

**NOVEL FUNCTIONAL FORMS AND PARAMETERIZATION METHODS FOR AB
INITIO FORCE FIELD DEVELOPMENT**

by

Mary J. Van Vleet

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: 08/15/17

The dissertation is approved by the following members of the Final Oral Committee:

J.R. Schmidt, Associate Professor, Chemistry

Clark R. Landis, Professor, Chemistry

Qiang Cui, Professor, Chemistry

Arun Yethiraj, Professor, Chemistry

Reid Van Lehn, Assistant Professor, Chemical and Biological Engineering

© Copyright by Mary J. Van Vleet 2017
All Rights Reserved

Soli Deo gloria.

ACKNOWLEDGMENTS

It is customary for authors of academic books to include in their prefaces statements such as this: “I am indebted to ... for their invaluable help; however, any errors which remain are my sole responsibility.” Occasionally an author will go further. Rather than say that if there are any mistakes then he is responsible for them, he will say that there will inevitably be some mistakes and he is responsible for them....

Although the shouldering of all responsibility is usually a social ritual, the admission that errors exist is not — it is often a sincere avowal of belief. But this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold logically incompatible beliefs.

— DAVID C. MAKINSON (1965)

CONTENTS

Contents iii

List of Tables xi

List of Figures xiv

Abstract xix

Published Work and Work in Preparation xx

| | |
|-----------------------|--|
| I Introduction | 1 |
| 1 | Introduction 2 |
| 1.1 | <i>Molecular Simulation: History and Importance</i> 2 |
| 1.2 | <i>Molecular Simulation: Challenges and Unanswered Questions</i> 5 |
| 2 | Background 9 |
| 2.1 | <i>Molecular Mechanics and the Theory of Intermolecular Forces</i> 9 |
| 2.1.1 | The Many-Body Expansion 9 |
| 2.1.2 | Energy Decomposition Schemes 9 |
| | Intramolecular Interactions 9 |
| | Electrostatics 9 |
| | Exchange 9 |
| | Induction 10 |
| | Dispersion 10 |
| 2.2 | <i>Ab-Initio Force Field Development</i> 10 |
| 2.2.1 | Electronic Structure Benchmarks 10 |
| | SAPT 10 |
| | Coupled-Cluster Methods 10 |
| 2.3 | <i>ISA-based methods for force field development</i> 10 |

| | |
|--|-----------|
| II Published Work | 11 |
| 3 Beyond Born–Mayer: Improved Models for Short-Range Repulsion in Ab Intio Force Fields | 12 |
| 3.1 <i>Introduction</i> | 12 |
| 3.2 <i>Theory</i> | 14 |
| 3.2.1 Models for the exchange-repulsion between isolated atoms | 15 |
| 3.2.2 Models for other short-range interactions between isolated atoms | 20 |
| 3.2.3 Models for short-range interactions between molecules | 21 |
| 3.3 <i>Computational Methods</i> | 25 |
| 3.3.1 Construction of the 91 dimer test set | 27 |
| 3.3.2 BS-ISA Calculations | 28 |
| 3.3.3 Determination of B_i^{ISA} | 28 |
| 3.3.4 Force Field Functional Forms and Parameterization | 29 |
| 3.3.5 Potential Energy Surface Scans | 33 |
| 3.3.6 Molecular Simulations | 33 |
| 3.4 <i>Results and Discussion</i> | 34 |
| 3.4.1 Accuracy: Comparison with DFT-SAPT | 35 |
| Argon Dimer | 41 |
| Ethane Dimer | 43 |
| Acetone Dimer | 44 |
| 3.4.2 Accuracy: Comparison with experiment | 46 |
| 3.4.3 Transferability | 49 |
| 3.4.4 Robustness | 51 |
| 3.4.5 Next-Generation Born-Mayer Models: Born-Mayer-sISA FF . | 55 |
| 3.5 <i>Conclusions and Recommendations</i> | 58 |
| 3.A <i>Waldman-Hagler Analysis of B_{ij} Combination Rule</i> | 60 |
| 3.B <i>Force Field Fits for Homomonomeric Systems</i> | 62 |

| | |
|--|-----------|
| 4 MASTIFF: A General Approach for Incorporating Atomic-level Anisotropy in Ab Initio Force Fields | 71 |
| 4.1 Introduction | 71 |
| 4.2 Background | 74 |
| 4.2.1 Prior Models for Long-Range Interactions | 75 |
| 4.2.2 Prior Models for Short-Range Interactions | 76 |
| 4.3 Theory and Motivation | 77 |
| 4.3.1 Anisotropic Models for Short-Range Interactions | 78 |
| Exchange-Repulsion | 78 |
| Other Short-Range Effects | 82 |
| 4.3.2 Anisotropic Models for Long-Range Interactions | 82 |
| Electrostatics | 82 |
| Induction | 83 |
| Dispersion | 84 |
| 4.4 Technical Details | 85 |
| 4.4.1 The 91 Dimer Test Set | 85 |
| 4.4.2 Parameter Determination | 85 |
| Parameters Calculated from Monomer Properties | 87 |
| Parameters Fit to Dimer Properties | 88 |
| Local Axis Determination | 88 |
| CCSD(T) Force Fields | 89 |
| CO ₂ 3-body potential | 90 |
| 4.4.3 Simulation Protocols | 90 |
| ΔH _{sub} for CO ₂ | 90 |
| Other CO ₂ Simulations | 91 |
| 2 nd Virial Calculations | 91 |
| 4.5 Results and Discussion | 91 |
| 4.5.1 Overview | 91 |
| 4.5.2 Accuracy: Comparison with DFT-SAPT | 93 |
| 4.5.3 Transferability: Comparison to DFT-SAPT | 98 |
| 4.5.4 Comparison to Experiment: Second Virial Coefficients | 99 |

| | |
|--|-----|
| 4.5.5 Comparison to Experiment: Condensed Phase Properties of CO ₂ | 105 |
| 4.6 Conclusions and Recommendations | 107 |
| 4.A Motivation for $g(\theta_i, \phi_i, \theta_j, \phi_j)$ | 108 |
| 4.B Local Axis Definitions | 110 |
| 4.B.1 Acetone | 111 |
| 4.B.2 Ar | 111 |
| 4.B.3 Chloromethane | 111 |
| 4.B.4 Carbon Dioxide | 111 |
| 4.B.5 Dimethyl Ether | 112 |
| 4.B.6 Ethane | 112 |
| 4.B.7 Ethanol | 112 |
| 4.B.8 Ethene | 113 |
| 4.B.9 Water | 113 |
| 4.B.10 Methane | 114 |
| 4.B.11 Methanol | 114 |
| 4.B.12 Methyl Amine | 114 |
| 4.B.13 Ammonia | 115 |
| 4.C Homodimer Fits | 115 |
| 4.D 2- and 3-body MASTIFF-CC CO₂ energies | 123 |
| III Unpublished Work | 125 |
| 5 Ab Initio Force Fields using LMO-EDA | 99 |
| 5.1 Preface | 99 |
| 5.2 Introduction | 100 |
| 5.3 Background and Motivation | 101 |
| 5.4 Parameterizing Coordinatively-Unsaturated (CUS)-Metal-Organic Framework (MOF) force fields with LMO-EDA | 104 |
| 5.5 Computational Methods | 108 |
| 5.5.1 Partial Charge Determination | 108 |

| | | |
|---|---|------------|
| 5.5.2 | Force Field Fitting | 108 |
| 5.6 | <i>Results</i> | 109 |
| 5.6.1 | Initial Force Field and Cluster Model Analysis | 109 |
| 5.6.2 | Final Mg-MOF-74 CO ₂ Adsorption Isotherm | 113 |
| 5.6.3 | Transferability to Other Adsorption Isotherms | 114 |
| 5.6.4 | Transferability to Other M-MOF-74 systems | 115 |
| 5.7 | <i>Conclusions</i> | 116 |
| 5.8 | <i>Future Work</i> | 116 |
| 5.A | <i>Force Field Parameters for CO₂ and Mg-MOF-74</i> | 119 |
| 5.B | <i>Simulation Parameters CO₂ Adsorption in Mg-MOF-74</i> | 121 |
| 6 | Benchmark Database for Ab Initio Force Field Development | 123 |
| IV | Practical Matters | 124 |
| 7 | Workflow for Intermolecular Force Field Development | 125 |
| 7.1 | <i>Overview</i> | 125 |
| 7.2 | <i>Geometry Generation</i> | 128 |
| 7.2.1 | Guiding Principles | 128 |
| 7.2.2 | Theory | 130 |
| 7.2.3 | Practicals | 131 |
| 7.3 | <i>Symmetry-Adapted Perturbation Theory (SAPT) Benchmarks</i> | 131 |
| 7.4 | <i>CCSD(T) Calculations</i> | 132 |
| 7.5 | <i>Monomer-Based Parameterization</i> | 133 |
| 7.5.1 | Distributed Property Calculations using CamCASP | 133 |
| 7.5.2 | Multipoles | 134 |
| Practicals | 134 | |
| Advanced Multipole Parameterization Options | 135 | |
| 7.5.3 | ISA Exponents | 137 |
| 7.5.4 | Dispersion Coefficients | 138 |
| Theory | 138 | |

| | |
|---|------------|
| Iterative-Distributed Multipole Analysis (DMA)-pol | 140 |
| Theory | 140 |
| Practicals | 140 |
| Iterated Stockholder Atoms (ISA)-pol | 143 |
| Theory | 143 |
| Practicals | 144 |
| Comparison between iDMA-pol and ISA-pol | 144 |
| Dispersion Coefficient Post-processing | 146 |
| 7.5.5 Polarization Charges | 147 |
| Theory | 147 |
| Practicals | 147 |
| 7.6 Dimer-Based Parameterization | 148 |
| 7.A Input Scripts | 149 |
| 7.B Algorithm for Obtaining ISA Exponents | 152 |
| 8 POInter: A Program for Intermolecular Force Field Optimization | 154 |
| 8.1 Overview | 154 |
| 8.2 Parameterization Overview | 155 |
| 8.2.1 Theory | 155 |
| 8.3 The Parameter Optimizer for Inter-molecular Force Fields (POInter) Code | 160 |
| 8.3.1 Input | 160 |
| 8.3.2 Usage and Output | 162 |
| 8.4 Force Field Development: Principles and Practice | 163 |
| 8.4.1 General | 163 |
| Atom-typing | 163 |
| Anisotropy | 164 |
| Benchmark Energies and Correction Factors | 165 |
| 8.4.2 Exchange | 165 |
| Exponent Fitting | 165 |
| 8.4.3 Electrostatics | 166 |
| Off-site models | 166 |

| | | |
|--|--|------------|
| 8.4.4 | Induction | 167 |
| Polarization | 168 | |
| Polarization Damping | 169 | |
| Charge transfer and inductive charge penetration . . . | 170 | |
| 8.4.5 | Dispersion | 170 |
| 8.4.6 | Many-Body Effects | 172 |
| 8.5 | <i>Validation: Assessing Fit Quality</i> | 172 |
| 8.5.1 | Sanity Checks | 172 |
| Visualization | 172 | |
| Error Analysis of the Minimum Energy Region | 173 | |
| Error Analysis of the Asymptotic Region | 177 | |
| 8.5.2 | Validation | 177 |
| Trimer Configurations | 177 | |
| Simulations | 177 | |
| 8.6 | <i>Summary and Outlook</i> | 178 |
| 8.A | <i>POInter Input Files</i> | 178 |
| 8.B | <i>POInter Output Files</i> | 183 |
| 8.C | <i>Additional Fitting Options</i> | 186 |
| V | Conclusions and Future Work | 188 |
| 9 | Future Work | 189 |
| 10 | Conclusions | 190 |
| VI | Codes | 191 |
| A | Force Field Development Workflow | 192 |
| A.1 | <i>Monomer Geometries</i> | 196 |
| | Bibliography | 197 |

Acronyms 216

Glossary 217

LIST OF TABLES

| | |
|--|----|
| 3.1 Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Slater-ISA FF, Born-Mayer-IP FF and LJ FF. For the total energy, both characteristic RMSE and MSE have been shown, with only the magnitude of the MSE, $\ \text{MSE}\ $, displayed. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ($E_{\text{int}} < 0$), are shown in parentheses to the right of the total RMS errors; ‘attractive’ $\ \text{MSE}\ $ are likewise displayed for the total energy. As discussed in Section 3.4.3, the ‘Dimer-Specific Fits’ refer to force fields whose parameters have been optimized for each of the 91 dimers separately, whereas the ‘Transferable Fits’ refer to force fields whose parameters have been optimized for the 13 homodimers and then applied (without further optimization) to the remaining 78 mixed systems. Unless otherwise stated, a default weighting function of $\lambda = 2.0$ (see Eq. (3.37)) has been used for all force fields in this Chapter. | 37 |
| 3.2 Comparison of characteristic RMSE and $\ \text{MSE}\ $ over the 91 dimer test set for the various Lennard-Jones models. The LJ models are not parameterized on a component-by-component basis, thus RMSE/ $\ \text{MSE}\ $ values are only shown for the total FF energies. ‘Attractive’ errors, representing the characteristic RMSE/ $\ \text{MSE}\ $ for the subset of points whose energies are net attractive ($E_{\text{int}} < 0$), are shown in parentheses to the right of the total errors. ‘Dimer-Specific Fits’ and ‘Transferable Fits’ are as in Table 3.1. | 40 |
| 3.3 Characteristic RMS pairwise differences (RMSD) in force field total energies for different weighting functions with λ values as defined in Eq. (3.37); values shown are the (arithmetic mean, rather than geometric) RMSD across the 91 dimer test set. Characteristic ‘Attractive’ RMSD (as defined in Table 3.1) are shown in parentheses to the right of each overall RMSD. | 52 |

| | |
|--|-----|
| 3.4 Enthalpies of vaporization and liquid densities for ethane as a function of force field and weighting function. Values in parentheses include an estimation of the 3-body correction ($0.628 \text{ kJ mol}^{-1}$ and 0.034 g mL^{-1} for the enthalpy of vaporization and liquid density, respectively) as computed in Ref. 2. Experimental data taken from Ref. 3 and Ref. 4. | 53 |
| 3.5 Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Born-Mayer-sISA approximation compared with other methods. For the total energy, both RMSE and absolute mean signed errors (MSE) have been shown. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ($E_{\text{int}} < 0$), are shown in parentheses to the right of the total RMS errors; ‘attractive’ $\ \text{MSE}\ $ are likewise displayed for the total energy. Slater-ISA FF, Born-Mayer-ISA, and Born-Mayer-sISA FF are as described in the main text, and the ‘Dimer-Specific’ and ‘Transferable’ fits are as described in Table 3.1. | 56 |
| 4.1 ‘Improvement Ratios’ for each homomonomeric species in the 91 dimer test set. For each dimer and energy component, the improvement ratio is calculated as the ratio of aRMSE between Iso-Iso FF and MASTIFF; values greater than 1 indicate decreased errors in the anisotropic model. Entries have been ordered according to the improvement ratio for the total energy. | 97 |
| 4.2 Select densities for CO_2 across a range of experimental conditions. Experimental data taken from the EOS of Ref. 5. Entries ordered by increasing experimental density. | 107 |
| 4.3 Enthalpies of vaporization/sublimation for CO_2 at several temperatures. Experimental data taken from the EOS of Ref. 5. The uncertainty in the enthalpy of sublimation is due to ambiguity in the theoretical zero-point energy for CO_2 (see Section 4.4. | 107 |
| 7.1 Overview of ISA- and DMA-based methods for obtaining distributed monomer properties | 134 |

7.2 Comparison between the iDMA-pol and ISA-pol methods. 145

LIST OF FIGURES

| | |
|--|----|
| 1.1 Simple and complex potential energy surfaces for molecular systems. | 7 |
| 3.1 BS-ISA and fitted shape functions for each atom type in acetone: a) carbonyl carbon, b) oxygen, c) methyl carbon, d) hydrogen. BS-ISA shape functions (dotted line) for each atom type have been obtained at a PBE0/aug-cc-pVTZ level of theory. A modified BS-ISA shape function (dashed line) corrects the tail-region of the BS-ISA function to account for basis set deficiencies in the BS-ISA algorithm. A single Slater orbital of the form $D_i^{\text{ISA}} \exp(-B_i^{\text{ISA}}r)$ (solid line) is fit to the basis-corrected BS-ISA shape function, and the obtained B_i^{ISA} value is used as an atomic exponent in the functional form of Aniso-Iso FF. Results for acetone are typical of molecules studied in this Chapter. | 24 |
| 3.2 The 13 small molecules included in the 91 dimer (13 homomonomeric, 78 heteromonomeric) test set. Cartesian geometries for all of these molecules are given in Section A.1. | 27 |
| 3.3 Characteristic RMSE (as described in the main text) for the Born-Mayer-IP FF (orange) and the Slater-ISA FF (green) over the 91 dimer test set. The translucent bars represent total RMSE for each energy component, while the smaller solid bars represent ‘Attractive’ RMSE, in which repulsive points have been excluded. | 36 |
| 3.4 Potential energy surface for the argon dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. | 42 |

| | |
|---|----|
| 3.5 Force field fits for the ethane dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs. Fits for each energy component are displayed along with two views of the total interaction energy. The diagonal line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within $\pm 10\%$ agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each energy component. | 43 |
| 3.6 A representative potential energy scan near a local minimum for the ethane dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The ethane dimer configuration in this scan corresponds to the most energetically attractive dimer included in the training set; other points along this scan are not included in the training set. | 45 |
| 3.7 Force field fits for the acetone dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs, as in Fig. 3.5. | 46 |
| 3.8 A representative potential energy scan near a local minimum for the acetone dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The intermolecular distance is taken to be the internuclear distance between the two carbonyl carbons on each acetone monomer. The configuration in this scan corresponds to the most attractive dimer configuration included in the training set for the acetone dimer; other points along this scan have not explicitly been included in the training set. | 47 |
| 3.9 Second virial coefficients for argon. The Slater-ISA and the Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from Ref. 6. | 48 |

| | |
|--|----|
| 3.10 Second virial coefficients for ethane. The Slater-ISA and Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from Ref. 6. | 49 |
| 3.11 foo | 54 |
| 3.12 Mean absolute percent error of fitted overlap values as a function of the absolute difference between B_i and B_j values. Element pairs containing He, Li, Ne and/or Na are shown as empty circles. Deviations below 1% are seen for most element pairs, with noble gases and alkali metals posing a more significant challenge. Scatter in the plot is due to small variations in the absolute values of r_{ij} fit for each pair. As expected, $S_{B_i \neq B_j}^{ij}$ and $S_{B_i = B_j}^{ij}$ closely agree for $ B_i - B_j \approx 0$ | 62 |
| 3.13 Waldman-Hagler-style analysis of possible B_{ij} combination rules. Exact B_{ij} values are derived from fitting an approximate overlap density of the form $S_{ij} = A_{ij} K_2(r_{ij}) \exp(-B_{ij} r_{ij})$ to the exact overlap density (as given by Rosen and by Tai ⁷⁸) of two distinct Slater orbitals whose exponents correspond to atomic exponents for the elements H-Ar, Cl, Br, and I. For each overlap pair, a range of r_{ij} values was used from 0.8 to 1.2 times the sum of the pair's van der Waals radii. The geometric mean combination rule $B_{ij} = \sqrt{B_i B_j}$ models the exact B_{ij} values with near-perfect agreement, justifying our choice of combination rule. | 63 |
| 3.14 Force field fits for the homomeric systems using the Slater-ISA FF (green), Born-Mayer-IP FF (orange) and Born-Mayer-sISA FF (blue). Fits for each energy component are displayed along with two views of the total interaction energy. The $y = x$ line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within $\pm 10\%$ agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each energy component. | 70 |
| 4.1 Local axis system, shown for select atoms in molecules. | 79 |

| | |
|---|-----|
| 4.9 Force field quality for MASTIFF-CC in reproducing (left) two-body and (right) three-body CO ₂ interaction energies. The $y = x$ line (solid) and ± 1 kJ/mol boundaries (dashed lines) are shown for reference. All dimer/trimer configurations were taken from a snapshot of CO ₂ liquid simulated using MASTIFF-CC at 273.15 K and 100 bar. Reference energies are taken from the Kalugina et al. ¹³ PES for the two-body energies, and the Hellmann ¹² PES for the three-body energies. A total of 62,583 dimer configurations and 43,784 trimer configurations are represented in the two plots. | 124 |
| 5.1 Model potential energy surface (PES) for interactions between CO ₂ and Mg-MOF-74 | 103 |
| 5.2 LMO-EDA vs. SAPT PES for the CO ₂ dimer | 106 |
| 5.3 LMO-EDA vs. SAPT PES for the CO ₂ /Mg-MOF-74 dimer | 107 |
| 5.4 Force field fitting quality for the Mg-MOF-74-small cluster | 110 |
| 5.5 Model clusters for Mg-MOF-74 | 111 |
| 5.6 Force field fitting quality for Mg-MOF-74-Yu | 114 |
| 5.7 Predicted CO ₂ Adsorption Isotherm for Mg-MOF-74 | 115 |
| 7.1 Generalized form of a PES showing the repulsive wall, minimum energy, and asymptotic regions. | 128 |
| 7.2 Linear extrapolation algorithm for the methyl carbon in acetone. | 153 |
| 8.1 Required parameters for MASTIFF | 157 |
| 8.2 Pyridine and water – molecular examples of challenges in force field fitting | 163 |
| 8.3 Comparison with the pyridine dimer | 173 |
| 8.4 Errors with the pyridine dimer | 174 |
| 8.5 Comparison with the water dimer | 174 |
| 8.6 Comparison with the water dimer | 175 |
| 8.7 Errors with the water dimer | 175 |
| A.1 The Semi-Automated Workflow for Force Field Development | 195 |

**NOVEL FUNCTIONAL FORMS AND PARAMETERIZATION METHODS
FOR AB INITIO FORCE FIELD DEVELOPMENT**

Mary J. Van Vleet

Under the supervision of Professor J.R. Schmidt
At the University of Wisconsin-Madison

FIXME: basically a placeholder; do not believe

J.R. Schmidt

ABSTRACT

FIXME: basically a placeholder; do not believe

PUBLISHED WORK AND WORK IN PREPARATION

- [104] Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.

Part I

Introduction

1 INTRODUCTION

1.1 Molecular Simulation: History and Importance

What are the functions of proteins in the body? How can we identify new and better drugs for improved disease treatment, or optimal materials for designing efficient solar cells? What are the microscopic mechanisms by which chemicals interact, undergo phase transitions, or react to form entirely new species? Increasingly, these and other essential chemical questions can be addressed with the aid of computer simulation,^{14–18} enabling us to, for example, peer into the detailed mechanisms of enzyme catalysis,¹⁹ watch proteins fold,^{20–23} virtually screen for novel drug candidates,²⁴ improve industrial materials,^{25–27} and directly simulate hard-to-understand phase transformations at an atomistic level.^{28,29} The question, of course, is: how?

To understand the link between computer simulation and observable experimental properties of interest,[†] we first summarize several fundamental physical principles that help define the scope and importance of molecular simulation. Our first insight comes from the field of statistical mechanics, where by the mid-19th century it was discovered that experimental observables, which we denote \mathcal{O} , depend entirely on a system's temperature, T , and the energy states (*vida infra*) available to that system:

$$\langle \mathcal{O} \rangle = \frac{\sum_n \mathcal{O}_n \exp(-E_n/k_B T)}{\sum_n \exp(-E_n/k_B T)} \quad (1.1)$$

Here the angle brackets denote that we're interested in some *average* value of the observable, the subscript n indicates a possible state of the system (defined in terms of the positions and momenta of all of the particles in the system) with energy E_n and value of the observable \mathcal{O}_n , and k_B is Boltzmann's constant.³⁰ Technically, this expression only holds for classical systems at constant temperature, however conceptually-similar expressions can be derived outside of these assumptions.

Regardless of the exact expression, the essential point is as follow: given knowledge of all possible energy states of a system, we can have exact information for that system about any macroscopic properties of interest.

In practice, most chemical systems contain large numbers of particles, and the number of available system states become intractably large. Thus in all but the simplest cases, Eq. (1.1) is overly difficult to solve by mere enumeration and averaging of states.³⁰ Fortunately, several crucial scientific advances paved the way to reformulate the relationship between $\langle \mathcal{O} \rangle$ and E_n in terms of known — and tractable — problems in such disparate fields as classical dynamics and probability theory. The connection to classical dynamics began at the turn of the 20th century, when Ludwig Boltzmann proposed his now-famous ‘ergodic’ hypothesis.³¹ Over sufficiently long time periods, according to Boltzmann, the ‘ensemble average’ defined in Eq. (1.1) becomes identical to the ‘time-average’ that results from studying the system’s dynamical behavior:

$$\langle \mathcal{O} \rangle = \langle \mathcal{O} \rangle_{\text{time}} = \langle \mathcal{O}(\Gamma(t)) \rangle_{\text{time}} = \lim_{t_{\text{obs}} \rightarrow \infty} \frac{1}{t_{\text{obs}}} \int_0^{t_{\text{obs}}} \mathcal{O}(\Gamma(t)) dt. \quad (1.2)$$

Here we have used $\Gamma(t)$ to denote the collective positions and momenta that define the state of the system as a function of time, and t_{obs} to indicate the length of time over which we average the system’s properties.³⁰ The right-hand side of Eq. (1.2) proves remarkably useful: so long as we know and can solve for the equations of motion that govern a given system’s behavior, we can simulate the time evolution of that system until the r.h.s. of Eq. (1.2) converges, and in doing so obtain an average value for any desired property of that system.

Boltzmann’s ergodic hypothesis wasn’t fully taken advantage of until the 1960s and 70s, as the laws of motion governing quantum mechanical systems can often

[†] We’ve been intentionally vague about what these ‘experimental properties of interest’ might be, as the experimental properties one finds important vary considerably between applications. To offer some concrete examples, drug discovery studies are often interested in the binding free energies between proteins and prospective drug molecules,^{24,27} and the optimization of putative solar cell materials often focuses on open circuit voltages and/or short-circuit currents.²⁷

be incredibly complex. *Classical* mechanics, on the other hand, is governed simply by Newton's laws of motion,

$$\frac{dE(x)}{dx} \equiv F(x) = m\ddot{x}, \quad (1.3)$$

with x as a position vector, m a mass, and F the force acting on a particle. Thus for classical particles, we need only know the forces operating on each particle (which is a direct function of the system's energy) to solve for the time-dependent positions, momenta, and (ultimately) properties of a system. To quote Karplus, who along with Levitt and Warshel won the Nobel Prize in 2013 for their work in molecular simulation, the important insight required to solve the dynamics problem for molecular systems was as follows:

Although the laws governing the motions of atoms are quantum mechanical, the key realization that made possible the simulation of the dynamics of complex systems, including biomolecules, was that a classical mechanical description of the atomic motions is adequate in most cases.

—Karplus³², Nobel Lecture 2013

As a result of this important insight, it soon became possible to study both kinetic and thermodynamic properties of a wide range of molecular systems (historically beginning with monatomic liquids in 1964, and leading to polymer and protein simulations by 1975 and 1977, respectively),¹⁴ and the field of molecular dynamics (MD) has since been established as a preeminent tool in the investigation and prediction of chemical phenomena.

Around the same time as MD was being developed and popularized as a 'dynamics' solution to Eq. (1.1), a separate group of researchers³³ found a way to formulate Eq. (1.1) as a sampling problem based in probability theory:^{30,34}

$$\langle \mathcal{O} \rangle = \langle \mathcal{O} \rangle_{\text{ens}} = \lim_{\tau_{\text{obs}} \rightarrow \infty} \frac{1}{\tau_{\text{obs}}} \sum_{\tau=1}^{\tau_{\text{obs}}} \mathcal{O}(\Gamma(\tau)) \quad (1.4)$$

In contrast to Eq. (1.1), Eq. (1.4) is the average over a total number of observations, τ_{obs} , taken of the system and its properties randomly sampled, according to some probability distribution ρ , over all states of the system. The key technique that defines MC is known as ‘importance sampling’: provided we cleverly choose our probability distribution, ρ , to be identical to the Boltzmann distribution, $\rho \propto \exp(-E_n/k_B T)$, the right-hand side of Eq. (1.4) converges fairly rapidly as a function of τ_{obs} . The interested reader is directed to Allen and Tildesley³⁰ for detailed information on the exact techniques, algorithms, and practical concerns involved in importance sampling. Regardless of which specific technique is used, however, both MD and Metropolis Monte Carlo (MC) methods enable us to calculate thermodynamic (and sometimes kinetic) properties of interest based solely on knowledge of the energy of a system as a function of the system state.

1.2 Molecular Simulation: Challenges and Unanswered Questions

Recent successes in MD and MC simulations have shown great promise for using molecular simulation in the understanding, interpretation, and even prediction of experimental results,¹⁴ making simulation a powerful complement to traditional experimental tools.^{15–17,25–27,35,36} Nevertheless, accurate and insightful molecular simulation depends on success in the following three critical aspects of any MD/MC simulation:

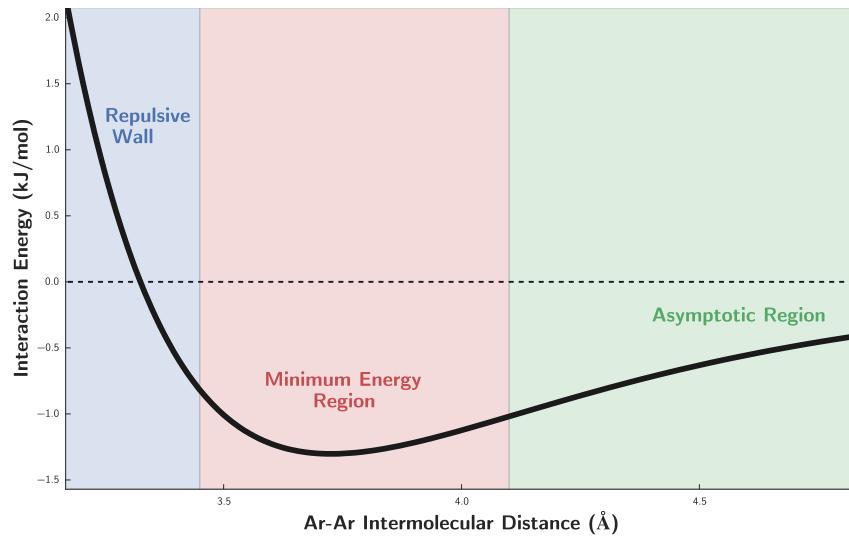
1. We must be able to accurately and efficiently quantify the potential energy, E_n , of any state n of the system that might get sampled by the MD/MC simulation.^{37–40} Henceforth we will collectively refer to these energies, given as a function of the system state Γ , as the potential energy surface (PES) of a system (Some visual examples of representative PESs are shown in Fig. 1.1).
2. We must be able to obtain a representative sample of all states of the system over a sufficiently long timescale (commensurate with the timescale(s)

of the chemical phenomena of interest) so as to obtain converged property predictions.^{41–43} This class of problems is often referred to simply as ‘sampling issues’.

