am. Chem. Soc.*, 125(31):9523–9530, 2003.
- [48] J. Dzubiella, J. M. J. Swanson, and J. A. McCammon. Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. *Phys. Rev. Lett.*, 96:087802, 2006-03.
- [49] C. Tan, Y.-H. Tan, and R. Luo. Implicit nonpolar solvent models. *J. Phys. Chem. B*, 111(42):12263–12274, 2007.

- [50] C. J. Fennell, C. Kehoe, and K. A. Dill. Oil/water transfer is partly driven by molecular shape, not just size. *J. Am. Chem. Soc.*, 132(1):234–240, 2010.
- [51] S. Riahi and C. N. Rowley. Why can hydrogen sulfide permeate cell membranes? *J. Am. Chem. Soc.*, 136(43):15111–15113, 2014.
- [52] E. Awoonor-Williams and C. N. Rowley. Molecular simulation of nonfacilitated membrane permeation. *Biochim. Biophys. Acta – Biomembranes*, 1858(7, Part B):1672–1687, 2016.
- [53] J. Wang, Y. Deng, and B. Roux. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.*, 91(8):2798–2814, 2006.
- [54] Y.-L. Lin, Y. Meng, W. Jiang, and B. Roux. Explaining why Gleevec is a specific and potent inhibitor of Abl Kinase. *Proc. Natl. Acad. Sci. U.S.A.*, 110(5):1664–1669, 2013.
- [55] M. Mohebifar, E. R. Johnson, and C. N. Rowley. Evaluating force-field London dispersion coefficients using the exchange-hole dipole moment model. *J. Chem. Theory Comput.*, 13(12):6146–6157, 2017.
- [56] G. Zeiss and W. J. Meath. The H_2O – H_2O dispersion energy constant and the dispersion of the specific refractivity of dilute water vapour. *Mol. Phys.*, 30(1):161–169, 1975.
- [57] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

- [58] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123(23):234505, 2005.
- [59] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. MacKerell Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.*, 31(4):671–690, 2010.
- [60] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general Amber force field. *J. Comput. Chem*, 25(9):1157–1174, 2004.
- [61] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6665–6670, 2005.
- [62] P. Lopes, B. Roux, and A. D. MacKerell Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: Theory and applications. *Theor. Chem. Acc.*, 124:11–28, 2009.
- [63] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158–6170, 1999.
- [64] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005.
- [65] F. Neese. The ORCA program system. *WIREs Comput. Mol. Sci.*, 2(1):73–78, 2012.

- [66] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- [67] A. L. Hickey and C. N. Rowley. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *J. Phys. Chem. A*, 118(20):3678–3687, 2014.
- [68] F. O. Kannemann and A. D. Becke. Atomic volumes and polarizabilities in density-functional theory. *J. Chem. Phys.*, 136:034109, 2012.
- [69] The postg program is available from <http://schooner.chem.dal.ca>.
- [70] PyPostg. <https://github.com/RowleyGroup/postg-parser>. (accessed March 17, 2017).
- [71] Virtual Chemistry. <http://virtualchemistry.org/>. (accessed March 17, 2017).

- [72] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, 2015.
- [73] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, 1980-10.
- [74] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.
- [75] C. L. Wennberg, T. Murtola, B. Hess, and E. Lindahl. Lennard-Jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *J. Chem. Theory Comput.*, 9(8):3527–3537, 2013.
- [76] E. R. Johnson and A. D. Becke. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.*, 124:174104, 2006.
- [77] A. Otero-de-la Roza and E. R. Johnson. A benchmark for non-covalent interactions in solids. *J. Chem. Phys.*, 137:054103, 2012.
- [78] R. W. Pastor. *Topics in stochastic dynamics of polymers*. PhD thesis, Harvard University, 1984.
- [79] M. R. Kennedy, A. R. McDonald, A. E. DePrince III, M. S. Marshall, R. Podeszwa, and C. D. Sherril. Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *J. Chem. Phys.*, 140:121104, 2014.
- [80] Y. Huang and G. J. Beran. Reliable prediction of three-body intermolecular