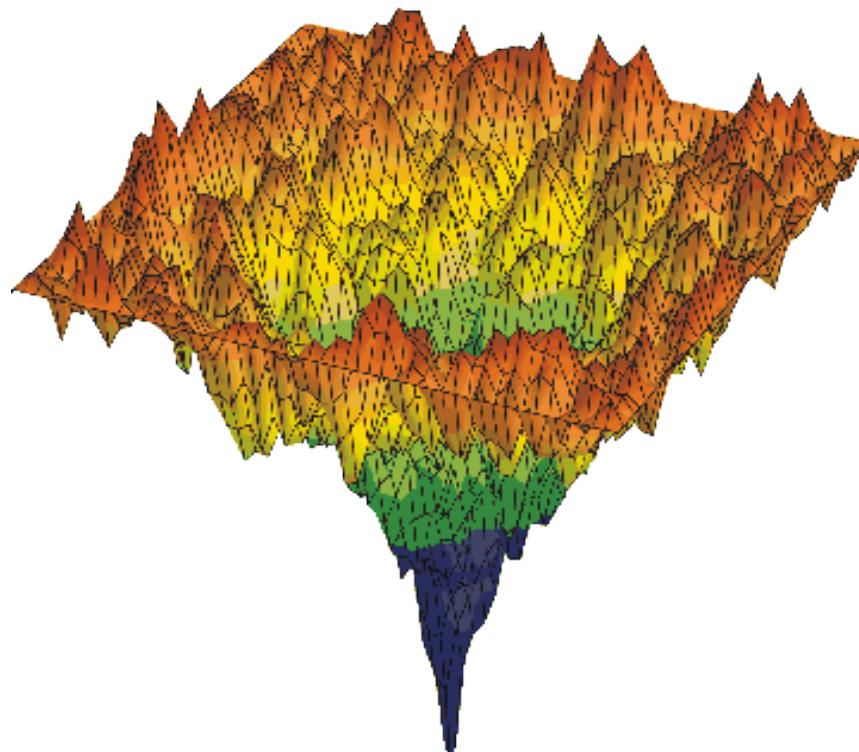
3. Especially when interested in *interpreting* chemical phenomena, we must be able to analyze the results of a simulation in such a way as to garner detailed, chemically-intuitive insight into the problem at hand.^{44–46} While this task is relatively straightforward in some cases, such as with studying the structural properties of homogeneous liquids, it can become decidedly difficult for analyzing complex systems and properties, such as with using simulation to investigate protein folding mechanisms.

Though our focus in this dissertation will be on the first point (that of computing potential energies for molecular simulation), all three aspects of molecular simulation are challenging in their own right, and form highly active and important areas of research.¹⁸ Moreover, there is significant interplay between these areas in terms of research development. As an example, improvements in sampling methods often lead to increased computational efficiency, thereby enabling use of more accurate (but more costly) representations of the PES. Conversely, the development of cost-effective potential energy functions is often necessary for running simulations over long enough time scales to ensure representative sampling and robust interpretation of the simulation results. Clearly, insofar as molecular simulation is concerned, both accuracy and computational efficiency are of paramount importance.

In the pursuit of increasingly accurate, insightful, and predictive molecular simulation, it is clear that we must be able to quantitatively represent the PES of any molecular system, and, furthermore, that our mathematical representation of this PES must be sufficiently accurate and cost-effective so as to enable simulation that is chemically insightful (given the type of simulation analysis required for a particular problem or application) and computationally affordable (in accordance with the length of molecular simulation that will need to be run in order to appropriately deal with any sampling issues). Bearing these stipulations in mind, we can now state the central question with which this dissertation is concerned: for the purposes



(a) One-dimensional PES for the argon dimer.



(b) Three-dimensional representation of an N-dimensional PES for a protein. Copied under a CC license from Chaplin⁴⁷

Figure 1.1: Simple and complex potential energy surfaces for molecular systems.

of running accurate and insightful molecular simulation, how can we optimally obtain a mathematical description of an arbitrary molecular PES?

Part II

Published Work

3 BEYOND BORN–MAYER: IMPROVED MODELS FOR SHORT-RANGE REPULSION IN AB INTIO FORCE FIELDS

3.1 Introduction

Molecular simulation is an essential tool for interpreting and predicting the structure, thermodynamics, and dynamics of chemical and biochemical systems. The fundamental inputs into these simulations are the intra- and intermolecular force fields, which provide simple and computationally efficient descriptions of molecular interactions. Consequently, the predictive and explanatory power of molecular simulations depends on the fidelity of the force field to the underlying (exact) potential energy surface.

In the case of intermolecular interactions, the dominant contributions for non-reactive systems can be decomposed into the following physically-meaningful energy components: electrostatic, exchange-repulsion, induction and dispersion.^{48–52} At large intermolecular distances, where monomer electron overlap can be neglected, the physics of intermolecular interactions can be described entirely on the basis of monomer properties (e.g. multipole moments, polarizabilities), all of which can be calculated with high accuracy from first principles.⁵³ In conjunction with associated distribution schemes that decompose molecular monomer properties into atomic contributions,^{48,51,54–58} these monomer properties lead to an accurate and computationally efficient model of ‘long-range’ intermolecular interactions as a sum of atom-atom terms, which can be straightforwardly included in common molecular simulation packages.

At shorter separations, where the molecular electron density overlap cannot be neglected, the asymptotic description of intermolecular interactions breaks down due to the influence of Pauli repulsion, charge penetration and charge transfer. These effects can be quantitatively described using modern electronic structure methods,^{50,59–62} but are far more challenging to model accurately using computationally inexpensive force fields. For efficiency and ease of parameterization, most

simple force fields use a single ‘repulsive’ term to model the cumulative influence of (chemically distinct) short-range interactions. These simple models have seen comparatively little progress over the past eighty years, and the Lennard-Jones⁶³ (A/r^{12}) and Born-Mayer^{64,65} ($A \exp(-Br)$) forms continue as popular descriptions of short-range effects in standard force fields despite some well-known limitations (*vide infra*).

Because the prediction of physical and chemical properties depends on the choice of short-range interaction model,^{66–79} it is essential to develop sufficiently accurate short-range force fields. This is particularly true in the case of ab initio force field development. A principle goal of such a first-principles approach is the reproduction of a calculated potential energy surface (PES), thus (ideally) yielding accurate predictions of bulk properties.⁸⁰ Substantial deviations between a fitted and calculated PES lead to non-trivial challenges in the parameterization process, which in turn can often degrade the quality of property predictions. The challenge of reproducing an ab initio PES becomes particularly pronounced at short inter-molecular separations, where many common force field functional forms are insufficiently accurate. For example, the popular Lennard-Jones (A/r^{12}) functional form is well-known to be substantially too repulsive at short contacts as compared to the exact potential.^{74–76,81,82} While the Born-Mayer ($A \exp(-Br)$) functional form is more physically-justified⁶⁵ and fares better in this regard,⁸¹ substantial deviations often persist.⁸³ In addition, parameterization of the Born-Mayer form is complicated by the strong coupling of the pre-exponential (A) and exponent (B) parameters, hindering the transferability of the resulting force field. These considerations, along with the observed sensitivity of structural and dynamic properties to the treatment of short-range repulsion,⁶⁶ highlight the need for new approaches to model short-range repulsive interactions.

Our primary goal in this Chapter is to derive a simple and accurate description of short-range interactions in molecular systems that improves upon both the standard Lennard-Jones and Born-Mayer potentials in terms of accuracy, transferability, and ease of parameterization. Our focus is on ab initio force field development, and thus we will use the fidelity of a given force field with respect to an accurate

ab initio PES as a principle metric of force field quality. We note that other metrics may be more appropriate for the development of empirical potentials, where Lennard-Jones or Born-Mayer forms may yield highly accurate ‘effective’ potentials when parameterized against select bulk properties. Nonetheless, we anticipate that the models proposed in this Chapter may prove useful for empirical force field development in cases where a more physically-motivated functional form is necessary.^{74–76}

The outline of this Chapter is thus as follows: first, we derive a new functional form capable of describing short-range repulsion from first principles, and show how the standard Born-Mayer form follows as an approximation to this more exact model. Our generalization of the Born-Mayer functional form allows for an improved description of a variety of short-range effects, namely electrostatic charge penetration, exchange-repulsion, and density overlap effects on induction and dispersion. Crucially, we also demonstrate how the associated atomic exponents can be extracted from first-principles monomer charge densities via an iterated stockholder atoms (ISA) density partitioning scheme, thereby reducing the number of required fitting parameters compared to the Born-Mayer model. Benchmarking this ‘Slater-ISA’ methodology (functional form and atomic exponents) against high-level ab initio calculations and experiment, we find that the approach exhibits increased accuracy, transferability, and robustness as compared to a typical Lennard-Jones or Born-Mayer potential. In addition, we show how the ISA-derived exponents can be adapted for use within the standard Born-Mayer form (Born-Mayer-sISA), while still retaining many of the advantages of the Slater-ISA approach. As such, our methodology also offers an opportunity to dramatically simplify the development of both empirically-parameterized and ab initio simulation potentials based upon the standard Born-Mayer form.

3.2 Theory

We begin with a formal treatment of the overlap model for the exchange-repulsion between two isolated atoms, and then extend these results to develop a general-

ized model for the short-range interactions in both atomic and molecular systems. Finally, we show how the conventional Born-Mayer model can be derived as an approximation to this more rigorous treatment.

3.2.1 Models for the exchange-repulsion between isolated atoms

It is well known that the exchange-repulsion interaction between two closed-shell atoms i and j is proportional, or very nearly proportional, to the overlap of their respective charge densities:⁸⁴

$$E_{ij}^{\text{exch}} \approx V_{ij}^{\text{exch}} = K_{ij}(S_{\rho}^{ij})^{\gamma} \quad (3.1)$$

$$S_{\rho}^{ij} = \int \rho_i(\mathbf{r}) \rho_j(\mathbf{r}) d^3\mathbf{r}. \quad (3.2)$$

Here and throughout, we use E to denote quantum mechanical energies, and V to denote the corresponding model/force field energies. Recently two of us have provided a theoretical justification for this repulsion hypothesis (or overlap model), and have shown that $\gamma = 1$ provided that asymptotically-correct densities are used to compute both the atomic densities and E_{ij}^{exch} .^{51,85} As this is the case for the calculations in this work, we assume $\gamma = 1$ throughout the Chapter.

The overlap model has frequently been utilized in the literature and has been found to yield essentially quantitative accuracy for a wide variety of chemical systems.^{84,86,87} Prior work exploiting the overlap model has generally followed one of two strategies. Striving for quantitative accuracy, several groups have developed approaches to evaluate Eq. (3.2) via either numerical integration or density fitting of ab-initio molecular electron densities, ρ_i (e.g. SIBFA, GEM, effective fragment potentials).^{88–97} These force fields, while often extremely accurate, lack the simple closed-form analytical expressions that define standard force fields (such as the Lennard-Jones or Born-Mayer models) and thus are often much more computationally expensive than conventional models.

In contrast, and similar to our objectives, the overlap model has also been used in the development of standard force fields. In this case, the molecular electron

density as well as the overlap itself is drastically simplified in order to yield a simple closed-form expression that can be used within a conventional molecular simulation package.^{84,86,87} As we show below, the Born-Mayer model can be ‘derived’ via such an approach. At the expense of some accuracy, the resulting overlap-based force fields exhibit high computational efficiency and employ well-known functional forms.

Building on this prior work, our present goal is to derive rigorous analytical expressions and improved approximations for both ρ_i and Eq. (3.2), facilitating the construction of ab initio force fields that exhibit simplicity, high computational efficiency, fidelity to the underlying PES, and (with only trivial modifications) compatibility with standard simulation packages. We first start with the case of isolated atoms, where it is well-known that the atomic electron density decays asymptotically as

$$\rho_{r \rightarrow \infty}(r) \propto r^{2\beta} e^{-2\alpha r} \quad (3.3)$$

where the exponent $\alpha = \sqrt{2I}$ is fixed by the vertical ionization potential I , $\beta = -1 + \frac{Q}{\alpha}$, and $Q = Z - N + 1$ for an atom with nuclear charge $+Z$ and electronic charge $-N$.^{85,98–100} The exponential term dominates the asymptotic form of the density, and the r -dependent prefactor may be neglected^{58,86,87,101}. In this case, the density takes the even simpler form

$$\rho_{r \rightarrow \infty}(r) \approx D e^{-Br}, \quad (3.4)$$

where D is a constant that effectively absorbs the missing r -dependent pre-factor and B is an exponent that is now only approximately equal to 2α .

In the case of two identical atoms, substitution into Eq. (3.2) yields a simple

expression for the density overlap, S_ρ ,^{7,8}

$$\begin{aligned} S_\rho^{ii} &= \frac{\pi D^2}{B^3} P(Br_{ii}) \exp(-Br_{ii}) \\ P(Br_{ii}) &= \frac{1}{3}(Br_{ii})^2 + Br_{ii} + 1 \end{aligned} \quad (3.5)$$

as well as (via Eq. (3.1)) the exchange-repulsion energy^{87,102}:

$$V_{ii}^{\text{exch}} = A_{ii}^{\text{exch}} P(Br_{ii}) \exp(-Br_{ii}). \quad (3.6)$$

Here, r_{ii} represents an interatomic distance, and A_{ii}^{exch} indicates a proportionality constant that is typically fit to calculated values of the exchange-repulsion energy. The only approximations thus far are the use of the overlap model and the simplified asymptotic form of the atomic charge density.

For the general case of two hetero-atoms, substitution of Eq. (3.4) into Eq. (3.2) yields the more complicated expression^{7,8}

$$\begin{aligned} S_\rho^{ij} &= \frac{16\pi D_i D_j \exp(-\{B_i + B_j\}r_{ij}/2)}{(B_i^2 - B_j^2)^3 r_{ij}} \times \\ &\left[\left(\frac{B_i - B_j}{2} \right)^2 \left(\exp \left(\{B_i - B_j\} \frac{r_{ij}}{2} \right) - \exp \left(-\{B_i - B_j\} \frac{r_{ij}}{2} \right) \right) \right. \\ &\quad \times \left(\left(\frac{B_i + B_j}{2} \right)^2 r_{ij}^2 + (B_i + B_j)r_{ij} + 2 \right) \\ &\quad - \left(\frac{B_i + B_j}{2} \right)^2 \exp \left(\{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left(\left(\frac{B_i - B_j}{2} \right)^2 r_{ij}^2 - (B_i - B_j)r_{ij} + 2 \right) \\ &\quad \left. + \left(\frac{B_i + B_j}{2} \right)^2 \exp \left(-\{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left(\left(\frac{B_i - B_j}{2} \right)^2 r_{ij}^2 + (B_i - B_j)r_{ij} + 2 \right) \right], \end{aligned} \quad (3.7)$$

which is too cumbersome to serve as a practical force field functional form. However, since the above expression reduces to Eq. (3.5) in the limit $B_i = B_j$, and because $|B_i -$

$|B_j|$ is small for most atom pairs, we have found that Eq. (3.7) may be approximated using Eq. (3.5) with an *effective* atomic exponent B . An expansion of Eq. (3.7) about $B_i = B_j$ suggests that this effective exponent should be given by the arithmetic mean, $B_{ij} = \frac{1}{2}(B_i + B_j)$. However, a Waldman-Hagler style analysis¹⁰³ (Section 3.A) suggests instead that a more suitable exponent is given by the geometric mean combination rule,

$$B = B_{ij} \equiv \sqrt{B_i B_j}. \quad (3.8)$$

As shown in the Supporting Information of Ref. 104, this approximate overlap model (Eq. (3.5) and Eq. (3.8)) is of comparable accuracy to the exact overlap from Eq. (3.7). Thus the density overlap and (force field) exchange energies of arbitrary hetero-atoms take the simple forms

$$S_\rho^{ij} = D_{ij} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \quad (3.9)$$

$$D_{ij} = \pi D_i D_j B_{ij}^{-3} \quad (3.10)$$

$$P(B_{ij}, r_{ij}) = \frac{1}{3}(B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \quad (3.11)$$

and

$$V_{ij}^{\text{exch}} = A_{ij}^{\text{exch}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}). \quad (3.12)$$

Due to the connection with the overlap between two s-type Slater orbitals, we refer to Eq. (3.12) as the Slater functional form. Note that this expression reduces to the standard Born-Mayer function by making the further approximation $P(B_{ij}, r_{ij}) = 1$, although it is known^{87,105} that this is a poor approximation with the B_{ij} as defined above. Instead, as we shall demonstrate in Section 3.4, the exponents B_{ij} need to be scaled for accurate use with a Born-Mayer functional form.

Variants of the polynomial pre-factor from Eq. (3.9) have previously been recognized and used in intermolecular interaction models.^{65,102,105} Early work by Buckingham⁶⁵ hypothesized that the functional form of Eq. (3.12) would be more accurate

than the Born-Mayer form, though no attempt was made to provide a closed-form expression for P . More recent potentials have incorporated a low-order polynomial into the exchange repulsion term, either by direct parameterization^{106–110} or indirectly by fitting the exchange to S_ρ/r^2 rather than to S_ρ itself.^{86,87,111} Kita et al. have derived (but not tested) Eq. (3.6) for the homoatomic case.¹¹¹ Recently, and most similar to the spirit of the present Chapter, York and co-workers have derived a model based upon the overlap of Slater-type orbitals for use in QM/MM simulations, yielding an expression identical to Eq. (3.7).^{112–114} Those authors treated D_i and D_j as empirical fitting parameters and estimated atomic exponents (B_i and B_j) via atomic-charge dependent functions. In contrast, we will demonstrate that utilization of the far simpler functional form from Eq. (3.12), in conjunction with exponents calculated from analysis of the first-principles molecular electron density, yields much higher computational efficiency and simplifies the parameterization process without significant loss of accuracy.

For an arbitrary pair of interacting atoms, A_{ij}^{exch} can be obtained by fitting to calculated exchange-repulsion energies. However, assuming that the overlap proportionality factor K_{ij} is a universal constant (or, alternatively, separable with $K_{ij} = K_i K_j$), then

$$A_{ij}^{\text{exch}} = \left(K_i \sqrt{\frac{\pi}{B_i^3}} D_i \right) \left(K_j \sqrt{\frac{\pi}{B_j^3}} D_j \right) \equiv A_i^{\text{exch}} A_j^{\text{exch}}, \quad (3.13)$$

thus providing a combination rule for heteroatomic interaction in terms of purely atomic quantities. The universality and separability of K_{ij} are, at present, empirically rather than theoretically justified.^{51,115,116} The A_i^{exch} can then be obtained, for example, by a straightforward fitting of calculated ab initio *homoatomic* exchange-repulsion energies.

3.2.2 Models for other short-range interactions between isolated atoms

Beyond the exchange-repulsion, the density-overlap model may also be used to model other short-range interaction components, such as the electrostatic charge penetration energy and the short-range induction and dispersion energies (that is, the portion modulated by charge overlap). Indeed, it has been demonstrated that the electrostatic charge penetration energy is approximately proportional to the exchange-repulsion energy, and consequently to the charge density overlap,^{51,58} which has provided a successful basis for modeling the electrostatic charge penetration energy.^{117,118} While the relation between short-range induction and charge overlap is less clear, recent results have demonstrated that the charge-transfer energy, which is the dominant short-range component of the induction energy,¹¹⁹ is approximately proportional to the first-order exchange energy,^{85,120} and prior work has successfully used the overlap hypothesis to describe the short-range induction.^{51,117,118} We therefore model the electrostatic charge penetration and short-range induction interactions as

$$V_{ij}^{\text{pen}} = A_{ij}^{\text{pen}} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij}) \quad (3.14)$$

and

$$V_{ij}^{\text{ind,sr}} = A_{ij}^{\text{ind}} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij}). \quad (3.15)$$

Aside from the pre-factors A_{ij} , these expressions are identical to that for the exchange-repulsion term.

The behavior of the dispersion interaction at short distances poses a special challenge. In order to model the short-range dispersion and to resolve the unphysical, mathematical divergence of the $1/r^n$ terms as $r \rightarrow 0$, Tang and Toennies have shown that the terms in the dispersion expansion should be damped using an

appropriate incomplete gamma function

$$f_n(x) = 1 - e^{-x} \sum_{k=0}^n \frac{(x)^k}{k!} \quad (3.16)$$

$$x = -\frac{d}{dr} [\ln V^{\text{exch}}(r)] \quad r \quad (3.17)$$

that accounts for both exchange and charge penetration effects.^{121,122} Note that the form of this damping factor depends on the model used for exchange repulsion. For the Slater functional form (Eq. (3.12)),

$$x_{\text{Slater}} = B_{ij} r_{ij} - \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij} r_{ij} + 3} r_{ij}. \quad (3.18)$$

Alternatively, if we replace the Slater functional form with the less accurate Born-Mayer expression, x simplifies to the result originally given by Tang and Toennies:

$$x_{\text{Born-Mayer}} = B_{ij} r_{ij}. \quad (3.19)$$

3.2.3 Models for short-range interactions between molecules

The overlap repulsion hypothesis can be extended to molecules^{51,115,123–125} by writing the molecular density ρ_I as a superposition of atomic densities

$$\rho_I(\mathbf{r}) = \sum_{i \in I} \rho_i(\mathbf{r}) \quad (3.20)$$

where i represents an atom in molecule I . In this case,

$$V_{IJ}^{\text{exch}} = \sum_{i \in I} \sum_{j \in J} V_{ij}^{\text{exch}} \quad (3.21)$$

$$V_{ij}^{\text{exch}} = K_{ij} S_{\rho}^{ij} = \int \rho_i(\mathbf{r}) \rho_j(\mathbf{r}) d^3 \mathbf{r}. \quad (3.22)$$

Note that the form of Eq. (3.22) is identical to the corresponding expression between isolated atoms, but requires partitioning of the molecular charge density into atom-in-molecule densities, ρ_i , each decaying according to an effective atom-in-molecule density decay exponent, B_i .

In principle, such atom-in-molecule exponents could be estimated from the ionization potentials of the corresponding isolated atoms,^{102,117} but this approach neglects the influence of the molecular environment. A more appealing possibility is to directly evaluate the atom-in-molecule densities via partitioning of the calculated monomer densities. Density partitioning has not yet (to our knowledge) been applied in the context of the overlap model to solve for Eq. (3.22), however several successful efforts in force field development have recently relied on an atoms-in-molecule approach in order to obtain accurate scaling relationships for intermolecular force field parameters.^{126–128} In particular, Cole et al. utilized a density-derived electrostatic and chemical (DDEC) partitioning scheme^{129,130} to generate Lennard-Jones dispersion and short-range repulsion parameters, though the latter parameters were calculated implicitly by enforcing the coincidence of the potential minimum and the calculated atomic radius.

While no unique atom-in-molecule density partitioning scheme exists, an ideal approach should yield atom-in-molecule densities that strongly resemble those of isolated atoms, e.g. maximally spherical and asymptotically exponential.^{58,131–133} The recently developed iterated stockholder partitioning of Lillestolen and Wheatley obeys this first important constraint of sphericity.^{134,135} As a non-trivial extension of the original Hirshfeld method,¹³⁶ iterated stockholder atom (ISA) densities are defined as

$$\rho_i(\mathbf{r}) = \rho_I(\mathbf{r}) \frac{w_i(\mathbf{r})}{\sum_{a \in I} w_a(\mathbf{r})} \quad (3.23)$$

where the converged shape functions $w_i(\mathbf{r})$ are spherical averages of the atomic

densities $\rho_i(\mathbf{r})$:

$$w_i(\mathbf{r}) = \langle \rho_i(\mathbf{r}) \rangle_{\text{sph}}. \quad (3.24)$$

This formulation ensures, by construction, that the sum of atomic densities reproduces the overall molecular density. Furthermore, the maximally spherical nature of the atom-in-molecule densities naturally facilitates a description of short-range interactions via a simple isotropic site-site model.

Misquitta et al. have developed a rapidly convergent implementation of the ISA procedure (BS-ISA⁵⁸) using a basis set expansion which, in addition to exhibiting good convergence with respect to basis set, also leads to asymptotically-exponential atomic densities. Consequently, the BS-ISA method is our preferred density partitioning scheme. As an example, the spherically-averaged atomic densities for acetone are shown in Fig. 3.1. For simplicity, and because a full treatment of the anisotropy is beyond the scope of this Chapter, we subsequently refer to the spherically-averaged atomic densities (i.e. the shape functions, $w_i(\mathbf{r})$) as atomic or atom-in-molecule densities.

From Fig. 3.1 we see that the ISA atomic shape functions (that is, the spherically-averaged ISA atoms-in-molecule density) decay exponentially outside the core region. However, note that the exponents governing the spherical density decay, B_i^{ISA} , differ from those of the free atoms. The ISA densities have been observed to account for electron movement in the molecule, and the consequent density changes brought about by this movement tend to be manifested in the region of the density tails.⁵⁸ The ISA exponents can be obtained by a weighted least-squares fit to the BS-ISA atomic density (see Section 3.3 for details), with the resulting fitted atomic densities shown in Fig. 3.1. Note that even a single exponential is remarkably successful in reproducing the entirety of the valence atomic density.

Given these fitted ISA exponents, we can now apply our short-range interaction

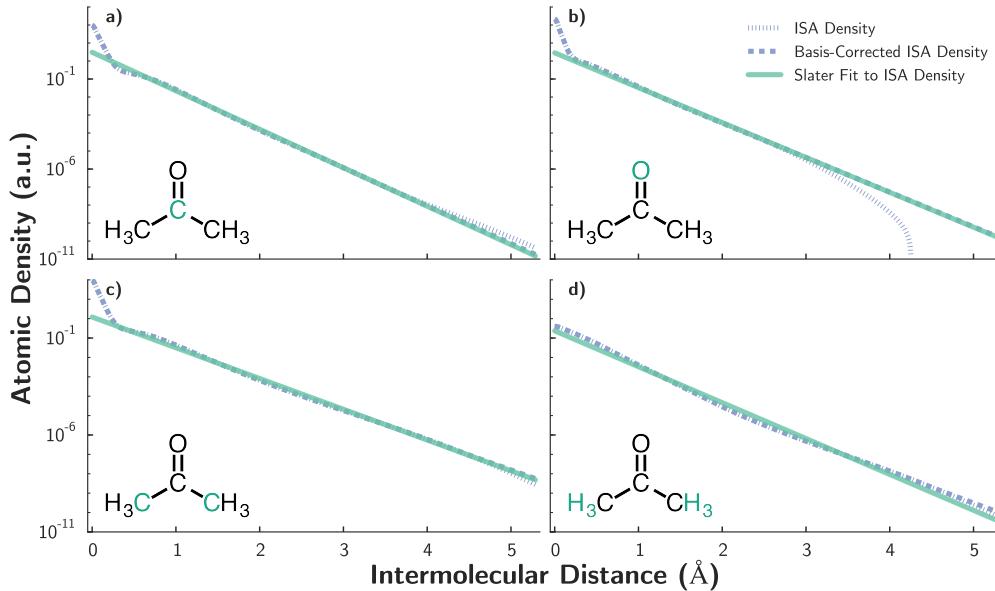


Figure 3.1: BS-ISA and fitted shape functions for each atom type in acetone: a) carbonyl carbon, b) oxygen, c) methyl carbon, d) hydrogen. BS-ISA shape functions (dotted line) for each atom type have been obtained at a PBE0/aug-cc-pVTZ level of theory. A modified BS-ISA shape function (dashed line) corrects the tail-region of the BS-ISA function to account for basis set deficiencies in the BS-ISA algorithm. A single Slater orbital of the form $D_i^{\text{ISA}} \exp(-B_i^{\text{ISA}}r)$ (solid line) is fit to the basis-corrected BS-ISA shape function, and the obtained B_i^{ISA} value is used as an atomic exponent in the functional form of Aniso-Iso FF. Results for acetone are typical of molecules studied in this Chapter.

formalism to polyatomics,

$$\begin{aligned}
 V^{sr} &= \sum_{ij} A_{ij}^{sr} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\
 P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\
 A_{ij}^{sr} &= A_i^{sr} A_j^{sr} \\
 B_{ij} &= \sqrt{B_i^{\text{ISA}} B_j^{\text{ISA}}}
 \end{aligned} \tag{3.25}$$

where the molecular short-range energy is now a sum of atom-atom contributions.

In conjunction with appropriately damped atomic dispersion (Eqs. (3.16) and (3.18)), Eq. (3.25) completely defines our new short-range force field. We refer to this new functional form and set of atomic exponents as the Aniso-Iso FF.

3.3 Computational Methods

To evaluate the Slater-ISA FF against conventional Born-Mayer and/or Lennard-Jones models, we compare the ability of each resulting short-range force field to reproduce benchmark ab initio intermolecular interaction energies for a collection of representative dimers. Such a metric is directly relevant for ab initio force field development. Even for an empirically-parameterized force field, however, fidelity to an accurate ab initio potential should be well correlated with the highest level of accuracy and transferability achievable with a given short-range methodology.

We have developed the Slater-ISA FF, Born-Mayer, and Lennard-Jones force fields using benchmark energies calculated using the symmetry-adapted perturbation theory based on density-functional theory (DFT-SAPT or SAPT(DFT)^{137–145}). DFT-SAPT provides interaction energies that are comparable in accuracy to those from CCSD(T) and which are rigorously free from basis set superposition error.^{50,146} Additionally, at second-order, DFT-SAPT also provides an explicit interaction energy decomposition into physically-meaningful contributions: the electrostatic, exchange-repulsion, induction, and dispersion energies. This decomposition is vital to the development of models as it allows the development of separate terms for each type of short-range interaction. Terms of third and higher order are estimated using the δHF correction¹⁴⁷ which contains mainly higher-order induction contributions. Following prior work,^{117,131} and for the purposes of fitting to the DFT-SAPT data, we keep the second-order induction term and the δHF term separate.

Since the Slater-ISA and Born-Mayer force fields describe only short-range interactions (i.e. those terms which are modulated by the overlap of the monomer electron densities), they must both be supplemented with additional long-range terms that describe the electrostatic, polarization, and dispersion interactions. Here

we have chosen a long-range potential of the form

$$V_{lr} = V_{multipole} + V_{dispersion} + V_{pol} \quad (3.26)$$

where

$$V_{multipole} = \sum_{ij} \sum_{tu} Q_t^i T_{tu} Q_u^j \quad (3.27)$$

includes distributed multipole contributions from each atom up to quadrupoles,

$$V_{dispersion} = - \sum_{ij} \sum_{n=3}^6 \frac{C_{ij,2n}}{r_{ij}^{2n}} \quad (3.28)$$

describes isotropic dispersion, and V_{pol} is the polarization energy modeled by Drude oscillators^{148,149} as in Ref. 117. The accuracy of each of these terms is expected to minimize errors in the long-range potential, simplifying the comparison between short-range force field functional forms. Nonetheless, we expect that our results will be qualitatively insensitive to the particular choice of long-range force field and acknowledge that simpler alternatives may be preferred for the development of highly efficient simulation potentials. In the case of the Lennard-Jones force field, we replace Eq. (3.28) with the simple $C_{ij,6}/r_{ij}^6$ dispersion term that is standard to the Lennard-Jones model.

We used a test set consisting of one atom (argon) and 12 small organic molecules (see Fig. 3.2) from which dimer potentials could be generated (we will use the term ‘dimer’ to mean two, potentially dissimilar, interacting molecules or atoms), yielding 91 dimer combinations (13 homo-monomeric, 78 hetero-monomeric). This wide range of systems allowed us to evaluate both the accuracy and transferability of the Slater-ISA model compared to conventional Born-Mayer and/or Lennard-Jones models.

A detailed description of this overall methodology is provided below.

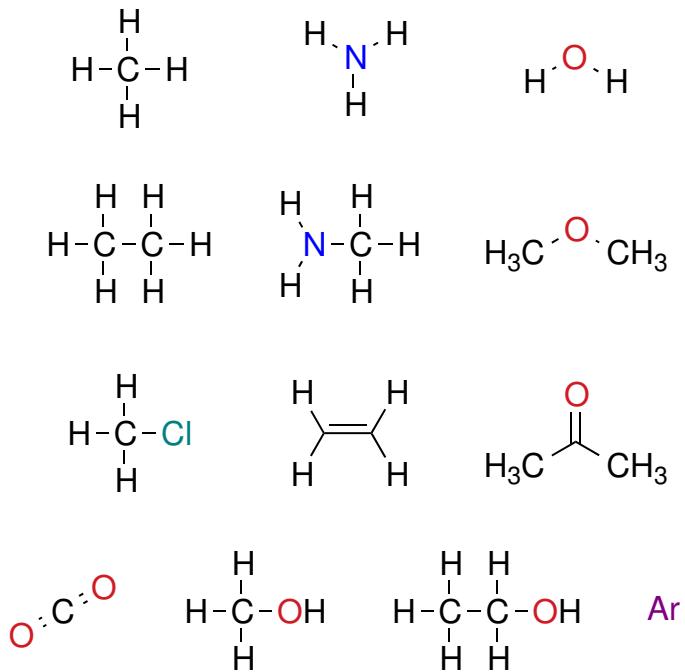


Figure 3.2: The 13 small molecules included in the 91 dimer (13 homomonomeric, 78 heteromonomeric) test set. Cartesian geometries for all of these molecules are given in Section A.1.

3.3.1 Construction of the 91 dimer test set

Monomer geometries for each of the 13 small molecules were taken from the experimental NIST [CCCBDB] database¹⁵⁰ and can be found in Section A.1. For acetone and methyl amine, experimental geometries were unavailable, and thus the computational NIST [CCCBDB] database was used to obtain geometries at a high level of theory (B3LYP/aVTZ for acetone, CCSD(T)/6-311G* for methyl amine). For each of the 91 dimers, a training set was constructed using DFT-SAPT (PBE0/AC) interaction energies calculated at 1000 quasi-random dimer configurations. These configurations were generated using Shoemaker's algorithm,¹⁵¹ subject to the constraint that the nearest atom pairs be separated by between 0.75 and 1.3 of the sum of their van der Waals radii. This ensured adequate sampling of the potential energy surface in the region of the repulsive wall. The DFT-SAPT interaction energies

were evaluated using an asymptotically corrected PBE0 functional (PBE0/AC) with monomer vertical (first) ionization potentials computed using the Δ -DFT approach at a PBE0/aVTZ level of theory. Unless otherwise noted, all DFT-SAPT calculations used an aVTZ basis set in the dimer-centered form with midbond functions (the so-called DC+ form), and were performed using the MOLPRO2009 software suite.¹⁵² The midbond set consisted of a 5s3p1d1f even-tempered basis set with ratios of 2.5 and centered at $\zeta = 0.5, 0.5, 0.3, 0.3$ for the s,p,d, and f shells, respectively. This set was placed near the midpoint of the centers of mass of the two interacting monomers.

A small fraction of DFT-SAPT calculations exhibited unphysical energies, which were attributed to errors in generating the optimized effective potential used during the DFT-SAPT (PBE0/AC) calculations; these points were removed from the test set.

3.3.2 BS-ISA Calculations

BS-ISA atomic densities were obtained using CamCASP 5.8¹⁵³⁻¹⁵⁵ following the procedure of Misquitta et al.⁵⁸ For the BS-ISA calculations, an auxiliary basis was constructed from an RI-MP2 aVTZ basis set with s-functions replaced by the ISA-set2 supplied with the CamCASP program; CamCASP's ISA-set2 basis was also used for the ISA basis set.⁵⁸ A custom ISA basis set for Ar was used (even tempered, $n_{\min} = -2, n_{\max} = 8$)⁵⁸ as no published basis was available. BS-ISA calculations were performed with the A+DF algorithm, which allows the ISA functional to be mixed with some fraction, ζ , of the density-fitting functional. Following the recommendations of Misquitta et al.⁵⁸, we have used $\zeta = 0.1$ for the multipole moment calculations, and $\zeta = 0.9$ for the density partitioning used to determine the B_{ij} coefficients.

3.3.3 Determination of B_i^{ISA}

The BS-ISA-derived atomic exponents, B_i^{ISA} , were obtained from a weighted least-squares fit to the spherically averaged BS-ISA atomic densities (shape functions),

$w_i(\mathbf{r})$. In some cases, numerical instabilities and basis-set limitations of the BS-ISA procedure yielded densities that exhibited non-exponential asymptotic behavior.⁵⁸ To correct for these unphysical densities, we extrapolated the exponential decay of the valence region to describe the BS-ISA tails also. Details of this procedure can be found in Section 7.B. The ISA atom-in-molecule exponents were then derived via a log-weighted fit to the tail-corrected shape-functions $w^a(\mathbf{r})$ for densities within the cutoff $10^{-2} > w^a > 10^{-20}$ a.u. This region was chosen to reproduce the charge density most accurately in the valence regimes most likely to be relevant to intermolecular interactions.

3.3.4 Force Field Functional Forms and Parameterization

The general structure of the force fields V_{FF} for both the Slater-ISA FF and the Born-Mayer-type models are given by the following equations:

$$\begin{aligned}
 V_{FF} &= \sum_{ij} V_{ij}^{\text{exch}} + V_{ij}^{\text{elst}} + V_{ij}^{\text{ind}} + V_{ij}^{\delta\text{HF}} + V_{ij}^{\text{disp}} \\
 V_{ij}^{\text{exch}} &= A_{ij}^{\text{exch}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\
 V_{ij}^{\text{elst}} &= -A_{ij}^{\text{elst}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + \sum_{tu} Q_t^i T_{tu} Q_u^j \\
 V_{ij}^{\text{ind}} &= -A_{ij}^{\text{ind}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(2)} \\
 V_{ij}^{\delta\text{HF}} &= -A_{ij}^{\delta\text{HF}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(3-\infty)} \\
 V_{ij}^{\text{disp}} &= - \sum_{n=3}^6 f_{2n}(x) \frac{C_{ij,2n}}{r_{ij}^{2n}} \\
 A_{ij} &= A_i A_j \\
 C_{ij,2n} &= \sqrt{C_{i,n} C_{j,n}} \\
 f_{2n}(x) &= 1 - e^{-x} \sum_{k=0}^{2n} \frac{(x)^k}{k!}
 \end{aligned} \tag{3.29}$$

For the Slater-ISA FF:

$$\begin{aligned} B_i &= B_i^{\text{ISA}} \\ B_{ij} &= \sqrt{B_i B_j} \\ P(B_{ij}, r_{ij}) &= \frac{1}{3}(B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\ x &= B_{ij} r_{ij} - \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij} r_{ij} + 3} r_{ij} \end{aligned} \quad (3.30)$$

For all Born-Mayer type models:

$$\begin{aligned} P(B_{ij}, r_{ij}) &= 1 \\ x &= B_{ij} r_{ij} \end{aligned} \quad (3.31)$$

For the Born-Mayer-IP FF:

$$\begin{aligned} B_i &\equiv B_i^{\text{IP}} = 2\sqrt{2I_i} \\ B_{ij} &= \frac{B_i B_j (B_i + B_j)}{B_i^2 + B_j^2} \end{aligned} \quad (3.32)$$

For the Born-Mayer-sISA FF:

$$\begin{aligned} B_i &= 0.84 B_i^{\text{ISA}} \\ B_{ij} &= \sqrt{B_i B_j} \end{aligned} \quad (3.33)$$

Of the parameters in these force fields, only the coefficients A_i were fit to reproduce DFT-SAPT dimer energies (details below). All other force field parameters were derived from first-principles atom or atom-in-molecule properties. Exponents for the Slater-ISA FF and the Born-Mayer-sISA FF were derived from BS-ISA calculations, while exponents for the Born-Mayer-IP FF were determined from vertical ionization potentials of the isolated atoms. Dispersion coefficients ($C_{ij,2n}$) were either used directly from Ref. 117 or were parameterized using analogous methods

in the case of argon. Distributed multipoles Q_t^i for each system were obtained from the BS-ISA-based distributed multipoles scheme (ISA-DMA)⁵⁸, with the expansion truncated to rank 2 (quadrupole). Note that here, $t = 00, 10, \dots, 22s$ denotes the rank of the multipole in the compact notation of Stone⁴⁸. (In addition to rank 2 ISA-DMA multipoles, we also tested the use of DMA4 multipoles⁵⁷ as well as the use of rank 0 charges obtained from the rank truncation or transformation¹⁵⁶ of either ISA-DMA or DMA4 multipoles; the effect of including a Tang-Toennies damping factor^{117,121} was studied in all cases. Each of these alternative long-range electrostatic models proved either comparably or less accurate for both the Slater-ISA FF and the Born-Mayer-IP FF in terms of their ability to reproduce the DFT-SAPT electrostatic energy, and are not discussed further.) Long-range polarization (V_{shell}) was modeled using Drude oscillators in a manner identical to Ref. 117. As in our prior work, during parameterization, the Drude energy was partitioned into 2nd ($V_{\text{pol}}^{(2)}$) and higher order ($V_{\text{pol}}^{(3-\infty)}$) contributions, where $V_{\text{pol}}^{(2)}$ is the Drude oscillator energy due to static charges (excluding intra-molecular contributions), and $V_{\text{pol}}^{(3-\infty)}$ is the difference between the fully converged Drude energy, V_{shell} , and $V_{\text{pol}}^{(2)}$. Force field parameters for all homo-monomeric systems are located in the Supporting Information of Ref. 104.

A weighted least-squares fitting procedure was used to fit A_i parameters to the benchmark DFT-SAPT (PBE0/AC) interaction energies on a component-by-component basis. That is, four separate optimizations¹¹⁷ were performed to directly fit V^{exch} , V^{elst} , V^{ind} , and $V^{\delta^{\text{HF}}}$ to, respectively, the following DFT-SAPT quantities (notation as in Ref. 140):

$$\begin{aligned} E^{\text{exch}} &\equiv E_{\text{exch}}^{(1)} \\ E^{\text{elst}} &\equiv E_{\text{pol}}^{(1)} \\ E^{\text{ind}} &\equiv E_{\text{ind}}^{(2)} + E_{\text{ind-exch}}^{(2)} \\ E^{\delta^{\text{HF}}} &\equiv \delta(\text{HF}). \end{aligned} \tag{3.34}$$

For V^{disp} , no parameters were directly fit to the DFT-SAPT dispersion,

$$E^{\text{disp}} \equiv E_{\text{disp}}^{(2)} + E_{\text{disp-exch}}^{(2)}, \quad (3.35)$$

but were instead obtained solely from monomer properties as described above. Finally, note that no parameters were directly fit to the total DFT-SAPT energy,

$$E_{\text{int}} = E^{\text{exch}} + E^{\text{elst}} + E^{\text{ind}} + E^{\delta^{\text{HF}}} + E^{\text{disp}}, \quad (3.36)$$

for either the Slater-ISA FF or the Born-Mayer-IP FF. Rather, V_{FF} was calculated according to Eq. (3.29).

Data points for each fit were weighted using a Fermi-Dirac functional form given by

$$w_i = \frac{1}{\exp((-E_i - \mu_{\text{eff}})/kT) + 1}, \quad (3.37)$$

where E_i is the reference energy and μ_{eff} and kT were treated as adjustable parameters. The parameter kT , which sets the energy scale for the weighting function, was taken to be $kT = \lambda|E_{\text{min}}|$; here E_{min} is an estimate of the global minimum well depth. Unless otherwise stated, we have used $\lambda = 2.0$ and $\mu_{\text{eff}} = 0.0$. These defaults were chosen to minimize overall average attractive RMSE for all 91 dimer sets. Increases or decreases in the λ factor correspond to the weighting of more or fewer repulsive configurations, respectively.

In the case of Lennard-Jones, the standard Lennard-Jones functional form was used for the van der Waals terms, with Coulomb and polarization terms modeled exactly as for the Slater-ISA FF:

$$V_{\text{FF}}^{\text{LJ}} = \sum_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij,6}}{r_{ij}^6} + V_{\text{pol}} + \sum_{tu} Q_t^i T_{tu} Q_u^j \quad (3.38)$$

Lorentz-Berthelot combination rules were used to obtain heteroatomic A_{ij} and C_{ij} parameters. Unlike with the Slater-ISA FF and Born-Mayer models, $V_{\text{FF}}^{\text{LJ}}$ was fit to

the total DFT-SAPT (PBE0/AC) energy, with A_{ij} and $C_{ij,6}$ as fitting parameters. The weighting function from Eq. (3.37) was used in fitting.

3.3.5 Potential Energy Surface Scans

In order to visually assess fit quality, representative one-dimensional scans of the potential energy surface were calculated for several dimer pairs along low-energy dimer orientations. For each dimer pair, the minimum energy configuration of the 1000 random dimer points was selected as a starting configuration, and additional dimer configurations (not necessarily included in the original 1000 points) were generated by scanning along some bond vector. In the case of the ethane dimer, two carbon atoms (one on each monomer) were used; for acetone, the carbonyl carbon on each monomer defined the bond vector.

3.3.6 Molecular Simulations

All bulk simulations were run using OpenMM release version 7.0.¹⁵⁷ Enthalpies of vaporization were computed from

$$\Delta H_{vap} = (E_{\text{pot}}(g) + RT) - E_{\text{pot}}(l)$$

where $E_{\text{pot}}(g)$ and $E_{\text{pot}}(l)$ were determined from NVT simulations at the experimental gas and liquid densities, respectively. Calculated liquid densities were determined from NPT simulations. In all cases, the OPLS/AA force field was used for the intramolecular potential.¹⁵⁸ All simulations used a Langevin integrator with a 0.5 fs time step and a 1 ps⁻¹ friction coefficient; NPT simulations used a Monte Carlo barostat with a trial volume step every 5th move. Periodic boundary conditions, particle-mesh Ewald, and a non-bonding cutoff of 1.2nm with added long-range corrections were used to simulate a unit cell of 222 molecules. After an equilibration period of at least 600ps, simulation data was gathered from production runs lasting at least 200ns.

3.4 Results and Discussion

The Slater-ISA methodology for short-range intermolecular interactions has been derived from a simple but rigorous physical model of overlapping monomer electron densities. In practice, this approach differs from the conventional Born-Mayer approach in both the choice of the short-range functional form (with the latter omitting the polynomial pre-factor) and the source of the exponents (with the former derived from ISA analysis of the monomer density). Our principal goal is to examine the influence of these modifications on the accuracy and transferability of the resulting force fields.

We initially benchmark the Slater-ISA FF against a conventional Born-Mayer potential, Born-Mayer-IP FF. The latter approach has been used extensively in prior work,^{80,117} and both approaches use identical numbers of fitted parameters. Following prior work, combination rules for the Born-Mayer-IP FF are as in Ref. 117. (We have tested the effect of using a geometric mean for the Born-Mayer-IP FF; results do not differ qualitatively from those presented below.) Owing to its popularity, we also compare the Slater-ISA FF to a Lennard-Jones functional form (LJ FF).

We first assess the accuracy of the Slater-ISA FF, Born-Mayer-IP FF, and LJ FF against benchmark ab initio intermolecular interaction energies and experimental 2nd virial coefficients, enthalpies of vaporization, and liquid densities. We next examine parameter transferability, assessing the extent to which parameters from pure homo-monomeric systems can be re-used (without further optimization) to describe mixed interactions. To assess parameter robustness, we also study the sensitivity of each methodology to changes in the weighting function (Eq. (3.37)). Finally, we explore the application of BS-ISA-derived exponents within the Born-Mayer functional form as a straightforward method for simplifying the parameterization (and potentially increasing the accuracy) of a wide variety of standard ab initio and empirically-parameterized force fields.

3.4.1 Accuracy: Comparison with DFT-SAPT

For each of the 91 molecule pairs described in the Computational Methods section, parameters for the Slater-ISA FF, Born-Mayer-IP FF, and LJ FF were fit to reproduce DFT-SAPT (PBE0/AC) interaction energies calculated for a set of 1000 dimer configurations. These 91,000 total configurations and corresponding DFT-SAPT energies are collectively referred to as the ‘91 dimer test set’. As a primary indication of accuracy, root-mean-square errors (RMSE) and mean signed errors (MSE), both with respect to DFT-SAPT, were computed for each methodology and for each dimer pair. Because these RMSE and MSE are dominated by repulsive contributions, and owing to the thermodynamic importance of attractive configurations, so-called ‘attractive RMSE/MSE’ were also computed by excluding net repulsive configurations (as measured by the DFT-SAPT total energy). The overall RMSE/MSE for all 91 dimers were then averaged to produce one ‘characteristic RMSE/MSE’ for the entire test set. Since these errors varied considerably in magnitude depending on the dimer in question, this overall average was taken in the geometric mean sense. (Results with an arithmetic mean do not differ qualitatively). Note that when computing the characteristic MSE, only the magnitude of each MSE, $\|\text{MSE}\|$, was considered.

Characteristic RMSE and $\|\text{MSE}\|$ across the 91 dimer test set are shown in Fig. 3.3 and Table 3.1. Overall, the Slater-ISA FF exhibits smaller errors compared to the Born-Mayer-IP FF. On average, the characteristic total energy RMSE for the Slater-ISA FF decrease by 33% relative to the Born-Mayer-IP FF. Even excluding repulsive configurations (dominated by short-range interactions), errors for the Slater-ISA FF are lower by 11% compared to the Born-Mayer-IP FF, demonstrating modest gains in accuracy even over the most energetically-relevant regions of the potential. A more detailed analysis of each of the 91 pairs of molecules shows that in an overwhelming 93% of such cases, force fields derived from the Slater-ISA method have smaller RMSEs compared to their Born-Mayer-IP counterparts (70% if only attractive configurations are considered). Regardless of the metric used, the Slater-ISA FF produces force fields with higher fidelity to the underlying benchmark interaction energies.

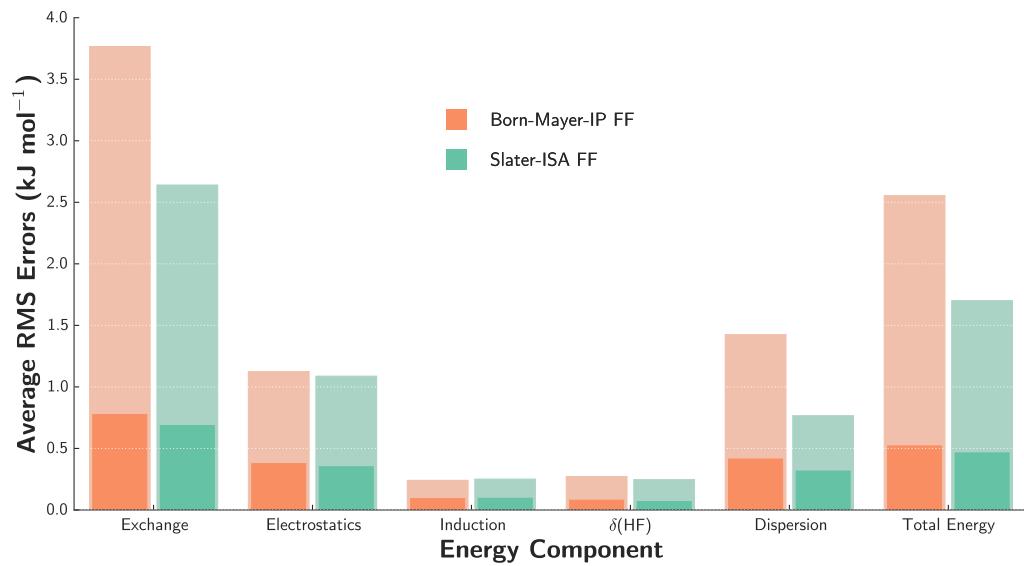


Figure 3.3: Characteristic RMSE (as described in the main text) for the Born-Mayer-IP FF (orange) and the Slater-ISA FF (green) over the 91 dimer test set. The translucent bars represent total RMSE for each energy component, while the smaller solid bars represent ‘Attractive’ RMSE, in which repulsive points have been excluded.

| Component | Dimer-Specific Fits | | | Transferable Fits | | |
|---------------------|--|---|----------------------------------|--|---|----------------------------------|
| | Slater-ISA FF (kJ mol ⁻¹) | Born-Mayer-IP FF (kJ mol ⁻¹) | LJ FF (kJ mol ⁻¹) | Slater-ISA FF (kJ mol ⁻¹) | Born-Mayer-IP FF (kJ mol ⁻¹) | LJ FF (kJ mol ⁻¹) |
| Exchange | 2.641 (0.686) | 3.766 (0.775) | — | 2.718 (0.720) | 4.033 (0.836) | — |
| Electrostatics | 1.087 (0.351) | 1.126 (0.377) | — | 1.134 (0.351) | 1.231 (0.378) | — |
| Induction | 0.251 (0.095) | 0.241 (0.093) | — | 0.278 (0.101) | 0.265 (0.098) | — |
| δ HF | 0.246 (0.068) | 0.272 (0.079) | — | 0.274 (0.076) | 0.304 (0.081) | — |
| Dispersion | 0.766 (0.317) | 1.425 (0.414) | — | 0.766 (0.317) | 1.425 (0.414) | — |
| Total Energy | | | | | | |
| RMSE | 1.701 (0.464) | 2.554 (0.520) | 1.984 (0.603) | 1.650 (0.456) | 2.698 (0.555) | 2.054 (0.640) |
| $\ MSE\ $ | 0.216 (0.057) | 0.539 (0.127) | 0.322 (0.345) | 0.175 (0.051) | 0.569 (0.112) | 0.311 (0.368) |

Table 3.1: Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Slater-ISA FF, Born-Mayer-IP FF and LJ FF. For the total energy, both characteristic RMSE and MSE have been shown, with only the magnitude of the MSE, $\|MSE\|$, displayed. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ($E_{int} < 0$), are shown in parentheses to the right of the total RMS errors; ‘attractive’ $\|MSE\|$ are likewise displayed for the total energy. As discussed in Section 3.4.3, the ‘Dimer-Specific Fits’ refer to force fields whose parameters have been optimized for each of the 91 dimers separately, whereas the ‘Transferable Fits’ refer to force fields whose parameters have been optimized for the 13 homodimers and then applied (without further optimization) to the remaining 78 mixed systems. Unless otherwise stated, a default weighting function of $\lambda = 2.0$ (see Eq. (3.37)) has been used for all force fields in this Chapter.

It is also instructive to consider each energy component individually. As might be expected, improvements in the description of E^{exch} are pronounced, with the characteristic RMSE from the Slater-ISA FF being 30% smaller than that from the Born-Mayer-IP FF. Examining each dimer pair separately (see Section 3.B for homomeric fits, representative of the entire test set) we also find that, in general, the Slater-ISA FF is far better at reproducing *trends* in the exchange energy compared to the Born-Mayer-IP FF. This qualitative result is also reflected in the smaller $\|\text{MSE}\|$ values for the Slater-ISA FF as compared to the Born-Mayer-IP FF. Nevertheless, there remains a fair amount of scatter in the exchange energies for several dimer pairs, particularly for molecules with exposed lone pairs or delocalized π systems. We hypothesize that this scatter is due to a breakdown of the isotropic approximation made in the Theory section, a conclusion supported by observations on the pyridine dimer system recently made by some of us.¹²⁰ It is therefore quite possible that the observed 30% RMSE reduction underestimates the true error reduction that might be observed if such anisotropy were accounted for.

From Fig. 3.3, we see that the dispersion energy model from the Slater-ISA FF is also a substantial improvement; for dispersion, characteristic RMSE are 46% smaller for the Slater-ISA FF compared to the Born-Mayer model. This should not be a counter-intuitive result: while both potentials use identical dispersion coefficients, they differ in the damping model used. In the Born-Mayer-IP FF, the standard Tang–Toennies damping model is employed, and the damping parameters only depend on free atom ionization potentials; in the Slater-ISA FF, on the other hand, the damping parameters are obtained from the ISA shape functions, and thus take molecular environment effects into account. Even when only considering attractive dimer configurations (solid bar in Fig. 3.3), errors in the dispersion energy component are reduced by 23%, demonstrating the importance of the damping function across the potential surface. From these results, and in agreement with related literature studies,¹⁵⁹ we conclude that use of the standard Tang-Toennies damping function based on atomic ionization potentials^{117,121,160–164} lacks quantitative predictive power compared to the Slater-ISA model. Note that neither the Slater-ISA FF nor the Born-Mayer-IP FF are directly fitted to the DFT-SAPT dispersion energies

(all parameters are determined from monomer properties), making this accuracy particularly striking. We hypothesize that the effect of the Slater-ISA approach is greater for dispersion than for first-order exchange because here (in contrast to the exchange energy) there are no fitted parameters to compensate for deficiencies in the exponents or functional form of the Born-Mayer-IP FF.

In contrast to the exchange and dispersion energies, the Slater-ISA FF and the Born-Mayer-IP FF show nearly identical errors for the electrostatic and the induction (2^{nd} order induction plus δHF) energies. In these cases, the two models differ only in the parameters and functional form used to represent the exponentially-dependent short-range terms of these energy components, namely the penetration component for the electrostatic term and the penetration/charge-transfer term for the induction. The lack of improvement between the Slater-ISA and Born-Mayer-IP models may imply that we are not able to capture the physics of these particular short-range interactions with either the Slater-functional or Born-Mayer functional forms. Alternatively, the assumption that the short-range components of the electrostatic and induction energies are proportional to the exchange-repulsion may need to be re-examined. As discussed in Section 3.2.2, this proportionality is known to be approximately valid, but as yet there does not seem to be a deeper theoretical understanding of these short-range terms that may lead to a better model. Nevertheless, absolute errors in the electrostatic and induction components are relatively small for both models. Thus overall, the Slater-ISA FF functional form is promising for treating a wide variety of short-range effects.

The comparison between the Slater-ISA FF and the LJ FF is slightly more complicated, owing to the differences in long-range potential and fitting methodology (see Section 3.3.4). As such, we compare the Slater-ISA FF to several versions of the LJ FF (for which characteristic RMSE and $\|\text{MSE}\|$ are shown in Table 3.2). Using the same weighting function and constraining the Coulombic and polarization terms to be identical to the Slater-ISA FF, we see that the resulting Lennard-Jones force field (LJ FF, $\lambda = 2.0$) is significantly worse than the Slater-ISA FF, both in terms of total RMSE and attractive RMSE. Furthermore, by comparing the $\|\text{MSE}\|$ of both force fields, we see that errors in LJ FF are much more *systematic* than in the

| | LJ FF Dimer-Specific Fits | | LJ FF Transferable Fits | |
|-----------|--|--|--|--|
| | $\lambda = 2.0$ (kJ mol ⁻¹) | $\lambda = 0.1$ (kJ mol ⁻¹) | $\lambda = 2.0$ (kJ mol ⁻¹) | $\lambda = 0.1$ (kJ mol ⁻¹) |
| RMSE | 1.984 (0.603) | 6.058 (0.413) | 2.054 (0.640) | 5.760 (0.457) |
| $\ MSE\ $ | 0.322 (0.345) | 1.610 (0.041) | 0.311 (0.368) | 1.410 (0.060) |

Table 3.2: Comparison of characteristic RMSE and $\|MSE\|$ over the 91 dimer test set for the various Lennard-Jones models. The LJ models are not parameterized on a component-by-component basis, thus RMSE/ $\|MSE\|$ values are only shown for the total FF energies. ‘Attractive’ errors, representing the characteristic RMSE/ $\|MSE\|$ for the subset of points whose energies are net attractive ($E_{int} < 0$), are shown in parentheses to the right of the total errors. ‘Dimer-Specific Fits’ and ‘Transferable Fits’ are as in Table 3.1.

Slater-ISA FF: in order to reproduce the repulsive wall correctly, the Lennard-Jones potential generally underestimates the well-depth by a considerable fraction (see the Supporting Information of Ref. 104 for ethane as a typical example).

Given the failure of the LJ FF ($\lambda = 2.0$) force field to reproduce the energetically important region of the PES, we also compared the Slater-ISA FF to a ‘best-case’ scenario Lennard-Jones force field which correctly reproduces the minimum energy region at the expense of the repulsive wall. These LJ FF ($\lambda = 0.1$) fits have total RMSE errors nearly 4 times that of the Slater-ISA FF; indeed, the LJ FF ($\lambda = 0.1$) reproduces the repulsive wall only qualitatively. Insofar as the repulsive wall is concerned, the Slater-ISA FF is far superior to the Lennard-Jones short-range model. Nevertheless (and much more importantly for molecular simulation), the attractive region of the potential is reproduced surprisingly well by LJ FF. Characteristic attractive RMSE for the LJ FF ($\lambda = 0.1$) are slightly lower than those for Slater-ISA FF, although the former has one additional free parameter per atom type and is also fit directly to reproduce the total energy. Likewise, attractive $\|MSE\|$ between the Slater-ISA FF and the LJ FF ($\lambda = 0.1$) are comparable. As we show in the Supporting Information of Ref. 104, however, and as is well known in the literature, weighting the Lennard-Jones potential in this manner does not necessarily capture important

information from the long-range attractive tail or repulsive wall of the PES, such that the LJ FF ($\lambda = 0.1$) is not always expected to yield good property predictions. This latter point will be demonstrated in Section 3.4.2.

In order to compare the performance of the Slater-ISA FF against popular standard force fields, we also developed a ‘best case scenario’ non-polarizable point charge Lennard-Jones model, results for which are shown in the Supporting Information of Ref. 104. Unsurprisingly, this force field is worse (in an RMSE and $\|\text{MSE}\|$ sense) than all other force fields studied in this Chapter, thus demonstrating how important accurate models for long-range electrostatics and polarization are to the overall accuracy of ab initio force fields.

Argon Dimer

We now turn to several specific case studies. The Ar dimer provides an interesting test case to examine directly the impact of the polynomial pre-factor included in the Slater-ISA FF functional form. Since Ar is an atomic species, we should have $B_{\text{Ar}}^{\text{ISA}} = B_{\text{Ar}}^{\text{IP}}$. For numerical reasons, the Slater-ISA FF and Born-Mayer-IP FF exponents differ by 0.03 a.u.; however, this difference is insignificant, and the two FFs differ mainly in the polynomial pre-factor. Fig. 3.4 shows the potential energy surface (PES) for the argon dimer computed using the Slater-ISA FF and the Born-Mayer-IP FF. Here the default weighting scheme has been used so as to best reproduce the energetically attractive region. Note that, while both potentials reproduce the minimum energy configurations correctly, the Born-Mayer-IP FF considerably overestimates the exchange energy (and thus the total energy) along the repulsive wall. The Slater-ISA FF, on the other hand, maintains excellent accuracy in this region of the potential. This result is particularly notable because the repulsive wall is not heavily weighted in the fit. (A point 10 kJ mol⁻¹ along the repulsive wall, for instance, is weighted only 3% as heavily as a point near the bottom of the well). A similar, though smaller, increase in accuracy is seen in the fit to the DFT-SAPT dispersion energies, where the Slater-ISA FF is better able to model the energies for shorter interatomic separations. This increased accuracy is entirely attributable to

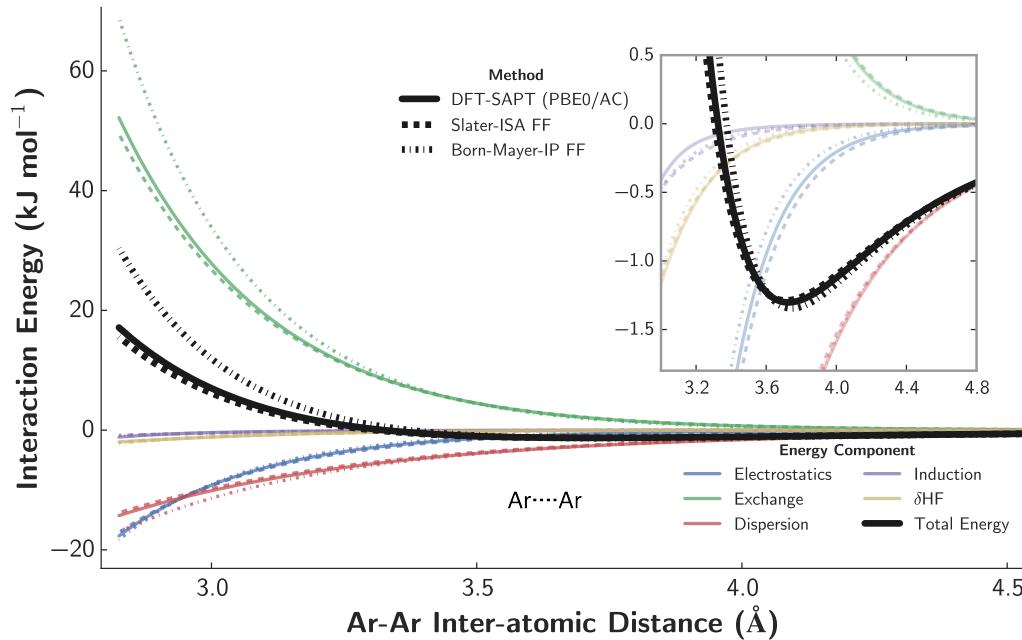


Figure 3.4: Potential energy surface for the argon dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference.

the functional form employed, as the dispersion parameters are identical between the two FFs.

Consistent with prior literature,^{87,105} these results suggest that neglect of the polynomial pre-factor P (as in standard Born-Mayer potentials) is *by itself* a poor approximation. However, as we show below, the Born-Mayer form can still be used as an accurate model in conjunction with appropriately scaled atomic exponents. Nonetheless, the more physically-motivated Slater form provides increased accuracy over a wider range of separations without recourse to empirical scaling.