- interactions using dispersion-corrected second-order Møller-Plesset perturbation theory. *J. Chem. Phys.*, 143:044113, 2015.
- [81] N. M. Fischer, P. J. van Maaren, J. C. Ditz, A. Yildirim, and D. van der Spoel. Properties of organic liquids when simulated with long-range Lennard-Jones interactions. *J. Chem. Theory Comput.*, 11(7):2938–2944, 2015.
- [82] M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B*, 111(45):13052–13063, 2007.
- [83] D. van der Spoel, P. J. van Maaren, and C. Caleman. GROMACS molecule and liquid database. *Bioinformatics*, 28(5):752, 2012.
- [84] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta*, 44:129–138, 1977.
- [85] P. Bultinck, C. V. Alsenoy, P. W. Ayers, and R. Carbó-Dorca. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.*, 126(14):144111, 2007.
- [86] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and J. Alexander D. MacKerell. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.*, 8(9):3257–3273, 2012.
- [87] G. Lamoureux and B. Roux. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119(6):3025–3039, 2003.

- [88] X. Zhu and A. D. MacKerell Jr. Polarizable empirical force field for sulfur-containing compounds based on the classical Drude oscillator model. *J. Comput. Chem.*, 31(12):2330–2341, 2010.
- [89] C. M. Baker, V. M. Anisimov, and A. D. MacKerell Jr. Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *J. Phys. Chem. B*, 115(3):580–596, 2011.
- [90] D. S. Patel, X. He, and A. D. MacKerell Jr. Polarizable empirical force field for hexopyranose monosaccharides based on the classical Drude oscillator. *J. Phys. Chem. B*, 119(3):637–652, 2015.
- [91] D. Braun, S. Boresch, and O. Steinhauser. Transport and dielectric properties of water and the influence of coarse-graining: Comparing BMW, SPC/E, and TIP3P models. *J. Chem. Phys.*, 140(6):064107, 2014.
- [92] R. B. Best, W. Zheng, and J. Mittal. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.*, 10(11):5113–5124, 2014.
- [93] L.-P. Wang, T. J. Martinez, and V. S. Pande. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.*, 5(11):1885–1891, 2014.
- [94] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.
- [95] H. W. Horn, W. C. Swope, J. W. Pitera, 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. *J. Chem. Phys.*, 120(20):9665–9678, 2004.

- [96] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B*, 119(16):5113–5123, 2015.
- [97] S. Izadi and A. V. Onufriev. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.*, 145(7):074501, 2016.
- [98] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.*, 418(1–3):245–249, 2006.
- [99] W. Yu, P. E. M. Lopes, B. Roux, and A. D. MacKerell Jr. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.*, 138(3):034508, 2013.
- [100] J. Huang and A. D. MacKerell. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.*, 34(25):2135–2145, 2013.
- [101] 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. *J. Phys. Chem. B*, 105(28):6474–6487, 2001.
- [102] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. the atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371, 1993.
- [103] Y. Deng and B. Roux. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy

- simulations with explicit water molecules. *J. Phys. Chem. B*, 108(42):16567–16576, 2004.
- [104] W. Jiang and B. Roux. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.*, 6(9):2559–2565, 2010.
- [105] Y.-L. Lin, A. Aleksandrov, T. Simonson, and B. Roux. An overview of electrostatic free energy computations for solutions and proteins. *J. Chem. Theory Comput.*, 10(7):2690–2709, 2014.
- [106] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.
- [107] S. Jo, T. Kim, V. G. Iyer, and W. Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.*, 29(11):1859–1865, 2008.
- [108] A. M. Buckle, G. Schreiber, and A. R. Fersht. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0 Å resolution. *Biochemistry (Mosc.)*, 33(30):8878–8889, 1994.
- [109] D. Wacker, S. Wang, J. D. McCorvy, R. M. Betz, A. Venkatakrisnan, A. Levit, K. Lansu, Z. L. Schools, T. Che, D. E. Nichols, B. K. Shoichet, R. O. Dror, and B. L. Roth. Crystal structure of an LSD-bound human serotonin receptor. *Cell*, 168(3):377–389.e12, 2017.
- [110] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp. Phys. Comm.*, 167(2):103–128, 2005.