Results for LJ FF are shown in the Supporting Information of Ref. 104; consistent with expectations for the Lennard-Jones model, the repulsive wall is overestimated by the $1/r_{ij}^{12}$ short-range functional form, and the magnitude of the attractive tail region is similarly overestimated by the effective $C_{ij,6}$ dispersion parameter. Note

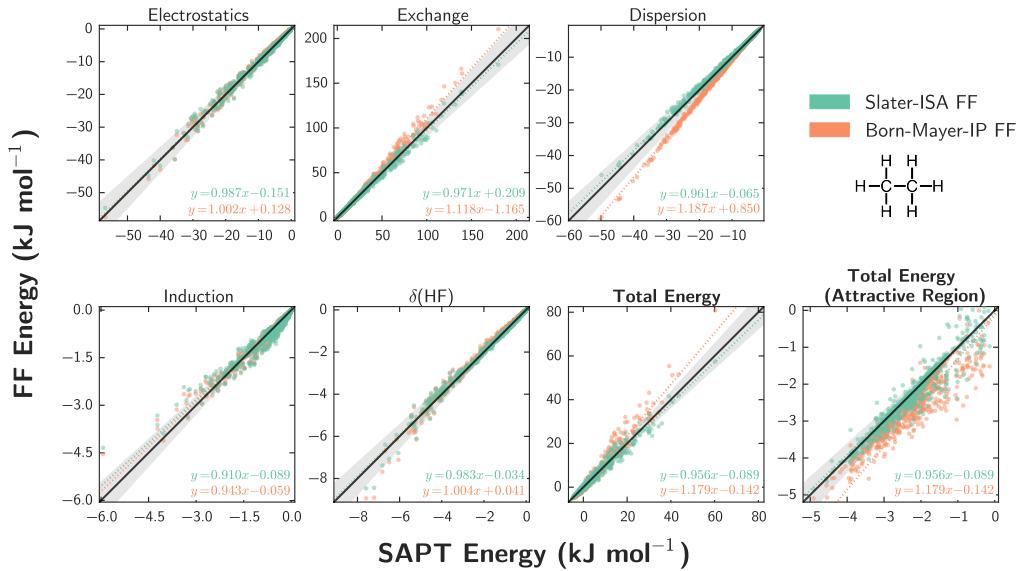


Figure 3.5: Force field fits for the ethane dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs. Fits for each energy component are displayed along with two views of the total interaction energy. The diagonal line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within $\pm 10\%$ agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each energy component.

that this $C_{ij,6}$ coefficient has been fit to the total energy, and thus differs from the asymptotically-correct $C_{ij,6}$ parameter used for both the Slater-ISA FF and the Born-Mayer-IP FF. An alternative parameterization strategy would have been to use the asymptotically-correct $C_{ij,6}$ parameter in the LJ FF, but this would have worsened predictions along both the repulsive wall and the minimum energy configurations.

Ethane Dimer

We next discuss the ethane dimer and show both a scatter plot of the 1000 dimer interactions (Fig. 3.5) and a cut through the potential energy surface near the minimum (Fig. 3.6) as indications of force field quality.

As with the argon dimer, for the ethane dimer the Slater-ISA FF produces more

accurate exchange and dispersion energies compared to the Born-Mayer-IP FF. Here, the effects of the Slater-ISA FF for dispersion are even more pronounced, likely because the conventional damping of the Born-Mayer-IP FF is systematically in error due to differences in both the form of the damping function and exponents. As for the total interaction energy, we again find that the Born-Mayer-IP FF exhibits large errors for repulsive contributions, while the Slater-ISA FF naturally reproduces interactions for both attractive and strongly repulsive configurations. Even in the attractive regime, the Born-Mayer-IP FF is systematically too attractive. These systematic errors are the result of imperfect error cancellation between the exchange and dispersion components of the fit, and are discussed in more detail in Section 3.4.4.

Examining a specific cut across the ethane-ethane PES (Fig. 3.6) visually confirms these results. Both potentials do an excellent job of reproducing the benchmark DFT-SAPT energies in the minimum energy region, though the Born-Mayer-IP FF is slightly too attractive. (Other cuts of the PES would show the Born-Mayer-IP predictions to be significantly more in error, consistent with the scatter plots). Along the repulsive wall, however, the Born-Mayer-IP FF predictions worsen in comparison to those from the Slater-ISA FF. Finally, the PES shows an increased reliance on error cancellation between the various energy components for the Born-Mayer-IP FF compared to the Slater-ISA FF.

As shown in the Supporting Information of Ref. 104, the Lennard-Jones force field models are incapable of reproducing the entirety of the ethane PES; depending on the weighting function, either the repulsive wall or the attractive well can be reproduced, however no set of parameters can predict both regions simultaneously.

Acetone Dimer

The acetone dimer provides a final interesting example involving a moderately sized organic molecule. From both the scatter plots (Fig. 3.7) and the PES cross section (Fig. 3.8), it is evident that both the Slater-ISA and Born-Mayer-IP force fields do an excellent job of reproducing DFT-SAPT energies for the low energy

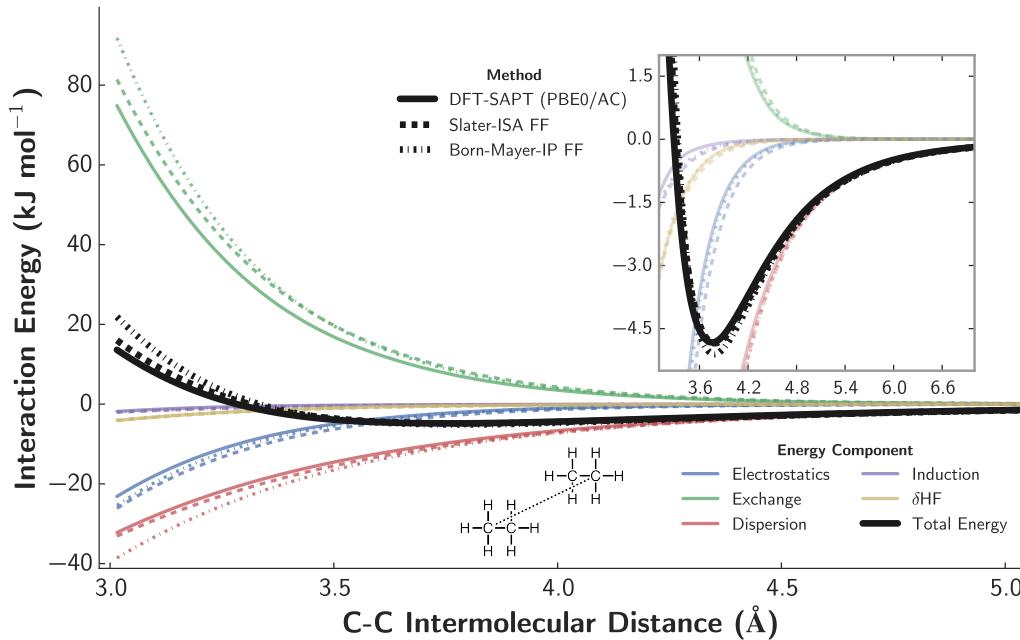


Figure 3.6: A representative potential energy scan near a local minimum for the ethane dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The ethane dimer configuration in this scan corresponds to the most energetically attractive dimer included in the training set; other points along this scan are not included in the training set.

dimers. Along the repulsive wall, however, the Born-Mayer-IP FF shows larger systematic errors in each energy component, and seems to rely on error cancellation to achieve good agreement in the total energy. This reliance on error cancellation has two negative effects: Firstly, the additional scatter in the total energy of the Born-Mayer-IP FF fit, especially prominent for attractive configurations, indicates that this error cancellation is imperfect in certain cases. MSE for the Slater-ISA FF ($-0.0115 \text{ kJ mol}^{-1}$) are an order of magnitude lower than for the Born-Mayer-IP FF ($0.182 \text{ kJ mol}^{-1}$) in the attractive region of the potential. Secondly, as we shall later explore, reliance on error cancellation likely contributes to the somewhat decreased transferability of the Born-Mayer-IP FF as compared to the Slater-ISA FF.

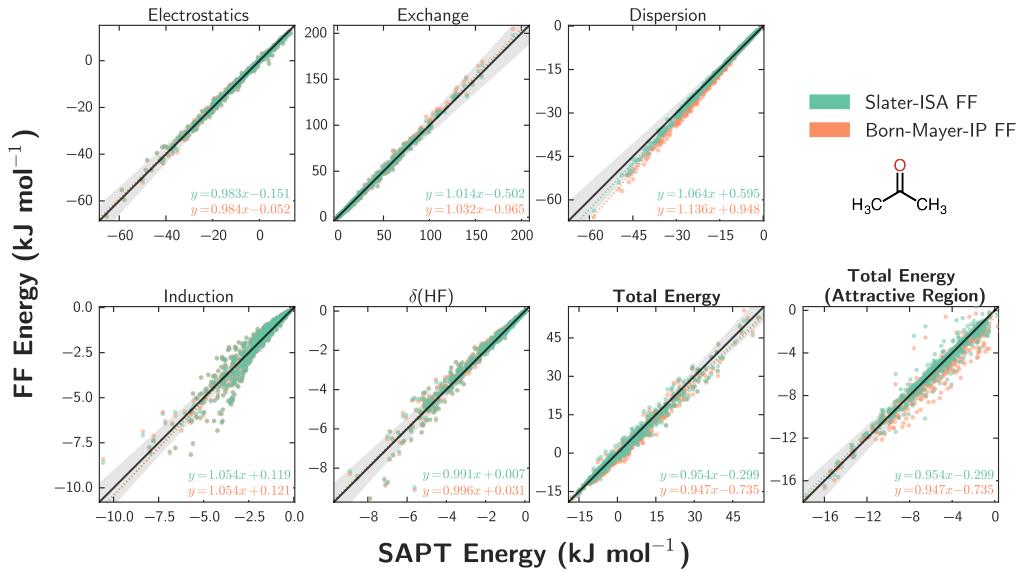


Figure 3.7: Force field fits for the acetone dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs, as in Fig. 3.5.

As shown in the Supporting Information of Ref. 104, the LJ FF predictions for acetone are reasonably good in both the tail and minimum energy regions of the potential, however the LJ FF grossly overpredicts the DFT-SAPT (PBE0/AC) energies along the repulsive wall.

3.4.2 Accuracy: Comparison with experiment

We have benchmarked the above force fields against experimental second virial coefficients and, in the case of ethane, enthalpies of vaporization and liquid densities. The classical 2nd virial coefficients were calculated for both argon and ethane using rigid monomer geometries, following the procedure described in Ref. 117. Enthalpies of vaporization and liquid densities were calculated using the OpenMM molecular simulation package¹⁵⁷ as described in Section 3.3. Higher-order multipole moments — which were negligible for these molecules — were neglected, and so only rank 0 terms were used in these calculations. Results are shown in Figures

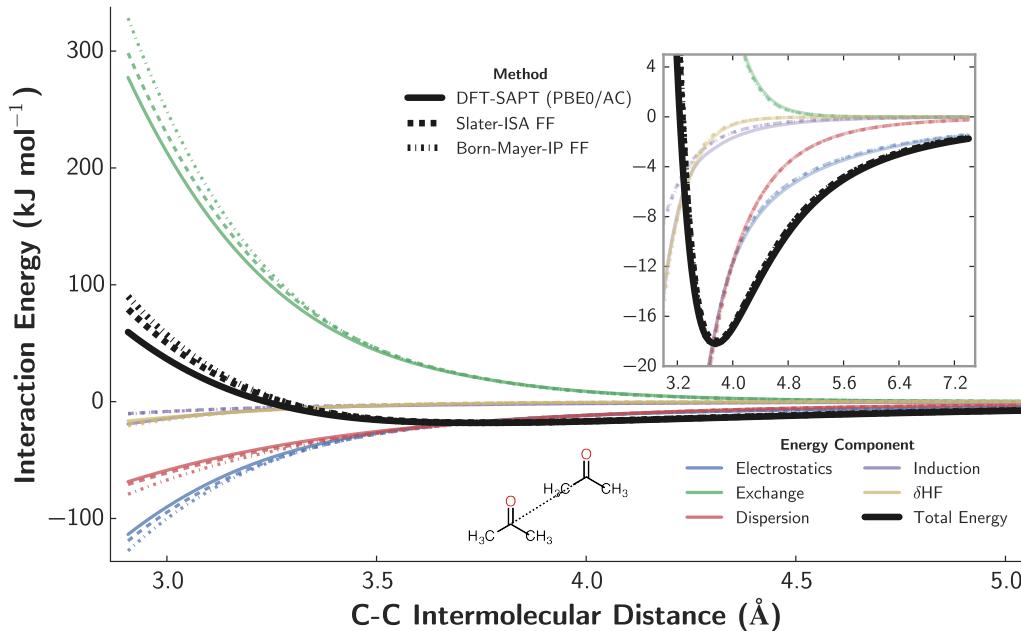


Figure 3.8: A representative potential energy scan near a local minimum for the acetone dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The intermolecular distance is taken to be the internuclear distance between the two carbonyl carbons on each acetone monomer. The configuration in this scan corresponds to the most attractive dimer configuration included in the training set for the acetone dimer; other points along this scan have not explicitly been included in the training set.

3.9 and 3.10 as well as Table 3.4.

For argon, since both Slater-ISA FF and Born-Mayer-IP FF accurately reproduce the energetics of low-energy configurations, it is unsurprising that both force fields yield accurate virial coefficients over a wide range of temperatures. Errors in computed B_2 coefficients (for both potentials) are likely attributable to small errors in the DFT-SAPT (PBE0/AC) potential itself,¹⁴⁶ and, to a much lesser extent, the neglect of nuclear quantum effects at lower temperatures.¹⁶⁵ Despite the good (in an RMSE sense) fit quality of the LJ FF ($\lambda = 0.1$), this force field overpredicts the magnitude of

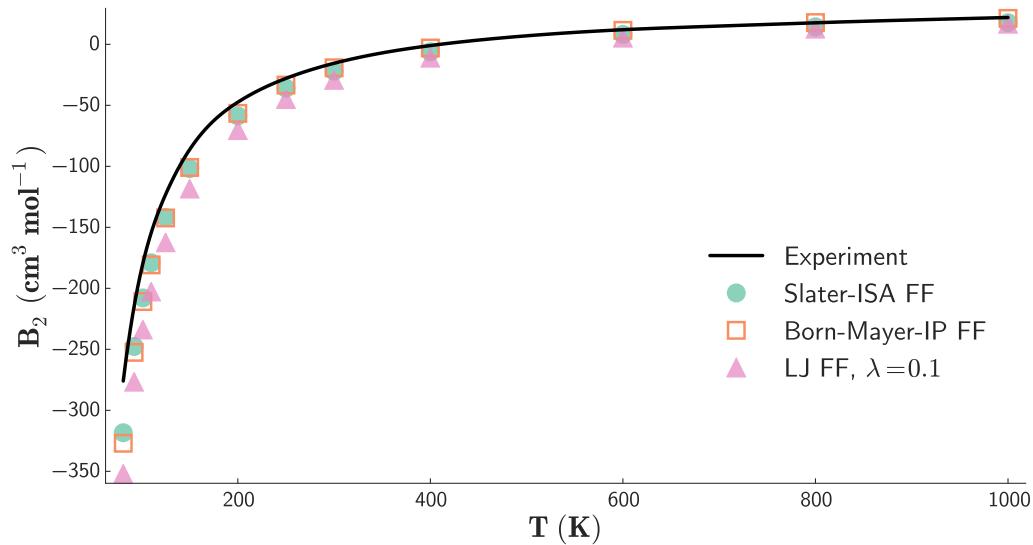


Figure 3.9: Second virial coefficients for argon. The Slater-ISA and the Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from Ref. 6.

the 2nd virial for argon, likely as a result of the effective dispersion coefficient, which overestimates the attraction in the tail region of the PES (see Supporting Information of Ref. 104). Although it is certainly possible to parameterize a Lennard-Jones model *empirically* for argon, such a force field would rely on a subtle cancellation of errors between the minimum energy- and tail-regions of the PES. As the proper balance is impossible to predict *a priori*, this result highlights one of the difficulties of using the less physical LJ model in the development of ab-initio force fields.

In the case of ethane, the Slater-ISA FF is in excellent agreement with experiment, whereas the Born-Mayer-IP FF underpredicts B_2 by as much as 20%. These results are indicative, not only of the more accurate functional form and parameterization of Slater-ISA FF, but also of the high accuracy of the underlying DFT-SAPT (PBE0/AC) benchmark energies. In this case, LJ FF also correctly predicts the virial. Using weighting functions for each model that are optimal for the 91 dimer test set as a whole ($\lambda = 2.0$ for the Slater-ISA FF and the Born-Mayer-IP FF, $\lambda = 0.1$ for the LJ FF), all force fields produce similar results for ΔH_{vap} and ρ (Table 3.4). These

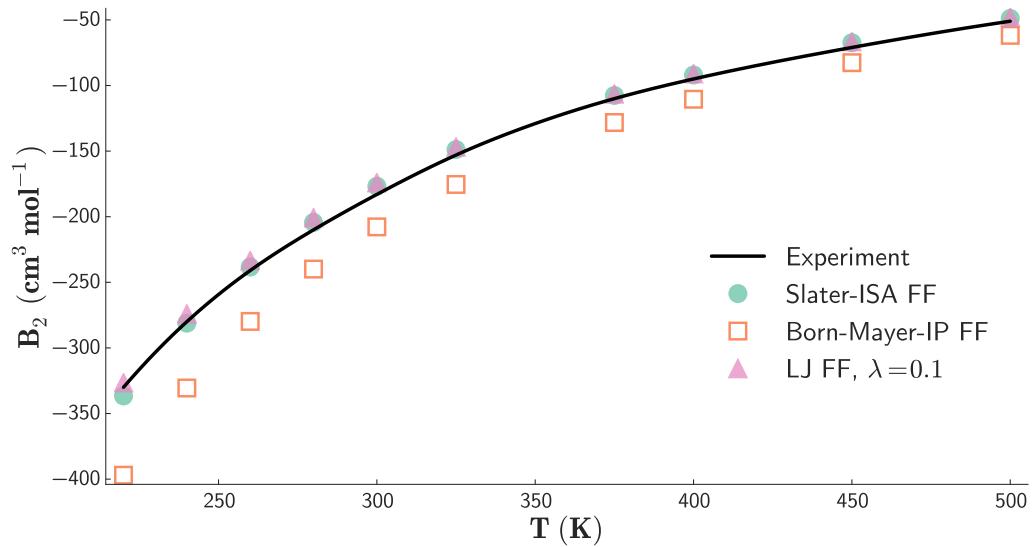


Figure 3.10: Second virial coefficients for ethane. The Slater-ISA and Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from Ref. 6.

values are slightly overestimated by all force fields (especially in the case of the Born-Mayer-IP FF), which is to be expected given our neglect of many-body effects. McDaniel and Schmidt have calculated the 3-body correction for the Born-Mayer-IP FF; using this value as a global 3-body correction for all force fields, we see that both the Slater-ISA and the Lennard-Jones force fields compare very favorably to experiment, with the Slater-ISA FF perhaps slightly more accurate.

3.4.3 Transferability

The transferability of interaction potentials is a crucial aspect of practical molecular simulations. Here we examine ‘parameter transferability’, by which we mean the extent to which parameters from two homo-monomeric systems can be combined to predict the intermolecular interactions of the resulting mixed hetero-monomeric system. As a measure of parameter transferability, we compared characteristic RMSE and $\|\text{MSE}\|$ relative to the benchmark data for two different parameterization

schemes. For the ‘Dimer-Specific Fits’, A_{ij} parameters were obtained for each of the 91 dimer pairs individually; these results are identical to those discussed in the previous two subsections. In contrast, for the ‘Transferable Fits’, the A_{ij} parameters were fit to the 13 homonomeric dimer pairs and were re-used (without any further optimization) to calculate energies for the 78 mixed systems using the combination rules listed in Section 3.3.4. Results for each parameterization scheme are shown in Table 3.1. From the RMSE and $\|\text{MSE}\|$ from the competing schemes, we see excellent parameter transferability for all force fields studied. For the Slater-ISA FF, characteristic RMSE and $\|\text{MSE}\|$ for each component increase by a very small fraction upon constraining the fit; due to small error cancellation, errors in the total energy actually *decrease* somewhat with these constraints. (This is possible since the total energy is not directly fit.) The Born-Mayer-IP FF also displays a significant degree of transferability, though errors in the total energy increase slightly upon constraining the fit. As in prior work, the observed parameter transferability for both force fields can be attributed to our use of a term-by-term parameterization scheme (Section 3.3.4), which serves to minimize error cancellation between energy components and generate a more physically-meaningful (and thus transferable) set of parameters.^{117,166} Finally, note that for four of the five interaction energy components the relative change in RMSE on constraining the fit is smaller for the Slater-ISA FF than the Born-Mayer-IP FF. The δHF term is the exception, but even here the relative change in errors from the two methods are comparable. This suggests that the Slater-ISA FF may be the more transferable of the force fields studied. Nevertheless, the Lennard-Jones model is surprisingly transferable, likely in part due to the same accurate and transferable ‘long-range’ electrostatics and polarization as the Slater-ISA FF. The non-polarizable, point-charge Lennard-Jones model (results for which are shown in the Supporting Information of Ref. 104) displays the least transferability (in both an RMSE and $\|\text{MSE}\|$ sense) of all force fields studied.

Although we do not examine it here, we expect that the previously demonstrated success^{2,105,117,166} of the Born-Mayer-IP FF with respect to ‘environment transferability’ — the extent to which a single set of parameters can model a variety of

phases and molecular environments — and ‘atom type transferability’ — the extent to which atoms in chemically similar environments can accurately be grouped together into ‘types’ and treated using one parameter set — would also apply to, or even be improved by, Slater-ISA FF. These issues are under investigation in our groups.

3.4.4 Robustness

One of the practical challenges of ab initio force field development is the robustness of the resulting force field quality with respect to the choice of an appropriate training set and/or weighting function. To this end, the default weighting function (Eq. (3.37), $\lambda = 2.0$) was varied to produce unconstrained fits that were skewed either towards attractive ($\lambda = 0.5$) or repulsive ($\lambda = 5.0$) configurations, and pairwise differences in force field total energies were computed between each weighting scheme. Characteristic root-mean-square pairwise differences (RMSD) between each weighting function are shown in Table 3.3; as before, ‘attractive RMSD’ were calculated by excluding repulsive points from consideration. Note that, on average, the default $\lambda = 2.0$ weighting scheme is optimal (in an RMSE sense) for both the Slater-ISA and Born-Mayer-IP FFs.

Overall, both the Born-Mayer-IP FF and the LJ FF display significant weighting function sensitivity. This sensitivity is not surprising; as both force fields are unable to reproduce the entirety of the potential energy surface, changing the weighting scheme (or equivalently, the balance of configurations in the training set) alters the parameters in the Born-Mayer-IP FF or the LJ FF models quite substantially. Even excluding repulsive configurations, RMSD of $\sim 0.5 \text{ kJ mol}^{-1}$ are typical for the Born-Mayer-IP FF. RMSD are somewhat smaller for the LJ FF ($\sim 0.3 \text{ kJ mol}^{-1}$), however qualitatively we see that differences in computed force field energies are systematic: smaller weighting functions capture the minimum energy region of the potential while overestimating the magnitudes of both the repulsive and tail regions of the potential, whereas larger weighting functions tend to underestimate the minimum energy region in order to correctly reproduce the repulsive wall.

| Characteristic RMSD | $\lambda = 0.5$ vs 2.0 (kJ mol ⁻¹) | $\lambda = 0.5$ vs 5.0 (kJ mol ⁻¹) | $\lambda = 2.0$ vs 5.0 (kJ mol ⁻¹) |
|---------------------|---|---|---|
| Slater-ISA FF | 0.742 (0.207) | 0.990 (0.273) | 0.306 (0.086) |
| Born-Mayer-IP FF | 1.866 (0.409) | 2.632 (0.550) | 0.797 (0.153) |
| LJ FF | 1.301 (0.216) | 1.605 (0.309) | 0.324 (0.099) |
| Born-Mayer-sISA FF | 0.611 (0.178) | 0.810 (0.236) | 0.293 (0.081) |

Table 3.3: Characteristic RMS pairwise differences (RMSD) in force field total energies for different weighting functions with λ values as defined in Eq. (3.37); values shown are the (arithmetic mean, rather than geometric) RMSD across the 91 dimer test set. Characteristic ‘Attractive’ RMSD (as defined in Table 3.1) are shown in parentheses to the right of each overall RMSD.

Consequently, the Lennard-Jones model shows weighting-function sensitivity in a manner that is not entirely captured by the RMSD, but is instead reflected in the greater sensitivity of the LJ FF (as compared to the Born-Mayer-IP FF) in the prediction of experimental properties (*vide infra*).

Note that for practical force field development (as opposed to minimization of overall RMSE), the default weighting scheme for the Born-Mayer-IP FF and the LJ FF is suboptimal for many dimers in the test set. Because both the Born-Mayer-IP FF and the LJ FF must inherently compromise between accuracy near the minimum and along the repulsive wall, the weighting function requires system-specific fine-tuning in order to achieve proper balance. This empiricism creates significant challenges in the development of ab initio force fields.

By contrast, we find the Slater-ISA FF to be robust with respect to the choice of weighting function due to its more balanced treatment of repulsive and attractive regions of the potential energy surface. Average RMSD for the Slater-ISA FF are between two to three *times* smaller compared to the Born-Mayer-IP FF, and the Slater-ISA FF is relatively insensitive to the choice of weighting function. These conclusions hold for both attractive and overall RMSD. As a result, the Slater-ISA model largely eliminates the need for empirical fine-tuning of the weighting function, which in turn greatly simplifies the parameterization process and allows for a more robust prediction of chemical and physical properties.

| Force Field | Weighting Function | | | | Experiment |
|--|--------------------|-----------------|-----------------|-----------------|------------|
| | $\lambda = 0.1$ | $\lambda = 0.5$ | $\lambda = 2.0$ | $\lambda = 5.0$ | |
| ΔH_{vap} (kJ mol ⁻¹); $\rho = 0.546$ g L ⁻¹ , T = 184 K | | | | | |
| Slater-ISA FF | 15.3 (14.7) | 15.3 (14.6) | 15.3 (14.7) | 15.2 (14.6) | |
| Born-Mayer-IP FF | 14.3 (13.7) | 15.1 (14.5) | 16.6 (15.9) | 18.6 (18.0) | 14.7 |
| LJ FF | 15.5 (14.9) | 14.6 (13.9) | 11.4 (10.7) | 10.1 (9.5) | |
| ρ (g L ⁻¹); P = 1 atm, T = 184 K | | | | | |
| Slater-ISA FF | 0.600 (0.566) | 0.602 (0.568) | 0.600 (0.566) | 0.593 (0.559) | |
| Born-Mayer-IP FF | 0.521 (0.487) | 0.567 (0.533) | 0.632 (0.598) | 0.678 (0.644) | 0.546 |
| LJ FF | 0.607 (0.573) | 0.610 (0.576) | 0.555 (0.521) | 0.494 (0.460) | |

Table 3.4: Enthalpies of vaporization and liquid densities for ethane as a function of force field and weighting function. Values in parentheses include an estimation of the 3-body correction (0.628 kJ mol⁻¹ and 0.034 g mL⁻¹ for the enthalpy of vaporization and liquid density, respectively) as computed in Ref. 2. Experimental data taken from Ref. 3 and Ref. 4.

For the ethane dimer, Fig. 3.11 shows overall force field energies for both the Slater-ISA and Born-Mayer-IP FFs for three weighting functions. Results for the Lennard-Jones models are shown in the SI, and are qualitatively similar to the Born-Mayer-IP FF results. The Born-Mayer-IP FF fits vary qualitatively with λ , leading to a relatively large uncertainty in calculated B_2 coefficients, enthalpies of vaporization, and liquid densities (see Table 3.4). By skewing the fits towards attractive configurations ($\lambda = 0.5$), the majority of attractive configurations are predicted without systematic error, though points along the repulsive wall (including those with net negative energies) are systematically too repulsive. Using a scheme which more heavily weights repulsive configurations, the Born-Mayer-IP FF regains semi-quantitative accuracy for repulsive configurations, albeit at the expense of a systematic increase in errors for the attractive dimer configurations. Finally, we reiterate that the optimal weighting function for the ethane dimer (here $\lambda = 0.5$ best reproduces the 2nd virial for the Born-Mayer-IP FF) is by no means universal for the molecules in the 91 dimer test set.

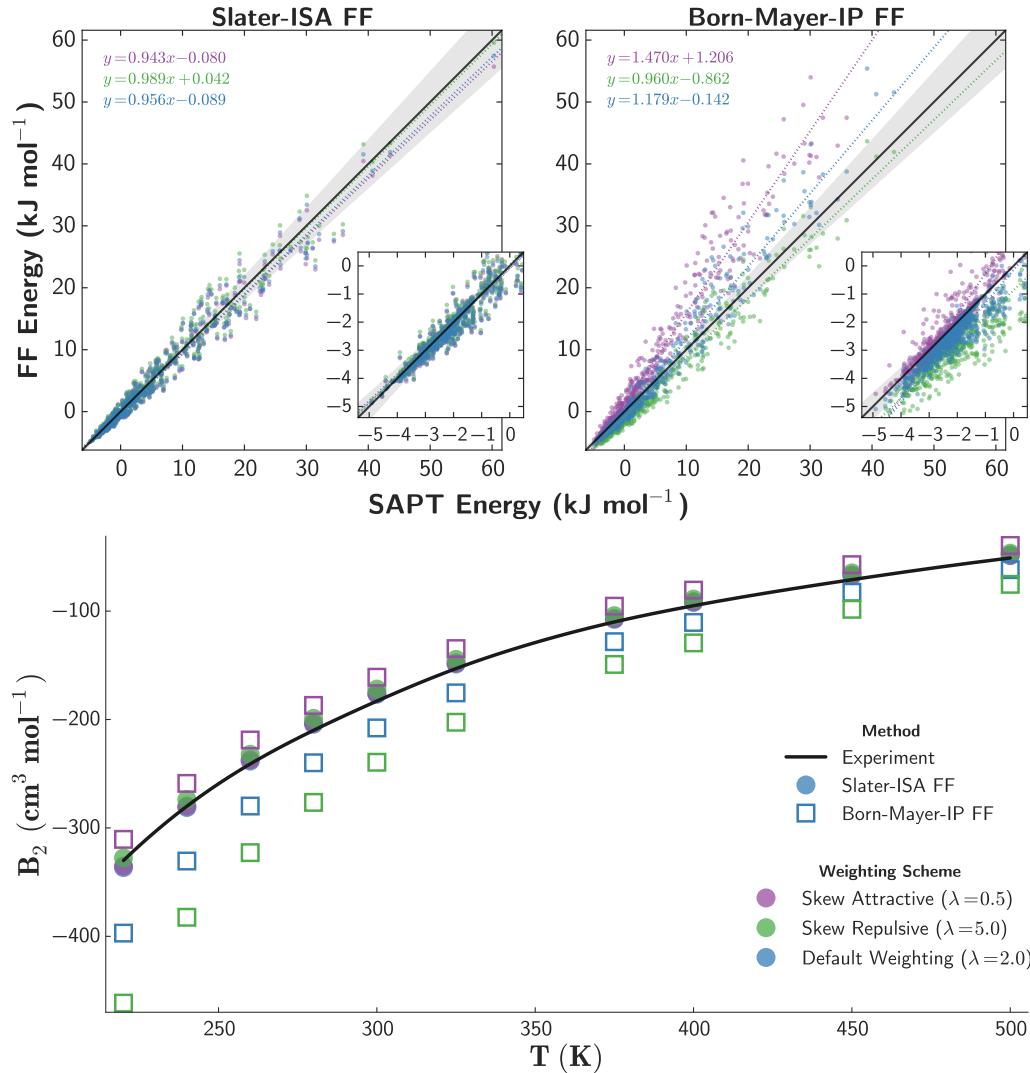


Figure 3.11: Comparison of the Slater-ISA FF and the Born-Mayer-IP FF in terms of sensitivity to the weighting function employed in parameter optimization for the ethane dimer. Three weighting functions, $\lambda = 0.5$ (purple), $\lambda = 2.0$ (blue), and $\lambda = 5.0$ (green) are shown, with higher λ values indicating more weighting of repulsive configurations.

(top) Total interaction energies for the Slater-ISA FF (left) and the Born-Mayer-IP FF (right) indicating the accuracy of each force field with respect to DFT-SAPT (PBE0/AC) benchmark energies. The diagonal line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within $\pm 10\%$ agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each weighting function.

(bottom) Computed 2nd virial coefficients for ethane. Data for the Slater-ISA FF and the Born-Mayer-IP FF are depicted using shaded circles and open squares, respectively; colors for the different weighting functions are as above. Experimental data from Ref. 6 (black line) is also shown.

The Slater-ISA FF fits for the ethane dimer, on the other hand, are nearly completely insensitive to the weighting function, leading to little intrinsic uncertainty in the determination of parameters or in the computation of macroscopic properties. Some other dimers, particularly those where atomic anisotropy would be anticipated (e.g., water), exhibited slightly larger sensitivity to the weighting function. Nevertheless, the vast majority of dimers in the test set are qualitatively insensitive to the choice of weighting function, and can be optimized with the default $\lambda = 2.0$ weighting function without yielding undue systematic error in the attractive region of the potential, thus proving the enhanced robustness of the Slater-ISA FF model relative to conventional force fields.

3.4.5 Next-Generation Born-Mayer Models: Born-Mayer-sISA FF

We hypothesize that the increased accuracy, transferability, and robustness of the Slater-ISA FF is a direct result of its more physically-motivated functional form and its use of ISA-derived atomic exponents that directly account for the influence of the molecular environment. Nonetheless, we recognize that the standard Born-Mayer functional form remains extremely common, both in simulation software and in existing force fields. It is therefore fruitful to explore the extent to which the BS-ISA exponents themselves could be used in conjunction with a Born-Mayer functional form. These results are shown in Table 3.5.

| Component | Dimer-Specific Fits | | | Transferable Fits | | |
|---------------------|--|---|--|--|---|--|
| | Slater-ISA FF (kJ mol ⁻¹) | Born-Mayer-ISA (kJ mol ⁻¹) | Born-Mayer-sISA (kJ mol ⁻¹) | Slater-ISA FF (kJ mol ⁻¹) | Born-Mayer-ISA (kJ mol ⁻¹) | Born-Mayer-sISA (kJ mol ⁻¹) |
| Exchange | 2.641 (0.686) | 7.030 (1.203) | 2.677 (0.686) | 2.718 (0.720) | 6.968 (1.228) | 2.764 (0.706) |
| Electrostatics | 1.087 (0.351) | 1.406 (0.589) | 1.083 (0.352) | 1.134 (0.351) | 1.461 (0.598) | 1.141 (0.352) |
| Induction | 0.251 (0.095) | 0.229 (0.097) | 0.250 (0.096) | 0.278 (0.101) | 0.257 (0.101) | 0.275 (0.101) |
| δ HF | 0.246 (0.068) | 0.327 (0.120) | 0.248 (0.068) | 0.274 (0.076) | 0.353 (0.122) | 0.274 (0.076) |
| Dispersion | 0.766 (0.317) | 3.584 (0.890) | 0.856 (0.336) | 0.766 (0.317) | 3.584 (0.890) | 0.856 (0.336) |
| Total Energy | | | | | | |
| RMSE | 1.701 (0.464) | 4.934 (1.054) | 1.751 (0.453) | 1.650 (0.456) | 4.555 (1.035) | 1.713 (0.446) |
| $\ MSE\ $ | 0.216 (0.057) | 1.127 (0.505) | 0.258 (0.063) | 0.175 (0.051) | 0.882 (0.516) | 0.245 (0.057) |

Table 3.5: Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Born-Mayer-sISA approximation compared with other methods. For the total energy, both RMSE and absolute mean signed errors (MSE) have been shown. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ($E_{int} < 0$), are shown in parentheses to the right of the total RMS errors; ‘attractive’ $\|MSE\|$ are likewise displayed for the total energy. Slater-ISA FF, Born-Mayer-ISA, and Born-Mayer-sISA FF are as described in the main text, and the ‘Dimer-Specific’ and ‘Transferable’ fits are as described in Table 3.1.

As expected, direct insertion of the BS-ISA exponents into the Born-Mayer functional form (Born-Mayer-ISA) does not yield promising results. Indeed, the Born-Mayer-ISA FF has significantly worse RMSE and $\|\text{MSE}\|$ than the Born-Mayer-IP FF. We reiterate that the $P = 1$ approximation from Eq. (3.25), yielding the conventional Born-Mayer form, is by itself a crude model. Rather, it becomes necessary to accompany this approximation by a corresponding exponent scale factor, ξ :

$$B_i = \xi B_i^{\text{ISA}}. \quad (3.39)$$

Following literature precedent,^{87,105} we hypothesized that ξ could be treated as a universal constant. To test this conjecture, we computed reference density overlaps for a variety of isolated atom pairs (details in the Supporting Information of Ref. 104), and fitted each of these overlaps to a Born-Mayer function of the form $S_{ij} \approx K_{ij} \exp(-\xi B_{ij}^{\text{ISA}} r_{ij})$, where $K_{ij} = \frac{K}{B_{ij}^3}$ in line with Eq. (3.13). To very good approximation, both K and ξ can be treated as universal constants; that is, neither K nor ξ is sensitive to the value of B^{ISA} . However, fitted values of K and ξ do depend strongly on the range of r_{ij} values used in the optimization, yielding estimates ranging from 0.74 to 0.88.

As an alternative, we optimized ξ directly by minimizing RMSE against the 91 dimer test set. Results from various choices of ξ can be found in the Supporting Information of Ref. 104. In agreement with prior literature and our ‘first-principles’ analysis of overlaps, we find $\xi = 0.84$ to be optimal for minimizing characteristic overall and attractive RMSE, though in practice the errors are insensitive to $\xi \in [0.82, 0.86]$. We henceforth use $\xi = 0.84$ and refer to this force field methodology (Born-Mayer functional form, ISA-derived exponents with scale factor $\xi = 0.84$) as the Born-Mayer-sISA FF. Parameters and homo-monomeric fits for the Born-Mayer-sISA FF can be found in the Supporting Information of Ref. 104 and in Section 3.B.

From Table 3.5 we see that the Born-Mayer-sISA FF is comparable in quality to our original Slater-ISA FF methodology. For all attractive configurations, the

Born-Mayer-sISA FF is equally accurate and transferable (Table 3.5). Furthermore, as shown in Table 3.3, Born-Mayer-sISA FF displays similar parameter robustness to Slater-ISA FF. These results suggest that many of the advantages of the Slater-ISA FF procedure can be captured simply by using the (scaled) ISA exponents. Note, however, that the optimal scale factor likely exhibits some system dependence, and furthermore that the enhanced Slater functional form may be important where an accurate description of highly repulsive configurations is crucial.

We also examined the Slater-ISA FF and the Born-Mayer-sISA FF against force fields where B_i values were instead treated as soft constraints, rather than fixed parameters. Using entirely unconstrained exponents yields unphysical parameters and a severe degradation in force field transferability. Using exponents from the Slater-ISA FF and the Born-Mayer-sISA FF as Bayesian priors (in the sense used in Refs. 85, 120), we generated two new force fields with optimized exponents, denoted Slater-OPT and Born-Mayer-OPT, respectively. Characteristic RMSE and $\|\text{MSE}\|$ for these force fields can be found in the Supporting Information of Ref. 104. We find that both methods yield only very minimal improvement, suggesting that the first-principles ISA exponents are already nearly optimal. Comparing the Born-Mayer-OPT exponents to those from Slater-ISA, we find a nearly identical average scale factor of $\gamma = 0.83 \pm 0.07$. Given that these optimal exponents can now be generated directly from first principles calculations of the molecular densities via the BS-ISA approach of Misquitta et al., we anticipate that the BS-ISA densities and resulting ISA exponents will be extremely useful in next-generation force field development in order to greatly simplify force field parameterization.

3.5 Conclusions and Recommendations

We have presented a new methodology for describing short-range intermolecular interactions based upon a simple model of atom-in-molecule electron density overlap. The resulting Slater-ISA FF is a simple extension of the conventional Born-Mayer functional form, supplemented with atomic exponents determined from an ISA analysis of the molecular electron density. In contrast to simple Born-Mayer or

Lennard-Jones models, the Slater-ISA FF is capable of reproducing ab initio interaction energies over a wide range of inter-atomic distances, and displays extremely low sensitivity to the details of parameterization. Furthermore, the Slater-ISA FF exhibits excellent parameter transferability. We thus recommend Slater-ISA FF for use in the development of future ab initio (and possibly empirically-parameterized) potentials, particularly where accuracy across wide regions of the potential surface is paramount.

More generally, we find that analysis of the ISA densities provides an excellent first-principles procedure for the determination of atomic-density decay exponents. This analysis improves upon existing approaches (which rely upon exponents derived from atomic radii or ionization potentials)^{102,167–169} and explicitly incorporates the influence of the molecular environment. These exponents can be used within Slater-ISA FF without further parameterization. Alternatively, in conjunction with an appropriate scale factor, the exponents can be used to enhance the accuracy of standard Born-Mayer potentials and/or Tang-Toennies damping functions. The resulting Born-Mayer-sISA FF retains many of the advantages of Slater-ISA FF, but also maintains compatibility with existing force fields and simulations packages that do not support the Slater functional form. Given that the BS-ISA exponents appear to be essentially optimal with respect to additional empirical optimization, we strongly recommend use of these first-principles exponents in order to simplify (both ab initio and empirical) future force field development involving Born-Mayer or related functional forms.⁶⁸

Overall, Slater-ISA FF enables a significantly increase in force field accuracy, particularly in describing short intermolecular contacts. Nevertheless, the neglect of atomic anisotropy remains, in some cases, a severe approximation.^{170–172} Indeed, it has been shown by many authors^{48,115,123,162} that quantitatively accurate A_{ij} parameters (and to a lesser extent, B_{ij} parameters) require incorporation of angular dependence for the generation of highly-accurate force fields. This anisotropy becomes crucial when describing systems containing lone pairs, hydrogen bonds, and/or π -interactions. Promisingly, BS-ISA densities naturally describe such anisotropy,^{85,120,173} and a straightforward method for its inclusion (where

essential) in ab initio force fields is the subject of Chapter 4.

3.A Waldman-Hagler Analysis of B_{ij} Combination Rule

The exact expressions for the overlap of two Slater densities $\rho_i = D_i \exp(-B_i r)$ and $\rho_j = D_j \exp(-B_j r)$ are shown here, first in the limiting case where the two exponents are equal ($B_i = B_j = B_{ij}$):

$$\begin{aligned} S_{B_i=B_j}^{ij} &= D_{ij} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\ D_{ij} &= \pi D_i D_j B_{ij}^{-3} \\ P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1, \end{aligned} \quad (3.40)$$

and second in the case where $B_i \neq B_j$:

$$\begin{aligned} S_{B_i \neq B_j}^{ij} &= \frac{16\pi D_i D_j \exp(-\{B_i + B_j\}r_{ij}/2)}{(B_i^2 - B_j^2)^3 r_{ij}} \times \\ &\left[\left(\frac{B_i - B_j}{2} \right)^2 \left(\exp \left(\{B_i - B_j\} \frac{r_{ij}}{2} \right) - \exp \left(-\{B_i - B_j\} \frac{r_{ij}}{2} \right) \right) \right. \\ &\times \left(\left(\frac{B_i + B_j}{2} \right)^2 r_{ij}^2 + (B_i + B_j) r_{ij} + 2 \right) \\ &- \left(\frac{B_i + B_j}{2} \right)^2 \exp \left(\{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left(\left(\frac{B_i - B_j}{2} \right)^2 r_{ij}^2 - (B_i - B_j) r_{ij} + 2 \right) \\ &\left. + \left(\frac{B_i + B_j}{2} \right)^2 \exp \left(-\{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left(\left(\frac{B_i - B_j}{2} \right)^2 r_{ij}^2 + (B_i - B_j) r_{ij} + 2 \right) \right]. \end{aligned} \quad (3.41)$$

Each overlap formula has been given a subscript to indicate limits on B_i and B_j .

Our goal is to ascertain the extent to which $S_{B_i \neq B_j}^{ij}$ can be accurately modeled by the functional form and variables of $S_{B_i = B_j}^{ij}$. D_i and D_j are pre-factors appearing in both equations, and we set these variables to unity without loss of generality. To find values of B_{ij} such that $S_{B_i \neq B_j}^{ij}(B_i, B_j, r_{ij}) \approx S_{B_i = B_j}^{ij}(B_{ij}, r_{ij})$, we first treat B_{ij} as a completely adjustable parameter, and later test for the existence of some simple combining function f such that $B_{ij} = f(B_i, B_j)$.

To optimize B_{ij} , we first require a training set of relevant $S_{B_i \neq B_j}^{ij}$ values. B_i , B_j , and r_{ij} are the only variables appearing in $S_{B_i \neq B_j}^{ij}$, and we could in principle fit B_{ij} values over a grid of B_i , B_j and r_{ij} combinations. However, we are only interested in the subset of points which are chemically relevant. Consequently, we developed a library of B_i values by deriving exponents from the ionization potentials of the first three rows of the periodic table (plus bromine and iodine). For each pair of elements, $B = 2\sqrt{2IP}$,¹³¹ and a range of r_{ij} values corresponding to 0.8-1.2 times the sum of the van der Waals radii of the two atoms was selected. B_{ij} values in $S_{B_i = B_j}^{ij}$ were then optimized (in a least-squares sense) for each element pair separately; Mean absolute percent errors (MAPE) for fitted overlaps are shown in 3.12 and in ??.

Relative errors for fitting are acceptably small for all element pairs. Excluding certain noble gases and alkali metals (He, Li, Ne, Na) from consideration, these being the elements with the most disparate B_i values compared to other elements, MAPE drops below 3% for all pairs, with the vast majority of MAPE below 1%. Our focus in this work is primarily on organic compounds where $|B_i - B_j|$ is small; empirically, these errors always translate to very small errors in the exchange energy itself. Use of an effective B_{ij} may require further testing in cases with extremely disparate B_i and B_j values.

We next tested whether the optimized B_{ij} could instead be modeled by a combination rule $B_{ij} = f(B_i, B_j)$. On the basis of symmetry and scaling considerations, Waldman and Hagler demonstrate that if a combination rule $f(B_i, B_j)$ exists, a plot of B_{ij}/B_i vs. B_j/B_i should lie on a single curve.¹⁰³ Remarkably (see 3.13), a geometric mean combination rule $B_{ij} = \sqrt{B_i B_j}$ models the fitted B_{ij} values with nearly quantitatively. This result allows the computation of Slater overlaps using the much

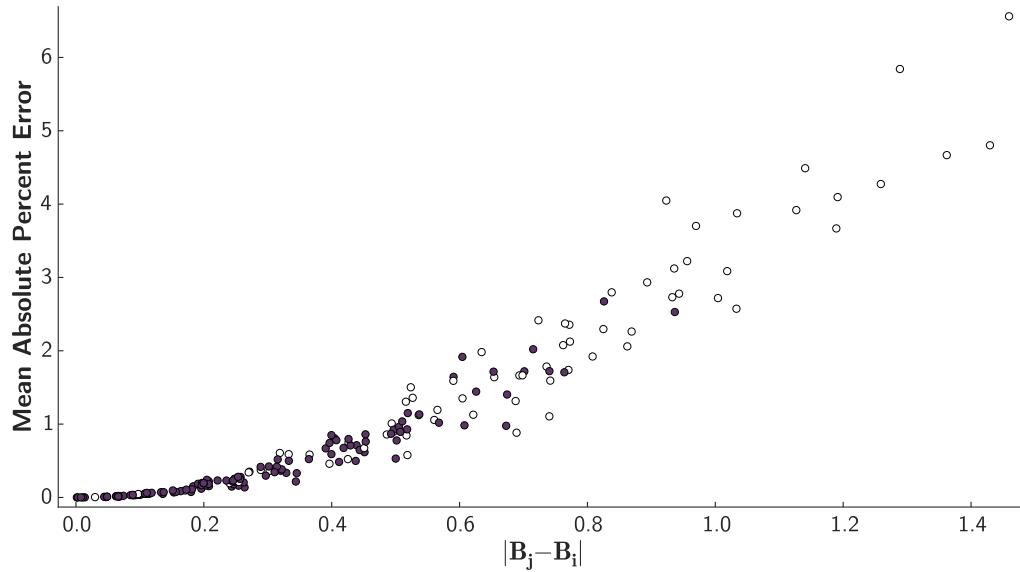


Figure 3.12: Mean absolute percent error of fitted overlap values as a function of the absolute difference between B_i and B_j values. Element pairs containing He, Li, Ne and/or Na are shown as empty circles. Deviations below 1% are seen for most element pairs, with noble gases and alkali metals posing a more significant challenge. Scatter in the plot is due to small variations in the absolute values of r_{ij} fit for each pair. As expected, $S_{B_i \neq B_j}^{ij}$ and $S_{B_i = B_j}^{ij}$ closely agree for $|B_i - B_j| \approx 0$.

simpler form of $S_{B_i = B_j}^{ij}$ (3.40) from individual atoms-in-molecule exponents B_i and B_j .

3.B Force Field Fits for Homomonomeric Systems

Scatter plots are shown for each homomonomeric system as an indication of force field quality with respect to DFT-SAPT (PBE0/AC) benchmark energies (4.7). As in the main text, fits for each energy component are displayed along with two views of the total interaction energy.

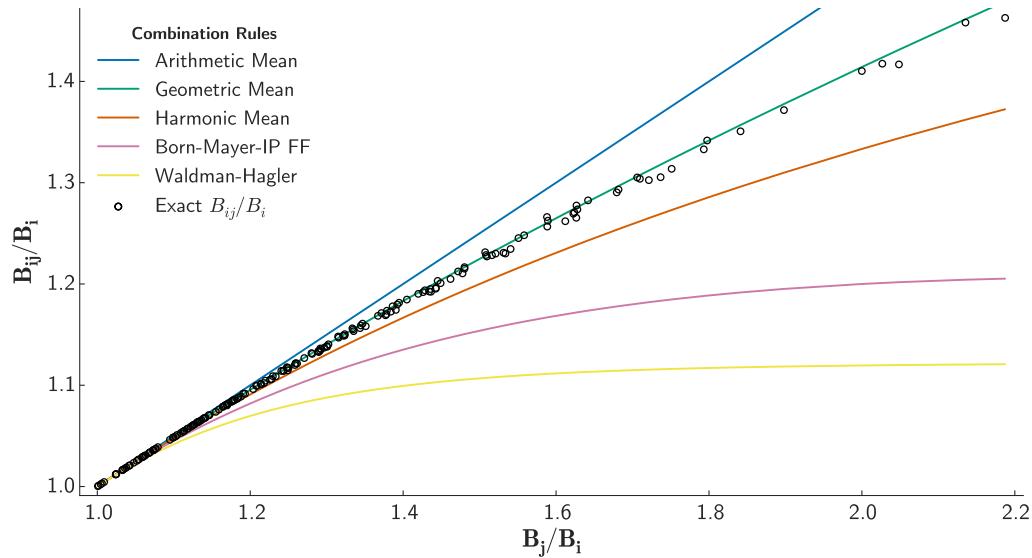
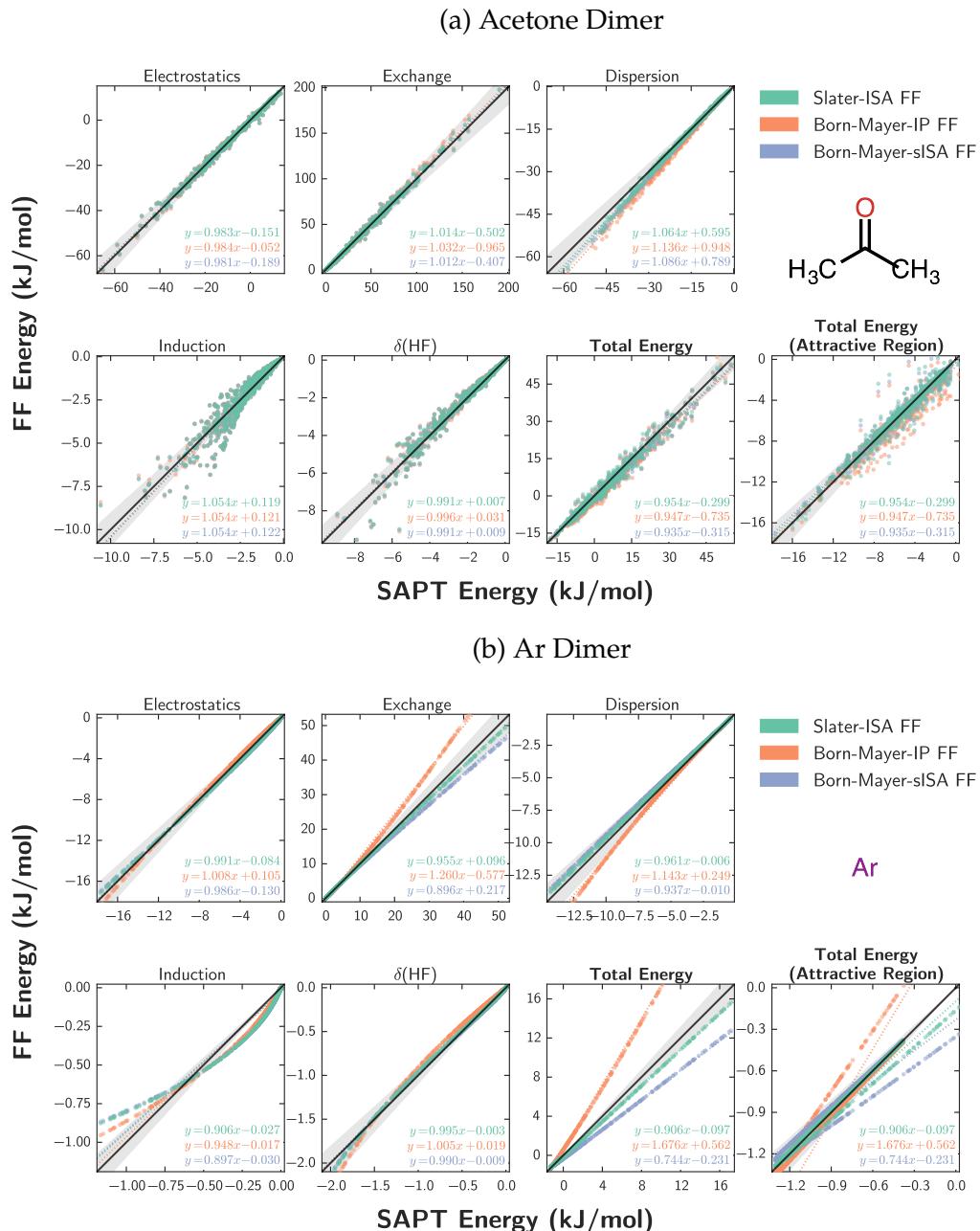
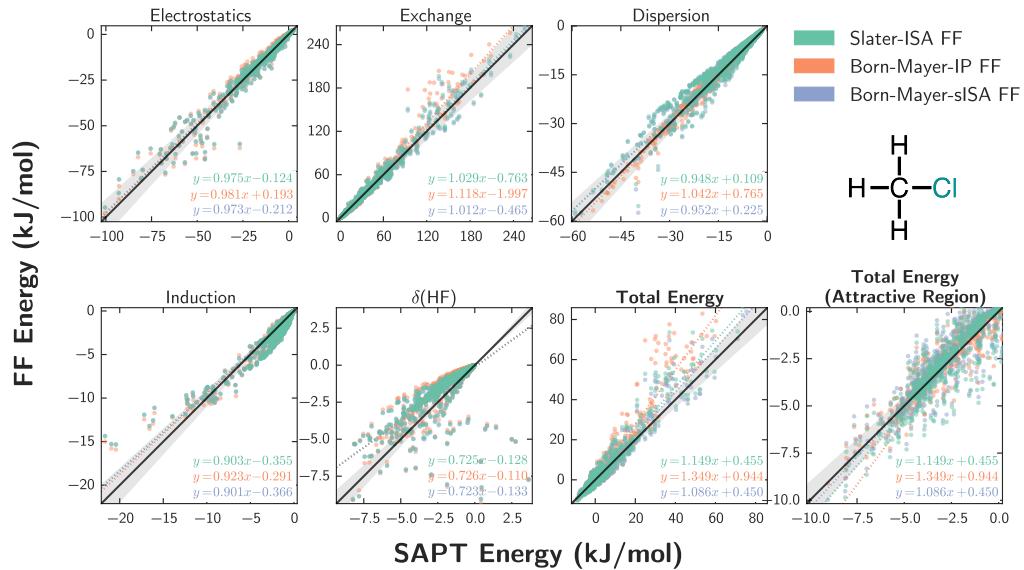
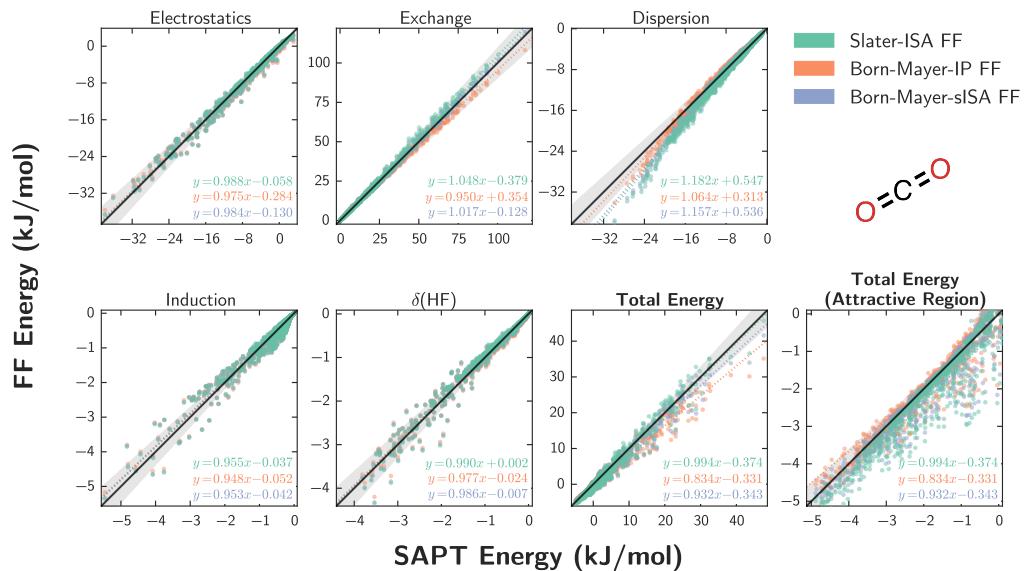


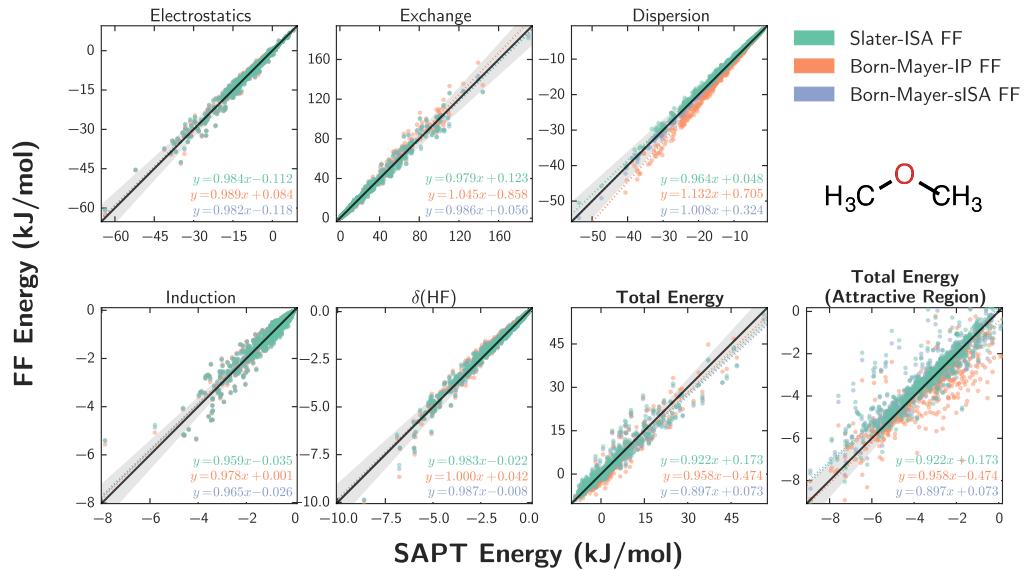
Figure 3.13: Waldman-Hagler-style analysis of possible B_{ij} combination rules. Exact B_{ij} values are derived from fitting an approximate overlap density of the form $S_{ij} = A_{ij} K_2(r_{ij}) \exp(-B_{ij}r_{ij})$ to the exact overlap density (as given by Rosen and by Tai^{7,8}) of two distinct Slater orbitals whose exponents correspond to atomic exponents for the elements H-Ar, Cl, Br, and I. For each overlap pair, a range of r_{ij} values was used from 0.8 to 1.2 times the sum of the pair's van der Waals radii. The geometric mean combination rule $B_{ij} = \sqrt{B_i B_j}$ models the exact B_{ij} values with near-perfect agreement, justifying our choice of combination rule.