- [111] Y. Zhang and W. Yang. Comment on “generalized gradient approximation made simple”. *Phys. Rev. Lett.*, 80:890, 1998-01.
- [112] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15):154104, 2010.
- [113] A. Bankura, A. Karmakar, V. Carnevale, A. Chandra, and M. L. Klein. Structure, dynamics, and spectral diffusion of water from first-principles molecular dynamics. *J. Phys. Chem. C*, 118(50):29401–29411, 2014.
- [114] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, 1996-07.
- [115] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.
- [116] B. A. C. Horta, P. T. Merz, P. F. J. Fuchs, J. Dolenc, S. Riniker, and P. H. Hünenberger. A GROMOS-compatible force field for small organic molecules in the condensed phase: The 2016h66 parameter set. *J. Chem. Theory Comput.*, 12(8):3825–3850, 2016.
- [117] A. Villa and A. E. Mark. Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.*, 23(5):548–553, 2002.
- [118] J. L. MacCallum and D. P. Tieleman. Calculation of the water–cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.*, 24(15):1930–1935, 2003.

- [119] B. Hess and N. F. A. van der Vegt. Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B*, 110(35):17616–17626, 2006.
- [120] J. Chang, A. M. Lenhoff, and S. I. Sandler. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B*, 111(8):2098–2106, 2007.
- [121] H. Zhang, C. Yin, Y. Jiang, and D. van der Spoel. Force field benchmark of amino acids: I. hydration and diffusion in different water models. *J. Chem. Inf. Model.*, 58(5):1037–1052, 2018.
- [122] P. Labute. The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.*, 29(10):1693–1698, 2008.
- [123] D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.*, 5(2):350–358, 2009.
- [124] S. B. Long, X. Tao, E. B. Campbell, and R. MacKinnon. Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature*, 450:376, 2007-11-15.
- [125] C. S. Schwaiger, S. I. Börjesson, B. Hess, B. Wallner, F. Elinder, and E. Lindahl. The free energy barrier for arginine gating charge translation is altered by mutations in the voltage sensor domain. *PLoS One*, 7(10):1–11, 2012-10.
- [126] M. A. Weiss and N. Narayana. RNA recognition by arginine-rich peptide motifs. *Biopolymers*, 48(2-3):167–180, 1998.

- [127] F. M. Floris, J. Tomasi, and J. L. P. Ahuir. Dispersion and repulsion contributions to the solvation energy: Refinements to a simple computational model in the continuum approximation. *J. Comput. Chem.*, 12(7):784–791, 1991.
- [128] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.*, 9(9):4046–4063, 2013.
- [129] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell Jr. Polarizable force field for peptides and proteins based on the classical Drude oscillator. *J. Chem. Theory Comput.*, 9(12):5430–5449, 2013.
- [130] A. V. Morozov, T. Kortemme, K. Tsemekhman, and D. Baker. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. U.S.A.*, 101(18):6946–6951, 2004.
- [131] J. Huang and M. Meuwly. Explicit hydrogen-bond potentials and their application to NMR scalar couplings in proteins. *J. Chem. Theory Comput.*, 6(2):467–476, 2010.
- [132] H. Choi, H. Kang, and H. Park. New angle-dependent potential energy function for backbone–backbone hydrogen bond in protein–protein interactions. *J. Comput. Chem.*, 31(5):897–903, 2010.
- [133] M. Korth. Empirical hydrogen-bond potential functions – an old hat reconditioned. *ChemPhysChem*, 12(17):3131–3142, 2011.
- [134] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell. Optimization of the CHARMM additive force field for DNA: Improved

- treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.*, 8(1):348–362, 2012.
- [135] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, and P. Jurečka. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.*, 11(12):5723–5736, 2015.
- [136] P. Su and H. Li. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.*, 131(1):014102, 2009.
- [137] K. T. Tang and J. P. Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.*, 80(8):3726–3741, 1984.
- [138] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, 13(7):1–17, 07 2017.
- [139] I.-C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B*, 108(40):15873–15879, 2004.
- [140] S. Fanetti, A. Lapini, M. Pagliai, M. Citroni, M. Di Donato, S. Scandolo, R. Righini, and R. Bini. Structure and dynamics of low-density and high-density liquid water at high pressure. *J. Phys. Chem. Lett.*, 5(1):235–240, 2014. PMID: 26276206.