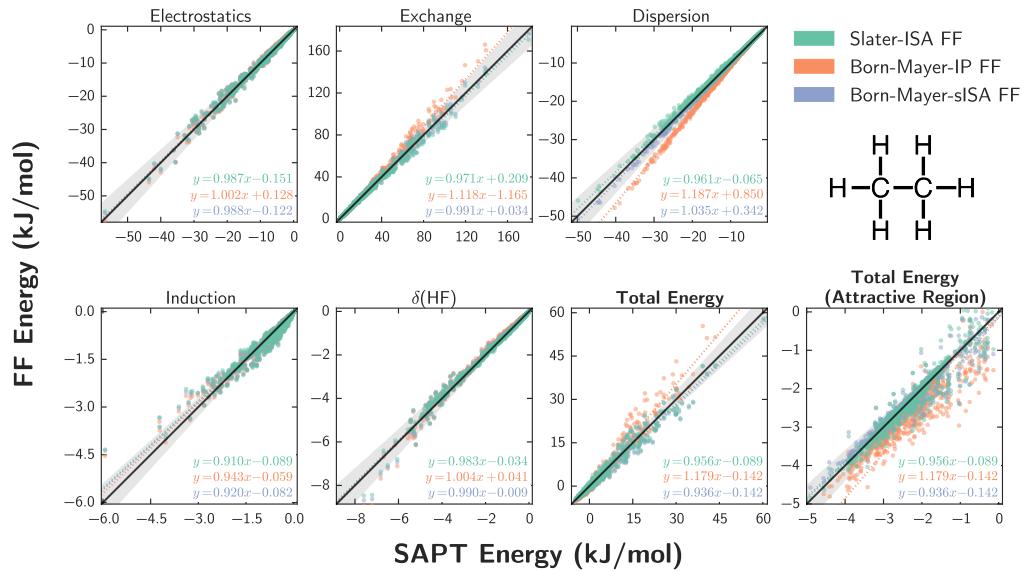
(c) Chloromethane Dimer

(d) CO₂ Dimer

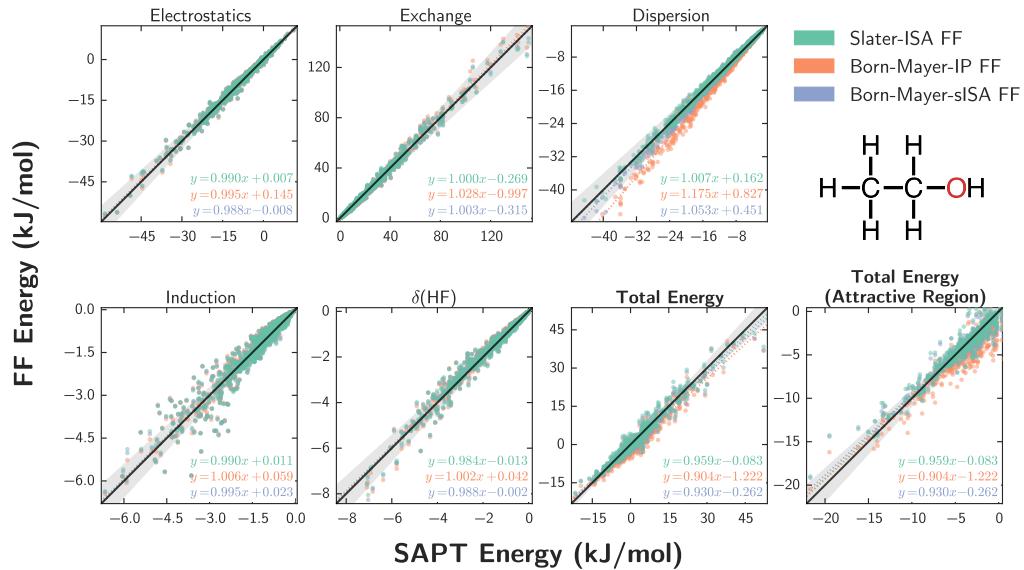
(e) Dimethyl Ether Dimer



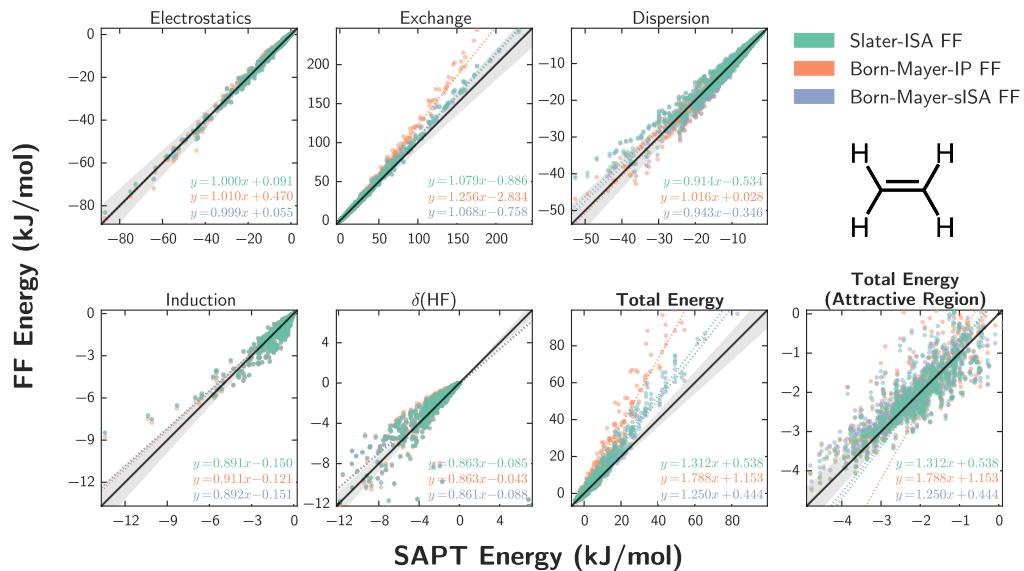
(f) Ethane Dimer

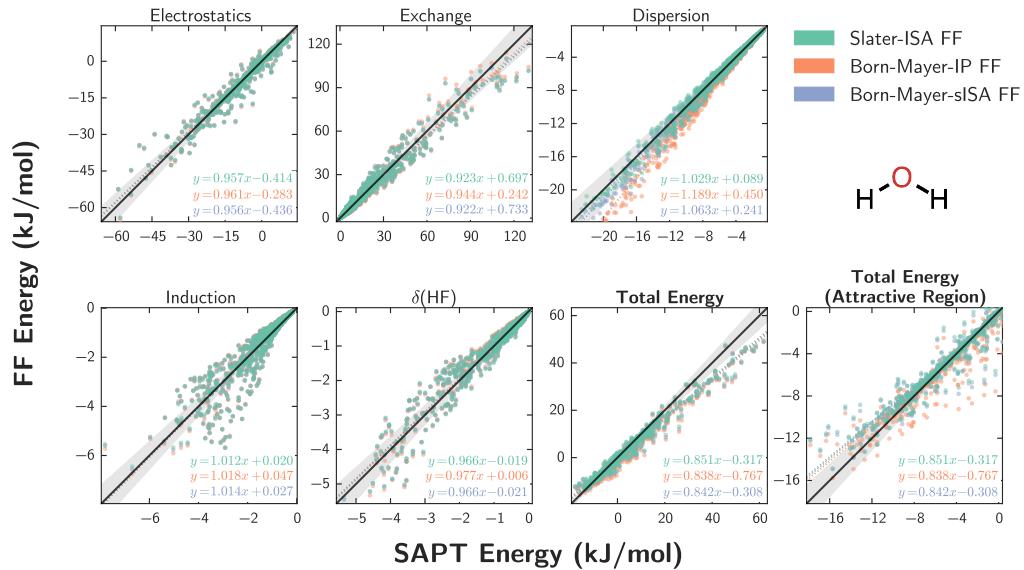


(g) Ethanol Dimer

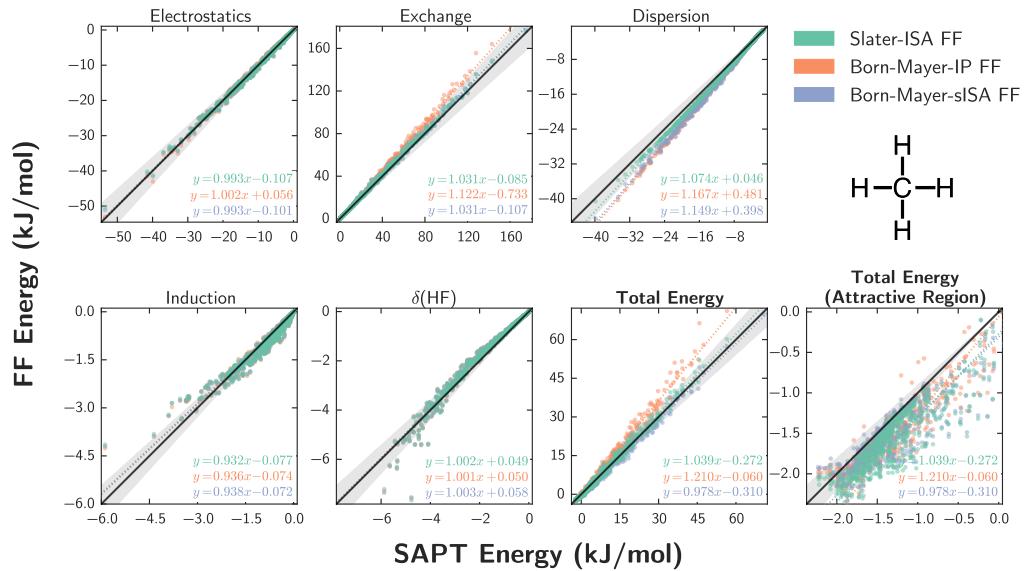


(h) Ethene Dimer

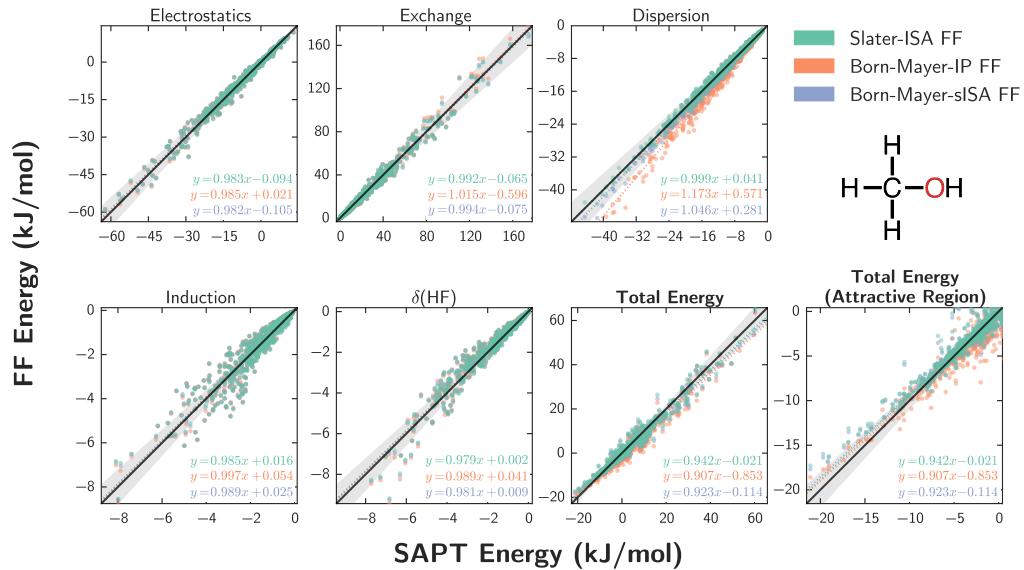


(i) H₂O Dimer

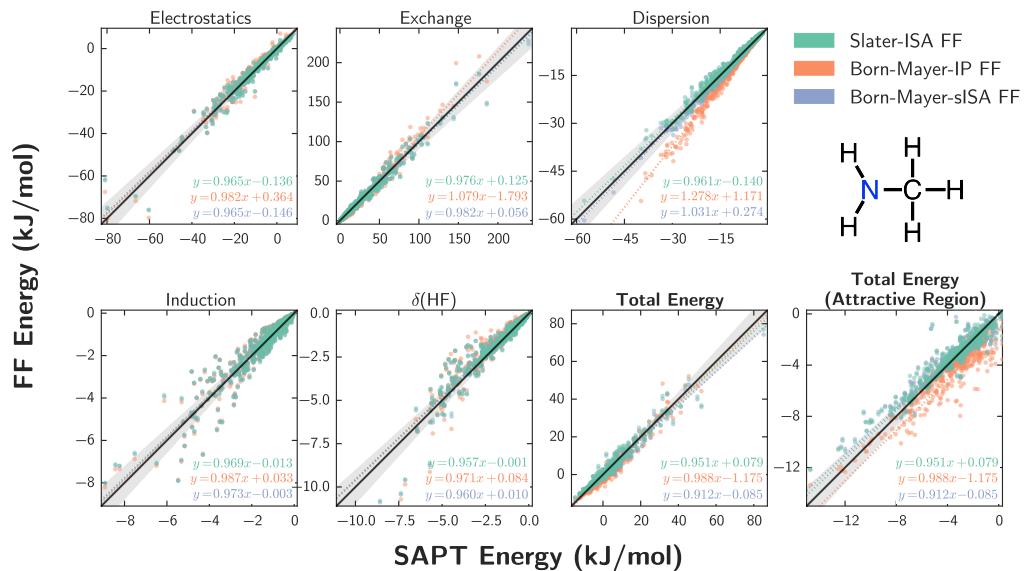
(j) Methane Dimer



(k) Methanol Dimer



(l) Methyl Amine Dimer



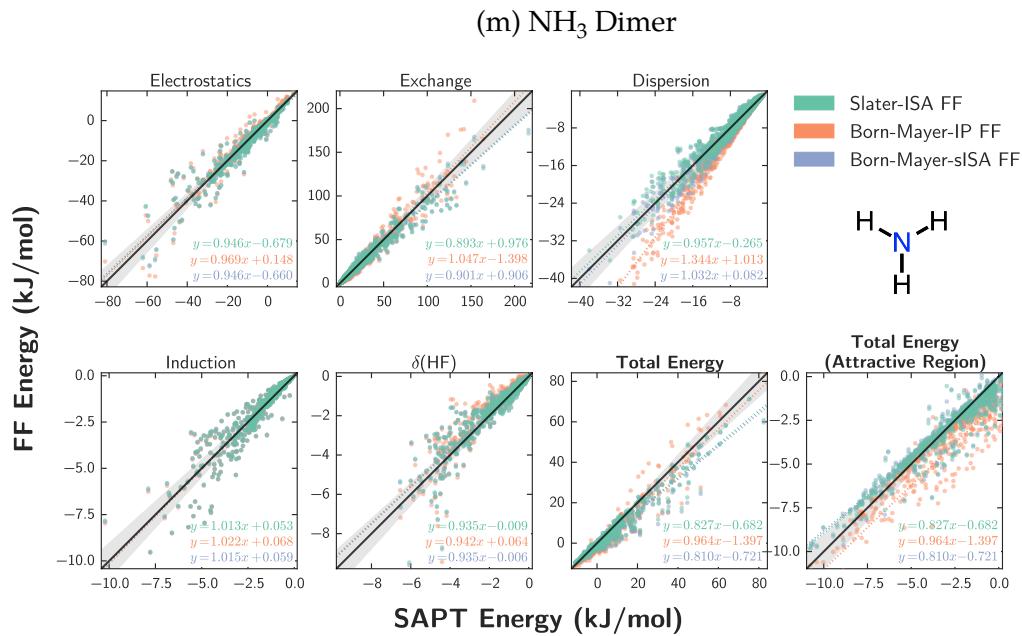


Figure 3.14: Force field fits for the homomeric systems using the Slater-ISA FF (green), Born-Mayer-IP FF (orange) and Born-Mayer-sISA FF (blue). Fits for each energy component are displayed along with two views of the total interaction energy. The $y = x$ line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within $\pm 10\%$ agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each energy component.

4 MASTIFF: A GENERAL APPROACH FOR INCORPORATING ATOMIC-LEVEL ANISOTROPY IN AB INITIO FORCE FIELDS

4.1 Introduction

Classical molecular simulation is a standard tool for interpreting and predicting the chemistry of an incredible host of systems ranging from simple liquids to complex materials and biomolecules. Such simulations always require, as input, a mathematical description of the system's potential energy surface (PES). In principle, the PES for most chemical systems can accurately be determined from one of several high-level electronic structure methods;^{174–176} nevertheless, these calculations are currently too expensive to use in simulations of large systems and/or long timescales.¹⁷⁷ Consequently, most routine molecular simulation today is performed with the aid of force fields: computationally-inexpensive, parameterized mathematical expressions that approximate the exact PES. Because the accuracy and predictive capabilities of molecular simulation are directly tied to the underlying force field, one of the central challenges of molecular simulation is the development of highly accurate force fields. For ab initio force field development, this accuracy is principally defined by a force field's fidelity to the underlying exact PES.

As of now, several common shortcomings⁷⁶ inhibit the accuracy and predictive capabilities of standard ab initio force fields, and these limitations must be systematically addressed in order to generate improved, 'next-generation' force fields. One important shortcoming, which will be the focus of this Chapter, is the so-called 'sum-of-spheres' approximation,¹⁷⁸ in which it is assumed that the non-bonding interactions between molecules can be treated as a superposition of interactions between pairs of spherically-symmetric atoms. Put differently, the sum-of-spheres, or 'isotropic atom-atom', approximation assumes that the exact PES, E_{int} (which depends both on the center of mass distance R and relative orientation Ω between

molecules), can be modeled as

$$E_{\text{int}}(R, \Omega) \approx \sum_{ij} f(r_{ij}) \equiv V_{\text{FF}}, \quad (4.1)$$

where the above sum runs over all non-bonded pairs of atoms i and j with interatomic separation r_{ij} , and $f(r_{ij})$ is an arbitrary, distance-dependent function that defines the pairwise interaction. Here and throughout, we use E to denote the true PES, and V to denote the corresponding model/force field prediction. With some exceptions (vida infra), nearly all standard intermolecular force fields — ranging from the popular “Lennard-Jones plus point charges” model to more complex functional forms⁸⁰ — explicitly make use of the isotropic atom-atom model.

Notwithstanding the popularity of the model, there is good experimental and theoretical evidence to suggest that the sum-of-spheres approximation does not hold in practice.^{48,178,179} Importantly, and as we argue in Section 4.5, models which include anisotropic (multipolar) electrostatics, but otherwise employ the sum-of-spheres approximation, are an improved but *still incomplete* model for describing the atomic-level anisotropy of intermolecular interactions. Experimentally, it has long been known that atom-in-molecule charge densities, as determined from x-ray diffraction, can exhibit significant non-spherical features, such as with lone pair or π electron densities.¹⁸⁰ Furthermore, statistical analyses of the Cambridge Structural Database have shown that the van der Waals radii of atoms-in-molecules (as measured from interatomic closest contact distances) are not isotropically distributed, but rather show strong orientation dependencies, particularly for halogens and other heteroatoms.^{170,181–185} These experimental studies are corroborated by a significant body of theoretical research on both the anisotropy of the atomic van der Waals radii as well as the non-spherical features of the atomic charge densities themselves.^{172,173,185–188} These studies suggest that the sum-of-spheres approximation is an insufficiently flexible model for the subset of intermolecular interactions that arise from atomically non-spherical charge densities, and may help explain known difficulties in generating accurate isotropic atom-atom force fields for such important chemical interactions as π -interactions,^{75,189,190} σ -bonding,^{191–193} and hydrogen

bonding,¹⁹⁴ (see Ref. 195 and references therein).

Motivated by these observations, a small but important body of work has been devoted to directly addressing the limitations of the isotropic atom-atom model in the context of ‘next-generation’ force field development. As will be discussed in detail below (see Section 4.2), the general conclusion from these studies is that many components of intermolecular interactions (specifically electrostatics, exchange-repulsion, induction, and dispersion) can be more accurately modeled by functional forms that go beyond the sum-of-spheres approximation.^{179,196,197} While few intermolecular potentials (and virtually no standard force fields amenable to routine molecular simulation) explicitly account for atomic-level anisotropy for each component of intermolecular interactions, several recent standard force fields have incorporated atomic-level anisotropy into their description of long-range electrostatics.¹⁹⁵ Some of these potentials (notably AMOEBA^{197–199} and some water potentials^{194,195}) are already employed in large-scale molecular simulation, often with very encouraging success.¹⁹⁵ Furthermore, others have shown that anisotropic potentials (some of which additionally model the anisotropy of exchange-repulsion and/or dispersion) lead to significant improvements in predicting molecular crystal structures.^{115,160,195,200–203} These and other results strongly suggest that a complete incorporation of atomic anisotropy into next-generation force fields will lead to increasingly accurate and predictive molecular simulations in a wider variety of chemical interactions.¹⁹⁶

Given the importance of atomic-level anisotropy in defining intermolecular interactions, and the critical role that computationally-affordable standard force fields play in enabling molecular simulation, our present goal is to develop a general methodology for standard force field development that can both universally account for atomic-level anisotropy in all components of intermolecular interactions *and* that can be routinely employed in large-scale molecular simulation. Furthermore, and in line with our usual goals for force field development,⁸⁰ our aim is to develop a first-principles-based model that is as accurate and transferable as possible, all while maintaining a simple, computationally-tractable functional form that allows for robust parameterization and avoids over-/under-fitting. Thus, building on

prior work (both our own^{51,58,80,104} and from other groups¹⁷⁹), we present here a general ansatz for anisotropic force field development that, at minimal computational overhead, incorporates atomic-level anisotropy into all aspects of intermolecular interactions (electrostatics, exchange, induction, and dispersion) and that accounts for this anisotropy, not only in the asymptotic limit of large intermolecular separations, but also in the region of non-negligible electron density overlap. After motivating and establishing the functional forms used in our anisotropic force fields, we next demonstrate, using a large library of dimer interactions between organic molecules, the excellent accuracy and transferability of these new force fields with respect to the reproduction of high-quality ab initio potential energy surfaces. Lastly, we showcase how these new force fields can be used in molecular simulation, and benchmark the accuracy of our models with regards to a variety of experimental properties. The theory and results presented in this Chapter should be of general utility in improving the accuracy of (particularly ab initio generated) force fields, such that the complex, inherently anisotropic details of intermolecular interactions may eventually be routinely incorporated into increasingly rigorous and predictive molecular simulation.

4.2 Background

Before presenting our development methodology for atomically-anisotropic potentials, we provide the reader with a summary of prior approaches to ab initio force field development and to models going beyond the sum-of-spheres approximation. In discussing the effects of anisotropic charge distributions on intermolecular potentials, here and throughout we employ the fairly standard²⁰⁴ decomposition of interaction energies into physically-meaningful components of electrostatics, exchange-repulsion, induction (which includes both polarization and charge-transfer), and dispersion. Many studies on atomically-anisotropic force field development have focused on incorporating anisotropy on a component-by-component basis, and so for clarity we discuss anisotropic modeling schemes for each energy component individually. As in Chapter 3,¹⁰⁴ we find it useful to separate and discuss in turn

the so-called ‘long-range’ effects (multipolar electrostatics, polarization, and dispersion) from those ‘short-range’ effects that arise only at shorter intermolecular separations due to the non-negligible overlap of monomer electron densities (e.g. charge penetration and exchange-repulsion). Finally, we take advantage of the many-body expansion^{51,205} to separately consider into two- and many-body contributions, and primarily focus our discussion on improvements to the two-body interaction energies themselves.

4.2.1 Prior Models for Long-Range Interactions

The importance of atomic-level anisotropy in modeling long-range interactions, particularly as it pertains to electrostatics, is quite well studied. A number of groups have found that using atomic multipoles (rather than simple point charges) greatly improves both the electrostatic potential^{186,206} and the resulting electrostatic interaction energies.^{89,114,195,197,199,207–209} Though not without additional computational cost, atomic multipoles are now routinely employed in a number of popular force fields.^{194,197,199} As an alternate, and often more computationally-affordable approach, other groups have used off-atom point charges to effectively account for anisotropic charge densities.^{92,192,210,211} In line with chemical intuition, improvements from use of atomic multipoles/off-site charges are often particularly important in describing the electric fields generated by heteroatoms and carbons in multiple bonding environments.^{212,213}

The induction and dispersion energies have also been shown to exhibit anisotropies that go beyond the sum-of-spheres model. For instance, it has been suggested that anisotropic polarizabilities (which effect both polarization and dispersion) are required to avoid an artificial over-stabilization of base stacking energies in biomolecules.¹⁹⁰ In order to more accurately treat polarization, several molecular mechanics potentials have made use of either off-site²¹⁴ or explicitly anisotropic polarizabilities.^{211,215} Similarly, the importance of anisotropic dispersion interactions has also been established,^{51,54,216–218} particularly for π -stacking interactions,^{76,190} and select potentials have incorporated directional dependence into the functional form

for dispersion by expanding dispersion coefficients in terms of spherical harmonics or, more generally, \bar{S} -functions (discussed in Section 4.A).^{203,216}

4.2.2 Prior Models for Short-Range Interactions

At closer intermolecular separations, where overlapping electron densities between monomers leads to exchange-repulsion and charge-penetration effects, anisotropy is also important. Exchange-repulsion has known orientation dependencies which can play a quantitative role in halogen bonding^{191,219} and other chemical interactions, and many authors have worked on developing different models for describing the anisotropy of exchange-repulsion. Some potentials (albeit not those that are amenable to large-scale molecular simulation) have numerically computed overlap integrals that can be used in conjunction with the density-overlap model popularised by Wheatley and Price^{84,86,87,111,123} to quantify anisotropic exchange-repulsion, charge transfer, and/or charge penetration interactions.^{89,208,209,220–222} Taking a more analytical approach, many other potentials have extended the Born-Mayer functional form⁶⁴ to allow for orientation-dependent pre-factors,^{51,115,118,178,179,200,203,222,223} and model short-range effects using an anisotropic functional form originally proposed by Stone and Price:

$$V_{ij}^{\text{exch}} = G \exp[-\alpha_{ij}(R_{ij} - \rho_{ij}(\Omega_{ij}))]. \quad (4.2)$$

Here G is not a parameter, but rather an energy unit,⁴⁸ and α and ρ represent, respectively, the hardness and shape of the potential. In principle, one might also allow α to have orientation dependence, however this seems unnecessary in practice, as the hardness of the potential has been empirically found to behave more isotropically than its shape.⁴⁸ Similar to treatments of anisotropic electrostatics, this functional form typically expresses orientation dependence, Ω_{ij} , in terms of spherical harmonics and/or \bar{S} -functions.⁴⁸

Finally, we note that, aside from exchange-repulsion, we are aware of relatively little research on the development of simple analytical expressions for the anisotropy

of other overlap effects, such as electrostatic/inductive charge penetration, charge-transfer, or short-range dispersion.

4.3 Theory and Motivation

Building on the extensive prior work that has led to a better understanding of atomic-level anisotropy and its effect(s) on intermolecular interactions, we now outline a methodology whereby atomic-level anisotropy can be incorporated into standard force fields amenable to large-scale molecular simulation. In particular, we aim to present a general methodology that optimally incorporates atomically-anisotropic effects given the following goals for ab initio force field development:

1. **Chemical accuracy with respect to ab initio benchmarks:** For systems that can be directly parameterized against high quality ab initio PES, the force field should exhibit chemical accuracy (average errors smaller than 1 kJ mol^{-1}) with respect to the ab initio benchmark; furthermore, any errors in the force field should be random rather than systematic
2. **Transferability across chemical environments:** Given force fields for two different pure systems, we should be able to accurately calculate (via simple combination rules and without additional parameterization) the PES of any system that is a mixture of the pure systems
3. **Simplicity:** The force field should be restricted to functional forms that are already compatible with, or could be easily implemented in, common molecular simulation packages
4. **Computational tractability:** The force field should be of minimal computational cost relative to existing polarizable multipolar force fields¹⁹⁹

Given these goals, we now outline a detailed methodology for incorporating atomic-level anisotropy into each component (electrostatic, exchange-repulsion, induction, and dispersion) of intermolecular interactions, beginning with a new

model for short-range overlap effects and concluding with some new and/or revised theories for treating long-range interactions.

4.3.1 Anisotropic Models for Short-Range Interactions

Exchange-Repulsion

We begin by considering the exchange-repulsion, E_{ij}^{exch} that arises from the overlap of electron densities from two non-spherical atoms-in-molecules, i and j . Here and throughout, we closely follow the notation and theory used by Stone.⁴⁸ Without loss of generality, we can express the exchange repulsion between these two atoms as a function of their interatomic distance, r_{ij} , and relative orientation, Ω_{ij} . Furthermore, we can mathematically describe this relative orientation by assigning local coordinate axes to each i and j , such that the exchange energy is given by

$$E_{ij}^{\text{exch}}(r_{ij}, \Omega_{ij}) \equiv E_{ij}^{\text{exch}}(r_{ij}, \theta_i, \phi_i, \theta_j, \phi_j), \quad (4.3)$$

where θ_i and ϕ_i are the polar coordinates, expressed in the local coordinate system of atom i , that describe the position of atom j . Correspondingly, θ_j and ϕ_j define the position of i in terms of the local coordinate system of j . In principle the choice of these local coordinate frames is arbitrary. However, for the models introduced below, parameterization can be dramatically simplified by taking advantage of the local symmetry of an atom in its molecular environment and aligning the local coordinate frame with the principal axis of this local symmetry.⁴⁸ Some examples of these local axes are shown in Fig. 4.1.

We next make an ansatz that Eq. (4.3) is separable into radial- and angular-dependent contributions,

$$E_{ij}^{\text{exch}}(r_{ij}, \theta_i, \phi_i, \theta_j, \phi_j) \approx V_{ij}^{\text{exch}}(r_{ij}, \theta_i, \phi_i, \theta_j, \phi_j) = f(r_{ij})g(\theta_i, \phi_i, \theta_j, \phi_j) \quad (4.4)$$

thus subdividing the problem of finding a general functional form for E_{ij}^{exch} into two more tractable tasks. First, we must find an ideal sum-of-spheres model to describe

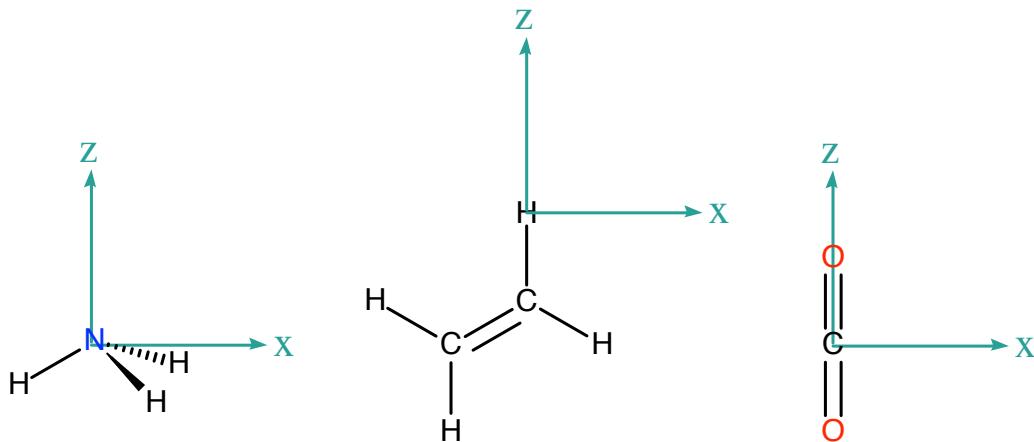


Figure 4.1: Local axis system, shown for select atoms in molecules.

the radial (isotropic) dependence of the force field, and second, we must find a way to model the orientation dependence as a multiplicative pre-factor to $f(r_{ij})$.

Given that the only requirement for $f(r_{ij})$ is that it be isotropic, how should a suitable model for $f(r_{ij})$ be chosen? Indeed, all standard isotropic force fields are of this general form, and thus might serve as a suitable starting point for anisotropic force field development. For reasons discussed below, in this Chapter we employ a simple and accurate model (Slater-ISA FF) from Chapter 3 for $f(r_{ij})$. This model can be derived from first-principles by approximating E_{ij}^{exch} as proportional to the overlap between spherically-symmetric atom-in-molecule (AIM) electron densities, each with density

$$\rho_i(r) = D_i \exp^{-B_i r}, \quad (4.5)$$

where D_i and B_i are both atom type-specific constants that can be parameterized from molecular electron densities and that represent, respectively, the shape and hardness of the AIM density. Using this approximation to the overlap model,^{51,84,86,87,115,123–125}

the exchange energy between two atoms is then modeled by

$$\begin{aligned} E_{ij}^{\text{exch}} &\approx V_{ij}^{\text{exch}} \propto S_{\rho}^{ij} \\ &\approx A_{ij}^{\text{exch}} \left(\frac{(B_{ij}r_{ij})^2}{3} + B_{ij}r_{ij} + 1 \right) \exp(-B_{ij}r_{ij}) \end{aligned} \quad (4.6)$$

with combining rules

$$\begin{aligned} A_{ij}^{\text{exch}} &\equiv A_i^{\text{exch}} A_j^{\text{exch}}, \\ B_{ij} &\equiv \sqrt{B_i B_j}. \end{aligned} \quad (4.7)$$

S_{ρ}^{ij} is the electron density overlap between atoms and A_{ij} is a fitted proportionality constant.

Here and throughout we use Eq. (4.6) as our model for $f(r_{ij})$. This choice is primarily justified in Chapter 3 by the previously-demonstrated accuracy of the Slater-ISA formalism as compared to other sum-of-spheres models for repulsion.¹⁰⁴ Furthermore, and especially for simple test cases where one might expect the sum-of-spheres approximation to hold (such as with argon, methane, or ethane), we have shown (see Chapter 3) that the Slater-ISA FF correctly models intermolecular potential energy surfaces for a sizable library of intermolecular interactions over the asymptotic, attractive, and repulsive regions of the PES.

In addition to this empirical motivation for using the Slater-ISA formalism, there are good theoretical grounds to utilize it as a model for $f(r_{ij})$. Specifically, the AIM densities used to parameterize Slater-ISA FF are partitioned using an iterated stockholder atoms (ISA) procedure, and the resulting density profiles are guaranteed to be maximally spherical.^{58,134,135} This condition of ‘maximum sphericity’ has two consequences. First, it suggests that the resulting Slater-ISA FF should be an optimal, or nearly optimal, isotropic atom-atom model. In other words, there is good reason to hope that our model for $f(r_{ij})$ completely accounts for the radial dependence of the potential, and consequently that models for $g(\theta_i, \phi_i, \theta_j, \phi_j)$ will truly represent the orientation dependence rather than simply over-fitting

residual errors from the radial functional form, thus retaining high transferability. Second, and relatedly, having maximally-spherical ISA densities suggests that anisotropic effects should be a minimal perturbation to the PES. This means that, to a first-order approximation, $g(\theta_i, \phi_i, \theta_j, \phi_j)$ is simply equal to 1. Furthermore, the non-spherical components of the ISA densities should provide us with guidance as to which atom types might require anisotropic treatment.

With the functional form for $f(r_{ij})$ determined, we now describe our model for $g(\theta_i, \phi_i, \theta_j, \phi_j)$. As motivated in Appendix 4.A, and under the ansatz of radial and angular separability, an approximate, transferable, and orientation-dependent expression for A_i^{exch} can be obtained by expanding A_i^{exch} in a basis of renormalized spherical harmonics,

$$C_{lm}(\theta, \phi) = \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\theta, \phi). \quad (4.8)$$

thus yielding

$$\begin{aligned} A_i^{\text{exch}}(\theta_i, \phi_i) &= A_{i,\text{iso}}^{\text{exch}}(1 + \xi^{\text{exch}}(\theta_i, \phi_i)), \\ \xi^{\text{exch}}(\theta_i, \phi_i) &\equiv \sum_{l>0,k} a_{lk}^{\text{exch}} C_{lk}(\theta_i, \phi_i) \end{aligned} \quad (4.9)$$

for A_i^{exch} and subsequently

$$V_{ij}^{\text{exch}} = A_{ij}^{\text{exch}}(\Omega_{ij}) \left(\frac{(B_{ij}r_{ij})^2}{3} + B_{ij}r_{ij} + 1 \right) \exp(-B_{ij}r_{ij}) \quad (4.10)$$

with

$$A_{ij}^{\text{exch}}(\Omega_{ij}) = A_i^{\text{exch}}(\theta_i, \phi_i) A_j^{\text{exch}}(\theta_j, \phi_j) \quad (4.11)$$

for the exchange-repulsion potential. Note that, with the exception of the now orientation-dependent A_i^{exch} , the atomically-anisotropic model in Eq. (4.10) is identical to our previously-defined isotropic model (Eq. (4.6)).

In terms of parameterization for our newly-developed anisotropic model, note

that the a_{lk}^{exch} are free parameters which must be fit to ab initio data. Still, we and others have found the expansion in Eq. (4.9) to be very quickly convergent,^{51,115,118,178,179,200,203,222,223} especially given a proper choice of coordinate system that eliminates many expansion terms via symmetry. In practice, only symmetry-allowed terms up to $l = 2$ seem to be required for heteroatoms, carbons in multiple bonding environments, and select hydrogens (see equations in Section 4.5), while many other atom types require no anisotropic parameters whatsoever. Encouragingly, isotropic atom types are easily modeled within this formalism simply by setting $\xi(\theta_i, \phi_i) = 0$.

Other Short-Range Effects

As in Chapter 3,¹⁰⁴ we have found that other short-range effects, namely charge penetration and short-range induction, can be modeled as proportional to exchange-repulsion. We take the same approach in the present Chapter, and the functional form for these two short-range effects is given by Eq. (4.10), with ‘exch’ superscripts replaced by the appropriate short-range energy term (see Section 4.4). Additionally, for induction, the long-range polarization must be damped, and for now this damping is modeled isotropically as in the AMOEBA force field.¹⁹⁹ Finally, to model short-range dispersion, we take the same Tang-Toennies^{121,122} damping approach as in Chapter 3.¹⁰⁴ Because the argument to the damping function is given by

$$\chi = -\frac{d}{dr} [\ln V^{\text{exch}}(r)] r,$$

and because anisotropy only enters into the functional form as a multiplicative pre-factor, our functional form for damping remains unchanged compared to our previously-derived isotropic model.¹⁰⁴

4.3.2 Anisotropic Models for Long-Range Interactions

Electrostatics

Theories for including anisotropy in long-range electrostatics are well established, and we refer the reader elsewhere for complete details on the required formalisms

for distributed multipole approaches.^{48,51} In the present Chapter,

$$V_{ij}^{\text{multipole}} = \sum_{tu} Q_t^i T_{tu} Q_u^j$$

with multipolar interaction tensor T and parameterized moments Q for all multipole moments tu up to (in the present Chapter) rank 2.

On the grounds of increased accuracy and ease of parameterization, here we have chosen to use a multipolar approach to describe the anisotropy of long-range electrostatics. However, for increased computational efficiency, off-site point charge models¹⁹⁵ could also be utilized.

Induction

Just as with electrostatics, long-range induction should properly be described by a distributed multipole expansion of interacting atomic polarizabilities.^{51,203} Indeed, it has been shown that inclusion of higher-order and/or anisotropic polarizabilities greatly reduces errors in the two-body induction potential relative to commonly-used isotropic dipole polarizability models.^{80,199,224–226} Because the model for the two-body induction also determines the many-body polarization energy, the proper treatment of induced multipoles becomes especially important in condensed phase simulation.^{48,80,199}

Owing to the increased computational cost of these higher-order and anisotropic polarizability models, and because such functional forms are (as of now, and to our knowledge) not fully implemented in common molecular simulation packages, we neglect both higher-order and anisotropic contributions to the long-range induction in the present Chapter. As we shall show, however, errors in the induction potential limit the overall accuracy of our force fields for extremely polar molecules (notably water), and further improvements will likely require us to generate improved models for long-range induction.

Dispersion

Past research⁴⁸ has motivated an anisotropic atom-atom model for dispersion of the form

$$V_{ij}^{\text{disp}} = - \sum_{n=6} \frac{C_{ij,n}(\Omega_{ij})}{r_{ij}^n} \quad (4.12)$$

Note that, in this equation, both odd and even powers of n are allowed in the dispersion expansion. In order to make this model both computationally efficient and maximally compatible with our previous isotropic model for dispersion, we choose (as an ansatz) to model the dispersion anisotropy as an orientation-dependent prefactor that effects all isotropic $C_6 - C_{12}$ dispersion coefficients equally:

$$V_{ij}^{\text{disp}} = -A_i^{\text{disp}} A_j^{\text{disp}} \sum_{n=3}^6 \frac{C_{ij,2n}}{r_{ij}^{2n}} \quad (4.13)$$

with

$$A_i^{\text{disp}} = 1 + \xi^{\text{disp}}(\theta_i, \phi_i) \quad (4.14)$$

and $\xi^{\text{disp}}(\theta_i, \phi_i)$ as in Eq. (4.9). Once again, Eq. (4.13) reduces to the isotropic case by setting $\xi^{\text{disp}}(\theta_i, \phi_i) = 0$. We must note that, though the functional form in Eq. (4.13) bears many similarities to Eq. (4.12), (unphysically) no odd powers of r show up in our proposed model for dispersion. Furthermore, the model utilizes the same anisotropic expansion for each dispersion coefficient. Nonetheless, we will show in Section 4.5 that this model yields significant accuracy gains in the dispersion energy with only minimal additional parameterization and model expense.

4.4 Technical Details

4.4.1 The 91 Dimer Test Set

Our benchmarking procedures are the same as in Chapter 3,¹⁰⁴ and we briefly summarize the relevant technical details. A full discussion of results and example calculations are presented in Section 4.5.

We have previously developed a large library of benchmark energies for interactions between the following 13 atomic and organic species: acetone, argon, ammonia, carbon dioxide, chloromethane, dimethyl ether, ethane, ethanol, ethene, methane, methanol, methyl amine, and water. Using these 13 monomers, we have generated a library of dimer interaction energies for each of the 91 possible unique dimer combinations (13 homomeric, 78 heteromeric). For each of these dimer combinations, interaction energies were computed at a DFT-SAPT^{137–145} level of theory for 1000 quasi-randomly chosen dimer configurations, representing 91,000 benchmark interaction energies in total. As described below, parameters for a given force field methodology are then fit on a component-by-component basis to reproduce the benchmark DFT-SAPT energies.

4.4.2 Parameter Determination

We will present three types of force field fitting methodologies in this Chapter, termed Iso-Iso FF, Aniso-Iso FF, and Aniso-Aniso FF (alternately referred to as MASTIFF, as discussed below). The nomenclature of each name refers to, first, the isotropic/anisotropic treatment of multipolar electrostatics and, second, the isotropic/anisotropic treatment of dispersion and short-range effects. All studied force fields use the following general functional form:

$$V_{\text{FF}} = \sum_{ij} V_{ij}^{\text{exch}} + V_{ij}^{\text{elst}} + V_{ij}^{\text{ind}} + V_{ij}^{\delta\text{HF}} + V_{ij}^{\text{disp}} \quad (4.15)$$

where

$$\begin{aligned}
V_{ij}^{\text{exch}} &= A_{ij}^{\text{exch}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\
V_{ij}^{\text{elst}} &= -A_{ij}^{\text{elst}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + \sum_{tu} Q_t^i T_{tu} Q_u^j \\
V_{ij}^{\text{ind}} &= -A_{ij}^{\text{ind}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(2)} \\
V_{ij}^{\delta^{\text{HF}}} &= -A_{ij}^{\delta^{\text{HF}}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(3-\infty)} \\
V_{ij}^{\text{disp}} &= -A_{ij}^{\text{disp}} \sum_{n=3}^6 f_{2n}(x) \frac{C_{ij,2n}}{r_{ij}^{2n}} \\
P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\
A_{ij} &= A_i A_j \\
B_{ij} &= \sqrt{B_i B_j} \\
C_{ij,2n} &= \sqrt{C_{i,2n} C_{j,2n}} \\
f_{2n}(x) &= 1 - e^{-x} \sum_{k=0}^{2n} \frac{(x)^k}{k!} \\
x &= B_{ij} r_{ij} - \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij} r_{ij} + 3} r_{ij}
\end{aligned} \tag{4.16}$$

For both Iso-Iso FF and Aniso-Iso FF, A_i is a fit parameter, and $A_{ij}^{\text{disp}} = 1$. For Iso-Iso FF (our completely isotropic model), the multipole expansion $\sum_{tu} Q_t^i T_{tu} Q_u^j$ is truncated to point charges, whereas Aniso-Iso FF and MASTIFF both use a multipole expansion up to quadrupoles. Finally, for our anisotropic model, MASTIFF, each A_i is treated as an orientation-dependent function, and is represented by the spherical harmonic expansion

$$\begin{aligned}
A_i(\theta_i, \phi_i) &= A_{i,\text{iso}}(1 + \xi(\theta_i, \phi_i)), \\
\xi(\theta_i, \phi_i) &\equiv \sum_{l>0,k} a_{i,lk} C_{lk}(\theta_i, \phi_i)
\end{aligned} \tag{4.17}$$

where $A_{i,\text{iso}}$ and $a_{i,\text{lk}}$ are fitted parameters with the exception that $A_{i,\text{iso}}^{\text{disp}} = 1$.

Because DFT-SAPT provides a physically-meaningful energy decomposition into electrostatic, exchange-repulsion, induction, and dispersion terms, parameters for each term in Eq. (4.15) are directly fit to model the corresponding DFT-SAPT energy (see Ref. 104 and references therein for details on the DFT-SAPT terminology):

$$\begin{aligned} V^{\text{exch}} &\approx E^{\text{exch}} \equiv E_{\text{exch}}^{(1)} \\ V^{\text{elst}} &\approx E^{\text{elst}} \equiv E_{\text{pol}}^{(1)} \\ V^{\text{ind}} &\approx E^{\text{ind}} \equiv E_{\text{ind}}^{(2)} + E_{\text{ind-exch}}^{(2)} \\ V^{\delta^{\text{HF}}} &\approx E^{\delta^{\text{HF}}} \equiv \delta(\text{HF}) \\ V^{\text{disp}} &\approx E^{\text{disp}} \equiv E_{\text{disp}}^{(2)} + E_{\text{disp-exch}}^{(2)}. \end{aligned} \quad (4.18)$$

Fitting parameters on a component-by-component basis helps ensure parameter transferability and minimizes reliance on error cancellation. Note that no parameters are fit to reproduce the total energy and that, because the DFT-SAPT energy decomposition is only calculated to second-order, third- and higher-order terms (mostly consisting of higher-order induction) are estimated by $E^{\delta^{\text{HF}}}$.

Parameters Calculated from Monomer Properties

Of the parameters listed in Eq. (4.16), most do not need to be fit to the DFT-SAPT energies, but can instead be calculated directly on the basis of monomer electron densities. In particular, all multipolar coefficients, Q , polarizabilities (involved in the calculation of V_{pol}), dispersion coefficients C , and atom-in-molecule exponents, B^{ISA} , are calculated in a manner nearly identical to Ref. 104. Note that, for our atom-in-molecule exponents, we tested the effects of treating B^{ISA} both as a hard- and as a soft-constraint in the final force field fit. While the conclusions from this study are rather insensitive to this choice of constraint methodology, we have found that the overall force field quality is somewhat improved by relaxing the B^{ISA} coefficients in the presence of a harmonic penalty function (technical details of which can be found in the Supporting Information of Ref. 104). The optimized B coefficients

in this Chapter are always within 5–10% of the calculated B^{ISA} coefficients from Chapter 3, demonstrating the good accuracy of the B^{ISA} calculations themselves.

As a second distinction from our prior work, and for reasons of compatibility with the OpenMM¹⁵⁷ software we use for all molecular dynamics simulations, here our molecular simulations use an induced dipole model to describe polarization effects. Numerical differences between this model and the drude model used previously are very minor. Additionally, the Thole-damping functions used in this Chapter follow the same functional form used in the AMOEBA model,¹⁹⁷ with a damping parameter of 0.39.

Parameters Fit to Dimer Properties

In addition to the soft-constrained B parameters, all other free parameters (A and a parameters from Eq. (4.15) and Eq. (4.17)) are fit to reproduce DFT-SAPT energies from the 91 dimer test set described above. For each dimer pair, 4–5 separate optimizations (for exchange, electrostatics, induction, δHF , and, for MASTIFF, dispersion) were carried out to minimize a weighted least-squares error, with the weighting function given by a Fermi-Dirac functional form,

$$w_i = \frac{1}{\exp(-E_i/kT) + 1}, \quad (4.19)$$

where E_i is the reference energy and the parameter kT , which sets the energy scale for the weighting function, is calculated from an estimate of the global minimum well depth, E_{\min} , such that $kT = 5.0|E_{\min}|$.

Local Axis Determination

Identically to AMOEBA and other force fields that incorporate some degree of atomic-level anisotropy,^{115,118,197} we use a z-then-x convention to describe the relative orientation of atomic species. By design, the z-axis is chosen to lie parallel to the principal symmetry axis (or approximate local symmetry axis) of an atom in its molecular environment, and the xz-plane is similarly chosen to correspond to a

secondary symmetry axis or plane. Based on the assigned symmetry of the local reference frame, many terms in the spherical expansion of Eq. (4.9) can then be set to zero, minimizing the number of free parameters that need to be fit to a given atom type. Representative local reference frames are shown for a few atom types in Fig. 4.1, and a complete listing of anisotropic atom types (along with their respective local reference frames and non-zero spherical harmonic expansion terms) are given in the Section 4.B.

CCSD(T) Force Fields

DFT-SAPT is known to systematically underestimate the interaction energies of hydrogen-bonding compounds, and can also exhibit small but important errors for dispersion-dominated compounds.²²⁷ Consequently, for simulations involving CO₂, CHCl₃, NH₃, and H₂O, we refit our SAPT-based force fields to reproduce benchmark supermolecular, counterpoise-corrected CCSD(T)-F12a/aVTZ calculations on the respective dimers. All calculations were performed using the Molpro 2012 software.²²⁸ Fits were still performed on a component-by-component basis, with the energy of most components matching the DFT-SAPT calculations used in Chapter 3.¹⁰⁴ However, so that the total benchmark energy corresponded to the total interaction energy calculated by CCSD(T)-F12a/aVTZ, the difference between coupled-cluster and SAPT energies was added to the SAPT dispersion energy,

$$\begin{aligned}
 V^{\text{exch}} &\approx E^{\text{exch}} \equiv E_{\text{exch}}^{(1)} \\
 V^{\text{elst}} &\approx E^{\text{elst}} \equiv E_{\text{pol}}^{(1)} \\
 V^{\text{ind}} &\approx E^{\text{ind}} \equiv E_{\text{ind}}^{(2)} + E_{\text{ind-exch}}^{(2)} \\
 V^{\delta^{\text{HF}}} &\approx E^{\delta^{\text{HF}}} \equiv \delta(\text{HF}) \\
 V^{\text{disp}} &\approx E^{\text{disp}} \equiv E_{\text{disp}}^{(2)} + E_{\text{disp-exch}}^{(2)} + \delta(\text{CC}),
 \end{aligned} \tag{4.20}$$

where $\delta(\text{CC}) \equiv E_{\text{int}}^{\text{CCSD(T)-F12a}} - E_{\text{int}}^{\text{DFT-SAPT}}$.

In fitting these CCSD(T)-f12a-based force fields, and to account for small errors in the original SAPT dispersion energy, we somewhat relaxed the constraint that

$A^{\text{disp}} = 1$ for all atom types, and instead let $0.7 \leq A^{\text{disp}} \leq 1.3$. This constraint relaxation led, in some cases, to modest improvements in the fitted potential.

CO₂ 3-body potential

For the CO₂ dimer, we developed a three-body model to account for three-body dispersion effects. This three-body model is based on the three-body dispersion Axilrod-Teller-Muto (ATM) type model developed by Oakley and Wheatley²²⁹. These authors fit the ATM term with the constraint that the total molecular C₉ coefficient be 1970 a.u. Based on our own calculations using a CCSD/AVTZ level of theory,²³⁰ we have obtained an isotropic molecular C₉ coefficient of 2246 a.u.; consequently, a 1.13 universal scale factor was introduced to the Oakley potential so as to obtain dispersion energies in line with this new dispersion coefficient.

4.4.3 Simulation Protocols

ΔH_{sub} for CO₂

For CO₂, the molar enthalpy of sublimation was determined according to

$$\begin{aligned}\Delta H_{\text{sub}} &= H_g - H_{\text{crys}} \\ &= (U_g + PV_g) - (U_{\text{el,crystal},0K} + \Delta U_{\text{el,crystal},0K \rightarrow T_{\text{sub}}} + PV_{\text{crys}} + E_{\text{vib,crystal}}) \quad (4.21) \\ &\approx (RT) - \left(U_{\text{el,crystal},0K} + \int_{0K}^{T_{\text{sub}}} C_p dT + E_{\text{vib,crystal}} \right)\end{aligned}$$

which assumes ideal gas behavior and $PV_g \gg PV_{\text{crys}}$. For the crystal, an experimental measure of C_p was obtained from Ref. 231 and numerically integrated to obtain a value $\Delta U_{\text{el,crystal},0K \rightarrow T_{\text{sub}}} = 6.70 \text{ kJ mol}^{-1}$. Theoretical measures of $E_{\text{vib,crystal}} \approx 2.24 - 2.6 \text{ kJ mol}^{-1}$ were obtained from (respectively) Ref. 232 and Ref. 233, and $U_{\text{el,crystal},0K}$ was determined from the intermolecular force field using a unit cell geometry taken from experiment.²³⁴

Other CO₂ Simulations

To determine the densities and enthalpies of vaporization used in this Chapter, simulations were run in OpenMM using NPT and NVT ensembles, respectively. After an equilibration period of at least 100ps, data was collected for a minimum of 500ps, and uncertainties were calculated using the block averaging method. Average densities were obtained directly from simulation, and the molar enthalpy of vaporization for CO₂ was determined from the following formula:

$$\begin{aligned}\Delta H_{\text{vap}} &= H_g - H_{\text{liq}} \\ &= U_g - U_{\text{liq}} + P(V_g - V_{\text{liq}})\end{aligned}\quad (4.22)$$

Note that, at the state points studied, the ideal gas approximation is insufficiently accurate, and thus simulations were run for both the gas and liquid phases at experimentally-determined densities and pressures.⁵

2nd Virial Calculations

Classical second virial coefficients were calculated for NH₃, H₂O, CO₂, and CHCl₃ using rigid monomer geometries and following the procedure described in Ref. 117.

4.5 Results and Discussion

4.5.1 Overview

We now turn to a discussion of the methods whereby we can compare our newly developed anisotropic force field methodology to various sum-of-spheres models. As is standard in ab initio force field development, we use a straightforward metric to evaluate force field quality: the accuracy with which a given force field functional form can reproduce high-quality ab initio benchmark energies. Furthermore, and because the functional forms introduced in Section 4.3 directly affect only the pairwise-additive portion of the intermolecular potential, we concentrate our

efforts on assessing force field with respect to benchmark calculations of dimer interaction energies, which directly measure the two-body portion of a system’s total intermolecular interaction energy. (When required, and as discussed in Section 4.5.5, many-body effects can be accounted for separately and systematically using known methods).^{2,235} In addition to this primary metric for force field quality, we also evaluate our force fields for their ability to reproduce select experimental properties. Importantly, however, experimental predictions from an ab initio force field significantly depend, not only on the fit quality of the pair potential, but also on the choice of benchmark electronic structure theory, treatment of many-body and/or quantum effects, etc. Because these factors complicate comparisons to experiment, here we treat experimental accuracy as an important, but secondary, metric for evaluating force field accuracy.

So as to systematically evaluate the effects of anisotropy on the development of intermolecular potentials, we compare three types of models. The first model, which we call Iso-Iso FF, uses a completely isotropic description of all energy components. Our second model, Aniso-Iso FF, accounts for long-range electrostatic anisotropy by including multipolar contributions (up to quadrupoles), but uses an isotropic model for all other terms in the intermolecular force field. Note that this model is virtually identical to the Slater-ISA FF model developed in Chapter 3, and that this manner of partially treating anisotropy is very similar in spirit to the popular AMOEBA^{197,199} methodology. Finally, we develop Aniso-Aniso FF, which selectively incorporates anisotropy into all energy components (aside from long-range polarization) of the intermolecular potential. This model, which we also refer to with the moniker MASTIFF (a **M**ultipolar, **A**nisotropic, **S**later-**T**ype **I**ntermolecular **F**orce **F**ield), treats all electrostatic interactions via a multipole expansion with up to quadrupolar contributions, and includes anisotropic parameters for other terms of the force field (short-range interactions plus dispersion) for heteroatoms, atoms in multiple bonding environments, and associated hydrogens. A complete list of anisotropic atom types is given in the Section 4.B.

4.5.2 Accuracy: Comparison with DFT-SAPT

For each of the 91 dimer combinations described in Section 4.4, parameters were fit to reproduce Symmetry-Adapted Perturbation Theory (SAPT) energies calculated for 1000 different relative orientations of the constituent monomers. From these ‘dimer-specific’ fits, and as in described in Chapter 3,¹⁰⁴ we then averaged the root-mean-squared (RMSE) and mean signed errors ($\|\text{MSE}\|$) from each of the 91 fits to produce so-called ‘characteristic RMSE/ $\|\text{MSE}\|$ ’, metrics representative of the errors associated with a given force field methodology. Typically, because the absolute magnitudes of the various energy components become large in the repulsive portion of the potential, these characteristic errors are dominated by repulsive configurations. As such, we have also calculated ‘attractive RMSE/ $\|\text{MSE}\|$ ’ ($a\text{RMSE}/a\|\text{MSE}\|$), defined as the characteristic errors for the subset of configurations with net attractive total interaction energies. All computed characteristic RMSE are shown in Fig. 4.2. Unless otherwise stated, results in this section refer exclusively to the ‘Dimer-specific’ fits in Fig. 4.2, with an explanation and full discussion of so-called ‘Transferable’ fits given in Section 4.5.3.

Based on the characteristic RMSE shown in Fig. 4.2, both Aniso-Iso FF and MASTIFF offer substantial improvements over the completely isotropic model Iso-Iso FF. Though unsurprising, given the well-studied importance of higher-order electrostatic multipole moments, Aniso-Iso FF shows reduced RMSE/aRMSE that are (depending on the exact error metric used) roughly 30% smaller than Iso-Iso FF. Both RMSE and aRMSE measures showing similar gains in accuracy, indicating that inclusion of higher-order multipoles (henceforth ‘multipolar electrostatic anisotropy’) is important in both attractive and repulsive regions of the potential. Crucially, inclusion of additional ‘short-range anisotropies’ (anisotropic interactions arising from overlap of monomer electron densities, namely exchange-repulsion and electrostatic/inductive charge penetration) and long-range ‘dispersion anisotropy’ yields a *further* 40% reduction in RMSE/aRMSE for MASTIFF as compared to the Aniso-Iso FF. This latter result is highly important, as it suggests that, for the generation of highly accurate ab initio potentials, the combination of short-range and

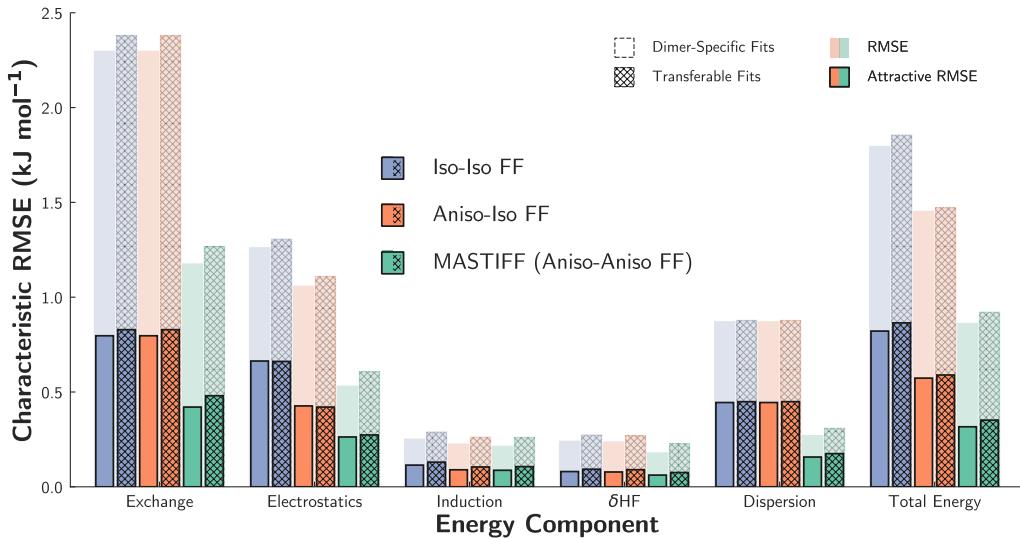


Figure 4.2: Characteristic RMSE (as described in the main text) for the Iso-Iso FF (purple), Aniso-Iso FF (orange), and MASTIFF (green) over the 91 dimer test set. The semi-transparent bars represent total RMSE for each energy component, while the smaller solid bars represent ‘Attractive’ RMSE, in which repulsive points have been excluded. For each force field, two types of fits, dimer-specific (solid) and transferable (hashed lines), are displayed; see Section 4.5.3 for details. Finally, note that, for Iso-Iso FF and Aniso-Iso FF, only the electrostatic and total energy RMSE’s differ.

dispersion anisotropies are just as important to include as multipolar electrostatic anisotropy. Indeed, this substantial increase in force field accuracy, which arises from a full treatment of anisotropic effects, and is independent of improvements from multipolar electrostatic anisotropy, is one of the most important findings in the present Chapter. In summary, and encouragingly, the combination of multipolar electrostatic, short-range, and dispersion anisotropies result in an overall 60% reduction in RMSE/aRMSE when comparing Iso-Iso FF to MASTIFF.

To see exactly how an inclusion of anisotropy impacts each component of the potential, Fig. 4.2 also displays characteristic RMSE/aRMSE for each term in the force field description as compared to DFT-SAPT. Immediately, one can see that (aside from induction, discussed below), an inclusion of atomic-level anisotropy

greatly improves the description of each energy component. Unless otherwise stated, here we report results for aRMSE and dimer-specific fits, though similar values are obtained for overall RMSE and for transferable fits. Compared to Iso-Iso FF, exchange errors in MASTIFF are reduced by 47%. Electrostatic errors are reduced by an even larger 60%. By evaluating the ratio of electrostatic errors between different models, we find that $aRMSE_{\text{Aniso-Iso FF}}/aRMSE_{\text{Iso-Iso FF}} = 0.64$ and $aRMSE_{\text{MASTIFF}}/aRMSE_{\text{Aniso-Iso FF}} = 0.62$, suggesting that *both* higher-order multipoles and anisotropic charge penetration terms are necessarily to obtain an accurate description of the DFT-SAPT electrostatic energy. Finally, via an inclusion of dispersion anisotropy, aRMSE for dispersion are reduced by a significant 65%.

Though the trends for exchange, electrostatics, and dispersion universally suggest the importance of including atomic-level anisotropy, trends for terms describing the physics of polarization and charge-transfer (represented in DFT-SAPT by induction and δHF) are less encouraging. On the one hand, including higher-order multipoles substantially lowers RMSE for induction, with $RMSE_{\text{Aniso-Iso FF}}/RMSE_{\text{Iso-Iso FF}} = 0.70$. Because both Iso-Iso FF and Aniso-Iso FF use isotropic polarizabilities, and because the induction energy fundamentally depends only on the polarizabilities and the static electric field, this result is clearly due to an improved treatment of the static electric field via anisotropy of the multipolar electrostatics. Once again, this suggests that an anisotropic treatment of long-range electrostatics is crucial for accurate force field development. On the other hand, our functional form for anisotropic short-range induction (Eq. (4.15) and Eq. (4.17)) leads to no improvement in the induction RMSE, with $RMSE_{\text{Aniso-Iso FF}}/RMSE_{\text{Iso-Iso FF}} = 0.97$. This observed lack of improvement is likely due to a combination of factors. First, and perhaps most importantly, we have chosen in this Chapter to use isotropically-averaged dipole polarizabilities, but as with electrostatics, anisotropy and higher-order terms have been shown to be important in the multipole expansion of atomic dipole polarizabilities.^{51,203,211,224,236} Second, and though probably a smaller source of error, it is also unclear how to optimally model the distance dependence of the induction energy at short intermolecular separations, where penetration and charge-transfer effects become important and the long-range polarization terms must be

damped.^{104,119,237,238} Given that the more elaborate short-range form of the MASTIFF induction model does not result in a tangible improvement, it is quite possible that alternative formulations are required for an accurate treatment of highly anisotropic induction.

To further analyze the effects of anisotropy on a molecule-by-molecule basis, we have calculated ‘improvement ratios’, defined as $aRMSE_{\text{Iso-Iso FF}}/aRMSE_{\text{MASTIFF}}$, for each energy component and for each homomonomeric species in the test set, results for which are shown in Table 4.1. (Improvement ratios for heteromonomeric species are given in the Supporting Information of Ref. 239, and we additionally provide scatter plots of each homomonomeric force field fit in Section 4.C.)