- [141] C. Moon, P. C. Taylor, and P. M. Rodger. Molecular dynamics study of gas hydrate formation. *J. Am. Chem. Soc.*, 125(16):4706–4707, 2003. PMID: 12696878.
- [142] J. W. Ponder, C. Wu, P. 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. *J. Phys. Chem. B*, 114(8):2549–2564, 2010. PMID: 20136072.
- [143] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.*, 98(7):1978–1988, 1994.
- [144] C. C. Chambers, G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.*, 100(40):16385–16398, 1996.

Appendix A

Conversion of Dispersion Coefficients

The postg XDM code directly reports C_6 dispersion coefficients in atomic units. Parameter files for GROMACS and CHARMM store the dispersion coefficients through the parameters for the Lennard-Jones potential (Figure A.1). In GROMACS, the LJ potential is defined as

$$\mathcal{V}_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (\text{A.1})$$

The C_6 coefficient, in terms of these σ and ε parameters, is

$$C_6 = 4\varepsilon\sigma^6. \quad (\text{A.2})$$

These LJ parameters are given in terms of kJ/mol for ε and nm for σ . The conversion to atomic units is

$$1(\text{kJ/mol})\text{nm}^6 = 17344.659 \text{ a.u.} \quad (\text{A.3})$$

CHARMM defines the Lennard-Jones potential in terms of the location of the

potential-energy minimum, R_{min} , instead of σ .

$$\mathcal{V}_{LJ}(r) = \varepsilon \left[\left(\frac{R_{min}}{r} \right)^{12} - 2 \left(\frac{R_{min}}{r} \right)^6 \right] \quad (\text{A.4})$$

$$= \varepsilon \frac{R_{min}^{12}}{r^{12}} - 2\varepsilon \frac{R_{min}^6}{r^6}, \quad (\text{A.5})$$

In terms of R_{min} , C_6 is defined as

$$C_6 = 2\varepsilon R_{min}^6. \quad (\text{A.6})$$

These LJ parameters are given in terms of kcal/mol for ε and Å for R_{min} .¹ The conversion to atomic units is

$$1(\text{kcal/mol})\text{Å}^6 = 0.07257 \text{ a.u.} \quad (\text{A.7})$$

¹CHARMM-format parameter files actually store $R_{min}/2$

Appendix B

Tables of Side Chain Hydration Energies

Table B.1: Hydration energies of side-chain models calculated using the CHARMM36 force field. The energies are in units of kcal/mol. Dispersion coefficients are in atomic units. Experimental values are taken from Refs. 143 and 144.

Residue	ΔG_{elec}	ΔG_{disp}	ΔG_{rep}	ΔG_{total}	ΔG_{exptl}	$C_{6,\text{FF}}$	$C_{6,\text{XDM}}$
Ile	-0.05 ± 0.00	-10.85 ± 0.00	13.38 ± 0.05	2.48 ± 0.05	2.08	1476.2	1154.4
Leu	-0.05 ± 0.00	-10.73 ± 0.00	13.17 ± 0.00	2.39 ± 0.00	2.28	1515.3	1164.5
Ala	0.00 ± 0.00	-4.41 ± 0.00	6.79 ± 0.07	2.38 ± 0.07	2.00	147.8	121.3
Val	-0.05 ± 0.00	-8.91 ± 0.00	11.35 ± 0.02	2.40 ± 0.02	1.96	877.5	692.3
Asn	-9.23 ± 0.02	-9.37 ± 0.02	10.78 ± 0.00	-7.82 ± 0.05	-9.72	950.7	598.0
Tyr	-6.09 ± 0.00	-15.56 ± 0.02	16.61 ± 0.05	-5.04 ± 0.07	-6.13	3736.0	2410.5
Thr	-6.12 ± 0.00	-8.01 ± 0.02	10.18 ± 0.02	-3.94 ± 0.05	-4.90	643.8	476.0
Cys	-1.12 ± 0.00	-8.45 ± 0.00	9.25 ± 0.05	-0.32 ± 0.05	-1.24	732.9	442.3
Ser	-6.26 ± 0.00	-5.94 ± 0.00	7.91 ± 0.02	-4.29 ± 0.02	-5.08	286.6	199.5
Met	-0.60 ± 0.00	-12.56 ± 0.00	13.67 ± 0.02	0.52 ± 0.02	-1.49	2161.1	1311.1
Trp	-5.62 ± 0.02	-18.81 ± 0.02	18.67 ± 0.05	-5.76 ± 0.10	-5.91	5967.1	3703.3
Hid	-11.52 ± 0.02	-11.88 ± 0.00	13.10 ± 0.02	-10.30 ± 0.05	-10.25	1689.7	1322.5
Gln	-9.15 ± 0.02	-11.24 ± 0.00	12.81 ± 0.05	-7.59 ± 0.07	-9.42	1570.8	1023.9
Phe	-1.74 ± 0.00	-14.38 ± 0.02	15.65 ± 0.05	-0.47 ± 0.07	-0.76	3115.0	2130.8
Asp	-87.33 ± 0.02	-8.42 ± 0.00	10.30 ± 0.02	-85.46 ± 0.05	-80.65	736.7	638.7
Lys	-67.72 ± 0.10	-12.69 ± 0.00	14.56 ± 0.05	-65.86 ± 0.14	-69.24	2082.6	1389.5
Arg	-60.45 ± 0.02	-16.48 ± 0.02	16.73 ± 0.02	-60.21 ± 0.07		3910.7	2046.1
Glu	-84.82 ± 0.02	-10.43 ± 0.00	12.38 ± 0.02	-82.87 ± 0.05	-79.12	1291.7	1069.7
NMA	-8.41 ± 0.00	-12.36 ± 0.00	13.12 ± 0.07	-7.65 ± 0.07	-10.10	1937.9	1019.9

Table B.2: Hydration energies of side-chain models calculated using the Amber ff14sb force field. The energies are in units of kcal/mol. Dispersion coefficients are in atomic units.

Residue	ΔG_{elec}	ΔG_{disp}	ΔG_{rep}	ΔG_{total}	ΔG_{exptl}	$C_{6,\text{FF}}$	$C_{6,\text{XDM}}$
Ile	-0.07 ± 0.00	-10.25 ± 0.00	12.55 ± 0.00	2.22 ± 0.00	2.08	1679.0	1154.4
Leu	-0.10 ± 0.00	-10.04 ± 0.02	12.33 ± 0.05	2.20 ± 0.07	2.28	1679.0	1164.5
Ala	0.00 ± 0.00	-4.00 ± 0.00	6.41 ± 0.05	2.40 ± 0.05	2.00	147.9	121.3
Val	-0.17 ± 0.02	-8.38 ± 0.02	10.61 ± 0.05	2.07 ± 0.10	1.96	984.3	692.3
Asn	-10.56 ± 0.00	-8.78 ± 0.00	9.85 ± 0.02	-9.50 ± 0.02	-9.72	995.9	598.0
Tyr	-5.98 ± 0.00	-13.89 ± 0.02	15.30 ± 0.02	-4.57 ± 0.05	-6.13	3619.7	2410.5
Thr	-5.69 ± 0.00	-7.74 ± 0.00	9.27 ± 0.05	-4.16 ± 0.05	-4.90	738.6	476.0
Cys	-1.84 ± 0.00	-6.37 ± 0.02	8.20 ± 0.02	-0.01 ± 0.05	-1.24	516.0	442.3
Ser	-5.81 ± 0.00	-5.53 ± 0.00	7.10 ± 0.02	-4.24 ± 0.02	-5.08	301.0	199.5
Met	-1.53 ± 0.00	-10.41 ± 0.02	12.48 ± 0.05	0.54 ± 0.07	-1.49	1813.7	1311.1
Trp	-6.12 ± 0.02	-16.38 ± 0.02	17.23 ± 0.24	-5.26 ± 0.29	-5.91	5532.4	3703.3
Hid	-9.25 ± 0.02	-11.74 ± 0.02	12.50 ± 0.07	-8.49 ± 0.12	-10.25	2218.5	1322.5
Gln	-11.26 ± 0.02	-10.64 ± 0.00	11.93 ± 0.05	-9.97 ± 0.07	-9.42	1694.1	1023.9
Phe	-2.25 ± 0.00	-12.59 ± 0.02	14.56 ± 0.07	-0.28 ± 0.10	-0.76	2940.3	2130.8
Asp	-77.06 ± 0.05	-8.53 ± 0.00	9.63 ± 0.10	-75.96 ± 0.14	-80.65	897.5	638.7
Lys	-63.49 ± 0.02	-11.62 ± 0.00	13.38 ± 0.02	-61.72 ± 0.05	-69.24	2165.7	1389.5
Arg	-58.04 ± 0.05	-14.34 ± 0.02	15.32 ± 0.05	-57.06 ± 0.12		3569.2	2046.1
Glu	-77.37 ± 0.02	-10.41 ± 0.00	11.59 ± 0.05	-76.18 ± 0.07	-79.12	1565.2	1069.7
NMA	-9.15 ± 0.00	-10.86 ± 0.02	12.07 ± 0.05	-7.94 ± 0.07	-10.10	1730.1	1019.9

Table B.3: Hydration energies of side-chain models calculated using the OPLS-AA force field. The energies are in units of kcal/mol. Dispersion coefficients are in atomic units.

Residue	ΔG_{elec}	ΔG_{disp}	ΔG_{rep}	ΔG_{total}	ΔG_{exptl}	$C_{6,\text{FF}}$	$C_{6,\text{XDM}}$
Ile	-0.05 ± 0.00	-10.28 ± 0.02	12.50 ± 0.02	2.17 ± 0.05	2.08	1529.4	1154.4
Leu	0.00 ± 0.00	-10.09 ± 0.00	12.28 ± 0.05	2.19 ± 0.05	2.28	1529.4	1164.5
Ala	0.00 ± 0.00	-4.15 ± 0.00	6.31 ± 0.00	2.16 ± 0.00	2.00	144.6	121.3
Val	-0.05 ± 0.00	-8.43 ± 0.00	10.56 ± 0.02	2.09 ± 0.02	1.96	905.0	692.3
Asn	-9.37 ± 0.00	-9.21 ± 0.00	10.04 ± 0.00	-8.54 ± 0.00	-9.72	1155.3	598.0
Tyr	-6.76 ± 0.00	-14.20 ± 0.00	15.37 ± 0.02	-5.59 ± 0.02	-6.13	3739.0	2410.5
Thr	-6.07 ± 0.00	-7.63 ± 0.00	9.37 ± 0.00	-4.33 ± 0.00	-4.90	691.7	476.0
Cys	-1.65 ± 0.00	-7.43 ± 0.00	8.58 ± 0.02	-0.49 ± 0.02	-1.24	732.3	442.3
Ser	-6.14 ± 0.00	-5.51 ± 0.00	7.31 ± 0.02	-4.34 ± 0.02	-5.08	298.5	199.5
Met	-2.15 ± 0.00	-11.23 ± 0.00	12.81 ± 0.02	-0.57 ± 0.02	-1.49	2048.8	1311.1
Trp	-5.54 ± 0.00	-16.80 ± 0.00	17.38 ± 0.05	-4.97 ± 0.05	-5.91	5859.6	3703.3
Hid	-9.46 ± 0.00	-11.99 ± 0.00	12.62 ± 0.00	-8.84 ± 0.00	-10.25	2293.3	1322.5
Gln	-9.56 ± 0.00	-11.01 ± 0.00	11.88 ± 0.02	-8.70 ± 0.02	-9.42	1850.6	1023.9
Phe	-2.32 ± 0.00	-13.17 ± 0.00	14.53 ± 0.02	-0.96 ± 0.02	-0.76	3149.3	2130.8
Asp	-83.03 ± 0.00	-9.20 ± 0.00	9.66 ± 0.02	-82.57 ± 0.02	-80.65	1075.9	638.7
Lys	-70.04 ± 0.02	-11.44 ± 0.00	13.55 ± 0.02	-67.93 ± 0.05	-69.24	2046.1	1389.5
Arg	-61.48 ± 0.00	-12.99 ± 0.00	15.44 ± 0.02	-59.04 ± 0.02		2805.7	2046.1
Glu	-83.44 ± 0.02	-11.03 ± 0.00	11.64 ± 0.05	-82.82 ± 0.07	-79.12	1749.3	1069.7
NMA	-7.48 ± 0.00	-11.51 ± 0.00	12.40 ± 0.05	-6.58 ± 0.05	-10.10	1984.5	1019.9