The most striking observation from the data presented in Table 4.1 is that the improvement ratios vary considerably with molecule. For example, with water the $aRMSE$ is improved by an order of magnitude when anisotropy is included. On the other hand, no improvement is seen for hydrocarbons such as ethane and methane (also see the Section 4.C). Consequently, anisotropy in the short-range expansions may be necessary for only some atoms types (see Section 4.6). For the molecules studied in our test set, and in line with chemical intuition, we have found anisotropy to be particularly important for heteroatoms, π -bonded atoms, and all hydrogens bonded to anisotropic heavy atoms. Appealingly, this distinction between anisotropic and isotropic atom types simplifies force field parameterization and can enable more efficient molecular simulation (via a more cost-effective treatment of multipolar electrostatics) without sacrificing force field accuracy. Note that the current empirically-determined definitions of anisotropic atom types match both chemical intuition and the more quantitative measures of atomic anisotropy proposed by other groups.^{173,186}

In general, the ordering of improvement ratios for exchange, electrostatics, dispersion, and the total energies are reasonably correlated. (As stated above, our model for anisotropic induction interactions is rather poor, and hence the improvement ratios for induction and δHF are relatively uncorrelated with the other components. A good model for anisotropic induction and δHF might easily change this result). Physically speaking, all atomically-anisotropic interactions

| | Exchange | Electrostatics | Induction | δHF | Dispersion | Total Energy |
|------------------------|-------------|----------------|-------------|-------------------|-------------|--------------|
| H ₂ O (O,H) | 4.96 | 13.12 | 1.69 | 1.88 | 8.20 | 11.54 |
| CO ₂ (C,O) | 3.83 | 9.13 | 0.99 | 0.64 | 4.91 | 8.62 |
| NH ₃ (N,H) | 3.15 | 5.36 | 0.90 | 2.86 | 2.45 | 5.78 |
| Ethene (C,H) | 1.44 | 1.46 | 1.00 | 1.00 | 7.59 | 4.16 |
| Chloromethane (Cl) | 3.17 | 4.03 | 1.36 | 1.04 | 4.20 | 4.08 |
| Methyl Amine (N,H) | 1.70 | 2.93 | 1.05 | 2.22 | 2.95 | 2.37 |
| Methanol (O,H) | 1.81 | 3.05 | 1.11 | 2.03 | 1.00 | 2.36 |
| Dimethyl Ether (O) | 1.30 | 2.07 | 1.38 | 1.19 | 1.85 | 2.30 |
| Ethanol (O,H) | 1.29 | 3.10 | 1.04 | 1.45 | 1.79 | 2.14 |
| Acetone (O) | 1.58 | 1.98 | 1.03 | 1.34 | 1.51 | 1.08 |
| Ethane () | 1.00 | 1.26 | 1.05 | 1.01 | 1.00 | 1.08 |
| Ar () | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Methane () | 1.00 | 0.93 | 0.99 | 1.01 | 1.00 | 0.94 |

Table 4.1: ‘Improvement Ratios’ for each homomonomeric species in the 91 dimer test set. For each dimer and energy component, the improvement ratio is calculated as the ratio of aRMSE between Iso-Iso FF and MASTIFF; values greater than 1 indicate decreased errors in the anisotropic model. Entries have been ordered according to the improvement ratio for the total energy.

arise from the same source (atomically-anisotropic electron densities), and so the observed correlation might have been expected. Nevertheless, there are some exceptions to this trend, such as with ethene and acetone. For ethene, relatively modest improvement ratios (roughly 1.4) are seen for exchange and electrostatics, whereas dispersion shows a much greater improvement ratio of 7.6. Since ethene homomonomeric interactions are dispersion-dominated, the improvement ratio for the total energy then roughly corresponds to that of dispersion. A larger test set (particularly one which includes more non-polar aromatic species) would be necessary to assess the generality of this result. For acetone, there is good correlation between the improvement ratios for exchange, electrostatics, and dispersion, which might lead one to suspect that the total energy improvement ratio would also be around 1.5-2.0. Nevertheless, for this molecule, the isotropic model benefits from error cancellation between energy components, and the total energy aRMSE between isotropic and anisotropic models are rather similar.

Crucially, electrostatics is most definitely not the only intermolecular interaction for which atomic-level anisotropy improves model quality. Indeed, for molecules like ethene, multipolar anisotropy in the electrostatic model is relatively unim-

portant, whereas dispersion anisotropy is essential for accurately modeling the π interactions. Thus, for a given system, multipolar electrostatic, dispersion, and/or short-range anisotropies may all be important, and all relevant anisotropies must be accounted for in order to obtain good intermolecular models.

4.5.3 Transferability: Comparison to DFT-SAPT

From the above results it is clear that, when explicitly parameterized, an inclusion of anisotropy can greatly enhance the accuracy of an intermolecular potential. Nevertheless, for standard force field development, force field parameters must be *transferable* in order to be useful in the accurate prediction of intermolecular interactions in new chemical and/or physical environments. Indeed, in comparing simpler models to ones that introduce additional complexity, there is an ever-present danger that any accuracy gains from the more complex functional form are simply due to over-fitting or error cancellation,²⁴⁰ ultimately resulting in an overly-complex model with poor predictive ability and limited transferability.

We have previously shown how, with models similar to Iso-Iso FF^{80,117} or Aniso-Iso FF,¹⁰⁴ it is possible to generate transferable potentials with applicability to a broad range of chemical and physical environments.⁸⁰ This transferability has been attributed to a combination of the physically-meaningful energy decomposition of DFT-SAPT, our choice to parameterize on a component-by-component basis (rather than to the total energy), our use of physically-motivated functional forms, and our recourse to parameters calculated on the basis of monomer properties.^{80,104,117}

MASTIFF largely shares this philosophy of force field development, and so we might also expect it to be transferable to heteromeric dimers. However, this transferability cannot be taken for granted because of the specific way in which we have included the anisotropy. First, we have relied on several separability ansatzes (Eq. (4.4) and Eq. (4.7)), and second, in doing so we have implicitly neglected potentially important interaction functions that depend on the relative orientation between monomers. Both of these assumptions may affect the transferability of the resulting force field.

To assess the transferability of the MASTIFF model, we analyze the extent to which parameters developed for the homomonomeric systems can be used, without modification, to describe the interactions of the mixed dimers. Such an out-of-sample prediction, which is easily accomplished with our test set, is a direct measure of the extent to which our pair potentials can be applied to new chemical environments. For these transferable fits, parameters were fit to the 13 homomonomeric systems, and the combination rules shown in Eq. (4.15) were used to generate force fields for the remaining heteromonomeric systems. Thus, with these transferable fits we have essentially generated 78,000 predictions from fits to 13,000 data points. RMSE and aRMSE for these fits are shown in Fig. 4.2, and we treat relative differences between these quantities for the ‘dimer-specific’ and ‘transferable’ fits as a measure of the extent of transferability for each force field methodology.

Remarkably, all three force fields — Iso-Iso, Aniso-Iso, and MASTIFF — perform similarly for the dimer-specific and transferable fits, both for the individual interaction energy components and for the total interaction energy. The degree of transferability of the MASTIFF model is very encouraging, and indicates that the manner in which we have chosen to include the anisotropy is meaningful and does not lead to overfitting, but rather increases the accuracy of the intermolecular potentials for both in-sample and out-of-sample systems.

4.5.4 Comparison to Experiment: Second Virial Coefficients

In addition to comparisons with DFT-SAPT, we have also benchmarked our force fields against experimental second virial coefficients, which offer a direct experimental measure of the pair potential without the complication of many-body effects. Still, such comparisons to experiment depend, not only on the quality of the force field, but also on the accuracy of the benchmark electronic structure theory used to fit the force field. As compared to gold-standard CCSD(T)/CBS calculations, small ($< 1 \text{ kJ mol}^{-1}$) but systematic inaccuracies can be present in DFT-SAPT/aVTZ+m¹⁰⁴ calculations, and so in this section we refit our potentials

to a CCSD(T)-F12a/aVTZ+m benchmark, which serves as a computationally affordable yet accurate prediction of the CCSD(T)/CBS limit.^{13,241} We refer to these coupled cluster-based models with a -CC suffix, e.g. MASTIFF-CC, and details of the refitting procedure (which minimally affect the dispersion energies) can be found earlier in Section 4.4. Thus, aside from quantum effects (which are negligible for CO₂²⁴² and well-benchmarked for H₂O²⁴³), our second virial predictions should offer a fairly clean comparison between different models and experiment.

Using the -CC potentials, we have calculated second virial coefficients for each Iso-Iso FF-CC, Aniso-Iso FF-CC, and MASTIFF-CC and for the following systems: H₂O (Fig. 4.3), NH₃ (Fig. 4.4), CHCl₃ (Fig. 4.5), and CO₂ (Fig. 4.6). First, we find that the MASTIFF-CC methodology predicts virial coefficients which closely corresponds to experimental data. Given the range of systems tested (CO₂ dimer interactions are dispersion dominated, while CHCl₃, NH₃, and H₂O have relatively larger electrostatic and polarization contributions), this suggests that, when benchmarked against high-quality electronic structure theory, our anisotropic methodology offers a general strategy for quantitatively accurate pairwise potential development. Second, we note that the Iso-Iso FF-CC predictions are much worse than their MASTIFF-CC or Aniso-Iso FF counterparts, suggesting that an accurate treatment of long-range electrostatics is essential to obtain accurate virial coefficients. Finally, though Aniso-Iso FF-CC gives equally good predictions for some systems (notably CHCl₃) compared to the MASTIFF-CC method, virial coefficients for other systems (especially H₂O) are less accurate, suggesting that dispersion and short-range anisotropies are also important in many systems for the accurate prediction of virial coefficients. Consequently, and in summary, the minimal additional computational overhead (compared to Aniso-Iso FF-CC) and excellent accuracy of MASTIFF-CC permits us to recommend this fully anisotropic MASTIFF methodology for the prediction of dimer interaction energies and second virial coefficients.

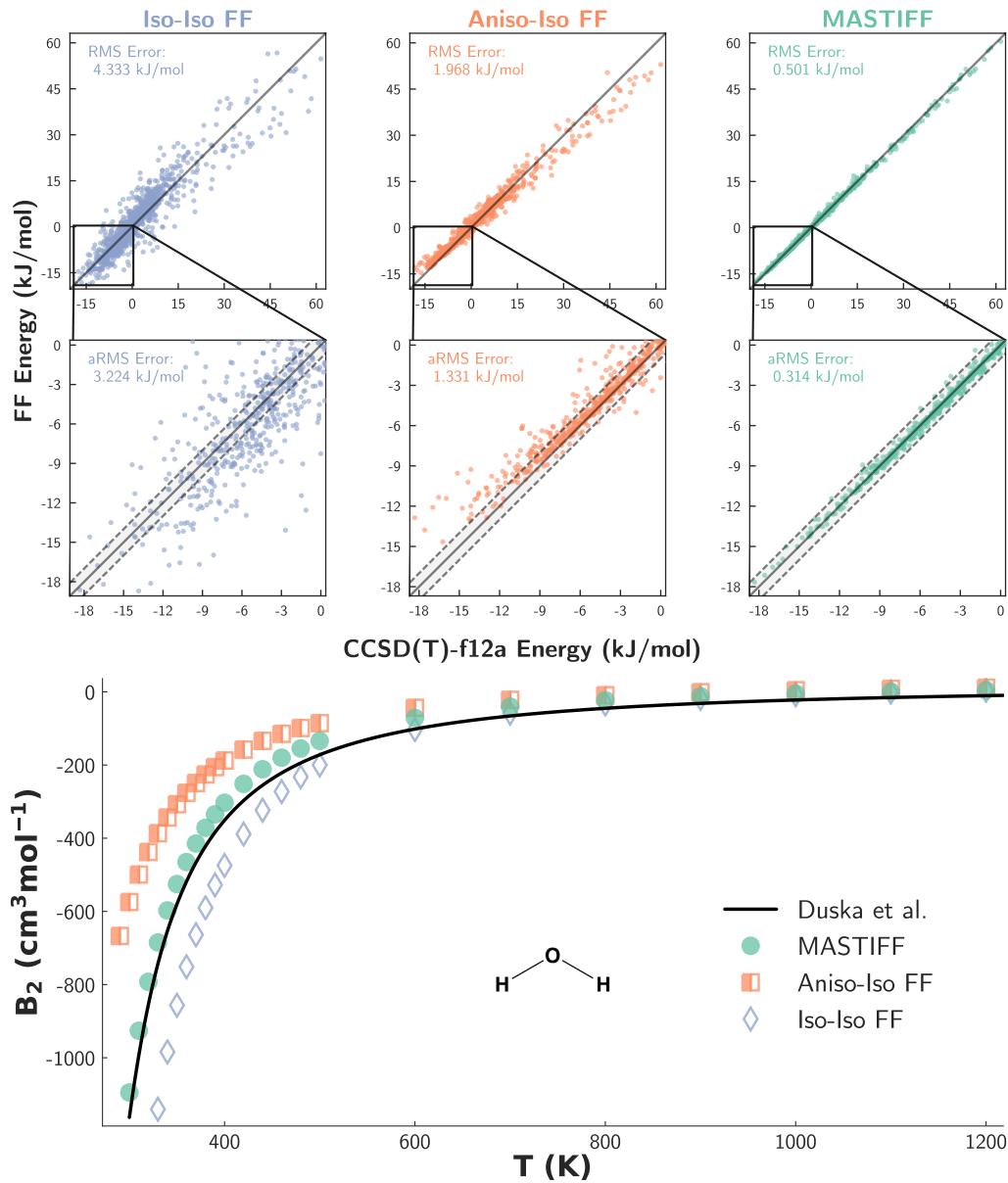


Figure 4.3: Classical second virial for water. Experimental data from Ref. 9. Note that some data points from Iso-Iso FF extend below the plot area.

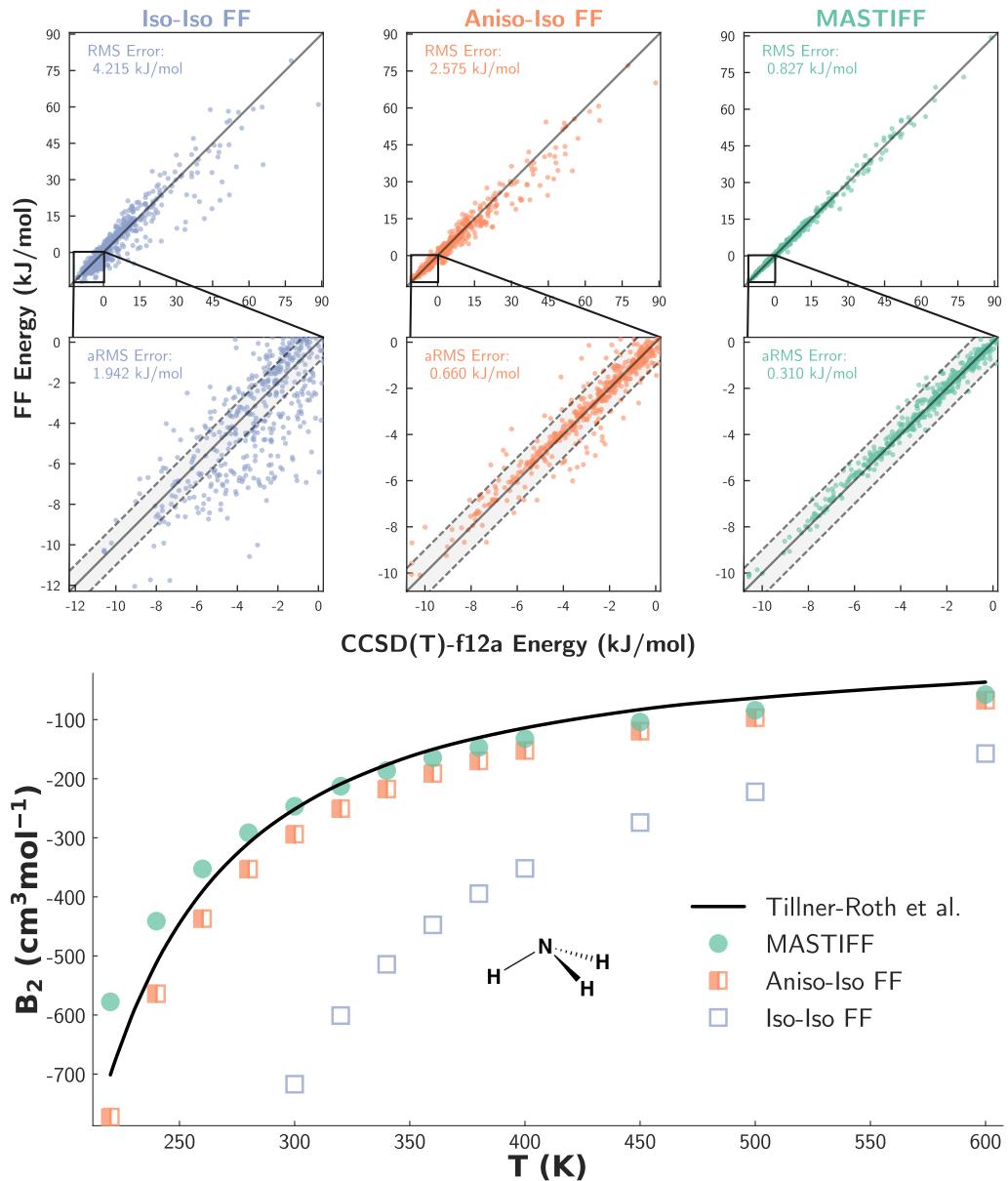


Figure 4.4: Classical second virial for ammonia. Experimental data from Ref. 10. Note that some data points from Iso-Iso FF extend below the plot area.

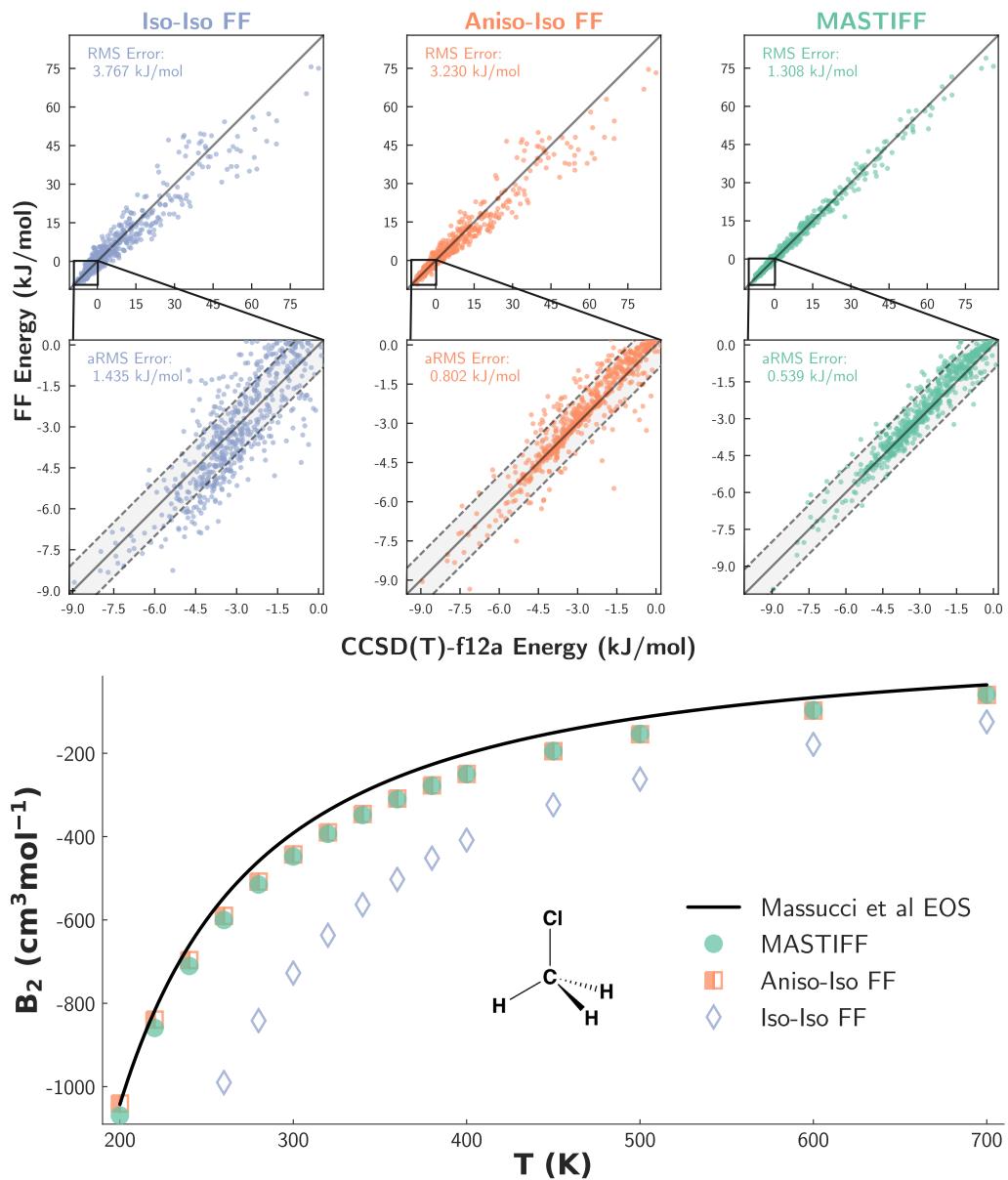


Figure 4.5: Classical second virial for chloromethane. Experimental equation of state (EOS) from Ref. 11. Note that some data points from Iso-Iso FF extend below the plot area.

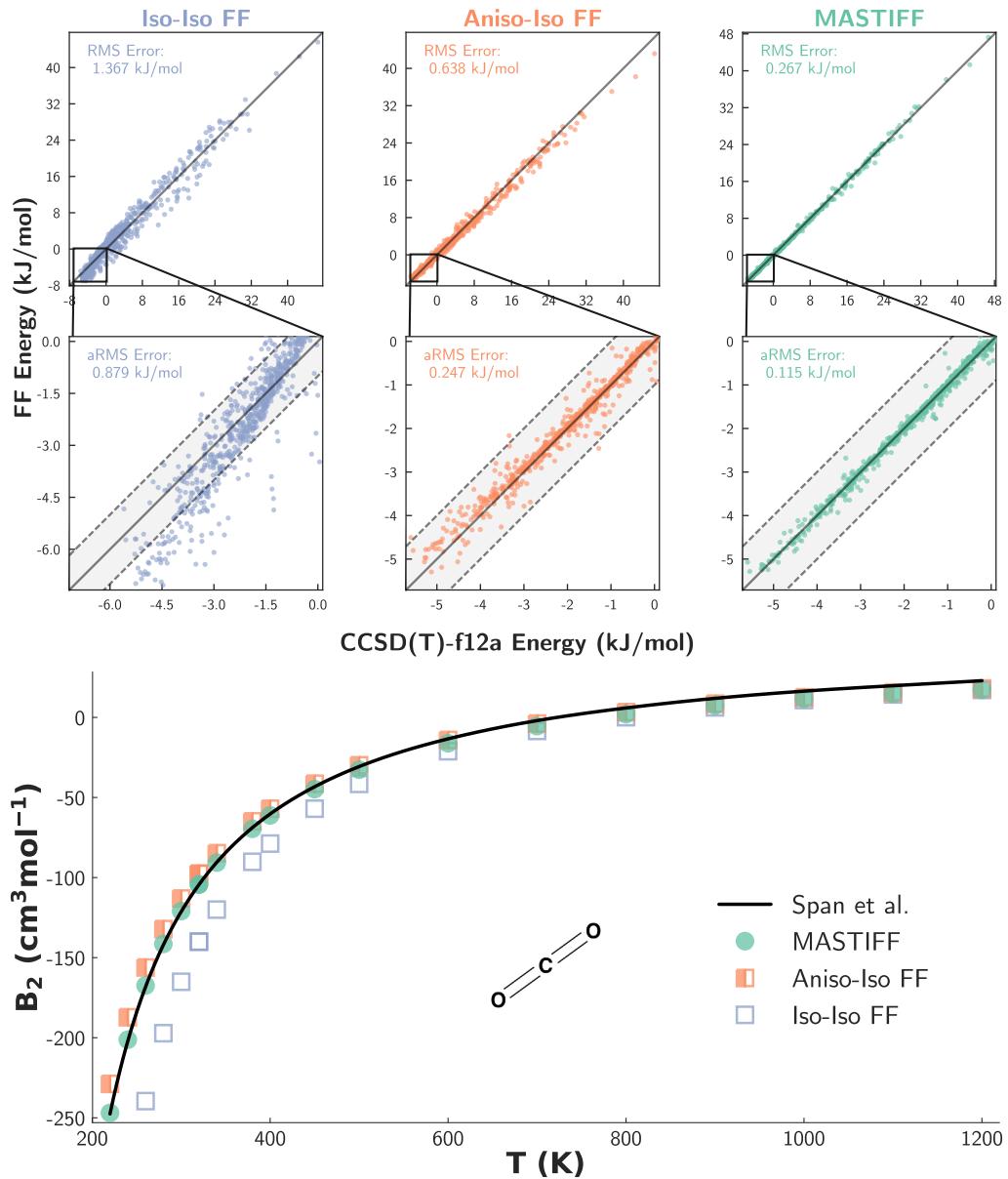


Figure 4.6: Classical second virial for CO₂. Experimental data from Ref. 5

4.5.5 Comparison to Experiment: Condensed Phase Properties of CO₂

To demonstrate the applicability of the MASTIFF methodology in condensed phase simulation, we have developed a complete many-body potential for CO₂, and have run bulk simulations involving a variety of vapor, liquid, supercritical, and solid phase points for preliminary comparisons to experiment. As above, we use the MASTIFF-CC potential to describe the pairwise potential and the many-body induction. As for other many-body effects, it is well-known^{12,229,235} that three-body dispersion, and to a lesser extent, three-body exchange, are also important.²⁴⁴ Thus, we model three-body dispersion via a modified version of the three-body dispersion potential developed by Oakley and Wheatley (see Section 4.4). Three-body exchange effects are not accounted for in our model, however prior work shows they are very small under the conditions studied here.²³⁵

Density predictions for the vapor, liquid, and supercritical phases of CO₂ are shown in Table 4.2, and enthalpies of sublimation and vaporization are shown in Table 4.3. We find it notable and highly encouraging that MASTIFF-CC reproduces *all* studied experimental properties to within a few percent. Of particular note is our excellent reproduction of the sublimation enthalpy, which critically depends on the lattice energy of the solid phase. Unlike with liquid or supercritical CO₂, where many dimer configurations are sampled, the solid consists of only four symmetry-unique configurations. Consequently, whereas an isotropic potential (which is in error for particular dimer configurations, but can take advantage of error cancellation to be accurate in an average sense) might yield good property predictions for the liquid phase, it would not be expected to correctly predict the solid phase (where beneficial error cancellation is unlikely). Indeed, most theories (including our previously developed SYM-3B model,²³⁵ nearly all popular empirically-developed CO₂ models,²⁴⁵ AMOEBA,²³³ and many electronic structure theories²³³) struggle to correctly predict the solid phase properties of CO₂! For this reason, the enthalpy of sublimation is an extremely stringent test of force field quality,²⁴⁵ and our accurate reproduction of this quantity is evidence for both the

excellent quality of the MASTIFF-CC potential in specific and of the importance of atomic-level anisotropy in general. Though more testing is needed to confirm the accuracy of our anisotropic force field for other phase points, our results suggest that, crucially, the MASTIFF-CC potential is transferable across the entire phase diagram of molecular CO₂, and is capable of describing the gas, liquid, supercritical, and solid phases.

Despite the excellent success of the MASTIFF-CC model, it is also worthwhile to address and understand its minor shortcomings. In particular, we have studied representative two- and three-body energies taken from a snapshot of the liquid at 273.15 K and 100 bar. For the two-body energies, we have compared against the extremely accurate Kalugina et al.¹³ potential, while for three-body energies we have benchmarked against the Hellmann¹² PES. From these results (Section 4.D), it is clear that our pairwise MASTIFF potential is highly accurate for all configurations present in the liquid, with very small RMSE and no systematic error in the potential, such that the total two-body energy is accurate to within 0.05% compared to the Kalugina et al. PES. Once again, this result argues strongly for the accuracy and transferability of the MASTIFF methodology, and suggests that an inclusion of anisotropy is essential, not only for gas-phase clusters, but also for simulations of the bulk. By contrast, our three-body potential is systematically in error compared to the Hellmann PES. Though some of this error may be due to inaccuracies in the benchmark potential itself as compared to coupled-cluster,¹² most of this error is likely due to inaccuracies in our model for many-body CO₂ interactions. The atomically-isotropic treatment of three-body dispersion, neglect of higher-order dispersion terms , and neglect of explicit three-body exchange may all contribute to this error, and an improved model for many-body CO₂ interactions will be the subject of future research. Indeed, it is well-known that the density can be extremely sensitive to the treatment of many-body effects,²⁴⁴ and it is highly probable that an improved many-body model would reduce the already small errors observed in our MASTIFF-CC predictions. Regardless, for now we conclude that, despite some small residual errors arising from the simplified treatment of many-body effects, our MASTIFF-CC methodology yields for an extremely accurate force field for CO₂

| Phase | T (K) | P (bar) | Density (g/ml) | Exp. | % Error |
|---------------|--------|---------|----------------|-------|---------|
| Gas | 300 | 50 | 0.131 | 0.128 | 2.34 |
| Supercritical | 320 | 140 | 0.728 | 0.703 | 3.56 |
| Liquid | 300 | 100 | 0.825 | 0.802 | 2.87 |
| Liquid | 273.15 | 100 | 1.000 | 0.974 | 2.67 |

Table 4.2: Select densities for CO₂ across a range of experimental conditions. Experimental data taken from the EOS of Ref. 5. Entries ordered by increasing experimental density.

| Phases | T (K) | ΔH (kJ mol ⁻¹) | Exp. | % Error |
|--------|--------|------------------------------------|------|---------|
| s → g | 194.76 | 25.0 ± 0.15 | 25.2 | -0.8 |
| l → g | 288 | 7.92 | 7.80 | -1.4 |

Table 4.3: Enthalpies of vaporization/sublimation for CO₂ at several temperatures. Experimental data taken from the EOS of Ref. 5. The uncertainty in the enthalpy of sublimation is due to ambiguity in the theoretical zero-point energy for CO₂ (see Section 4.4).

with applicability across a range of experimentally-important phases.

4.6 Conclusions and Recommendations

We have developed a comprehensive methodology for modeling atomic-level anisotropy in standard intermolecular force fields. By treating this anisotropy through a simple extension of standard isotropic force fields,¹⁰⁴ we have successfully demonstrated how this computationally-efficient treatment of atomic-level anisotropy leads to significant improvements in models for intermolecular interactions. Critically, and in contrast to popular assumption, we have shown how the accurate treatment of multipolar electrostatics does not *by itself* account for all energetically-important effects of atomic-level anisotropy. Rather, our results indicate that anisotropy may need to be included in the each electrostatic, exchange and

dispersion terms in order to obtain intermolecular force fields of the highest quality. In the present study, and in agreement with the more quantitative metrics proposed by others,^{173,186} we have found a comprehensive model of atomic-level anisotropy to be particularly important for obtaining sub- kJ mol^{-1} accuracy for describing molecules with heteroatoms (particularly ones with exposed lone pairs), carbons in multiple bonding environments, and hydrogens bound to anisotropic heavy atoms. Our new intermolecular ‘MASTIFF’ force fields show great promise, not only with respect to high-quality electronic structure benchmark energies, but also with respect to experimental property predictions. Importantly, MASTIFF maintains high efficiency and transferability, and can easily be implemented in common software packages such as OpenMM for use in condensed phase simulations.¹⁵⁷

Despite the advances presented in this Chapter, several aspects of our force field methodology require further improvement, and will be the subject of ongoing research. In particular, an improved description of induction effects will become essential for accurate bulk simulations of highly polarizable molecules such as water. We are now actively working to develop improved models that can describe both long-range anisotropic polarization and short-range polarization damping, as these aspects of the force field critically affect both the two- and many-body induction energies and can account for a sizable fraction of the total interaction energy in condensed phases. We anticipate that these improved models for induction will, in combination with an accurate description of three-body dispersion and exchange, yield a general approach to force field development that captures both the two- and many-body features of intermolecular interactions, in turn enabling highly accurate, ‘next-generation’ force field development capable of simulating a wide array of phases and chemical environments.

4.A Motivation for $g(\theta_i, \phi_i, \theta_j, \phi_j)$

As shown elsewhere,^{53,246} an exact (under the ansatz of radial and angular separability) model for $g(\theta_i, \phi_i, \theta_j, \phi_j)$ is given by Stone’s \bar{S} -functions, which form a complete basis set for describing any scalar function which depends on the relative

orientation between molecules, and are given (following Stone's notation⁴⁸) by the formula

$$\bar{S}_{l_1 l_2 j}^{k_1 k_2} = i^{l_1 - l_2 - j} \begin{pmatrix} l_1 & l_2 & j \\ 0 & 0 & 0 \end{pmatrix}^{-1} \sum_{m_1 m_2 m} [D_{m_1 k_1}^{l_1}(\Omega_1)]^* [D_{m_2 k_2}^{l_2}(\Omega_2)]^* C_{lm}(\theta, \phi) \begin{pmatrix} l_1 & l_2 & j \\ m_1 & m_2 & m \end{pmatrix}. \quad (4.23)$$

The general form of these \bar{S} -functions can be quite complicated, and involve both the Wigner D rotation matrices and Wigner 3j-symbols (quantities in parentheses) as well as the degree (l_1 , l_2 , and j) and order (m_1 , m_2 , and m for the global coordinate system, k_1 and k_2 for the various local coordinate systems) of the spherical harmonic tensors. Here subscripts reference either molecule 1 or molecule 2, and subscriptless quantities refer to the dimer as a whole.

In order to obtain a functional form for the exchange-repulsion that is amenable to simple combination rules (a necessary prerequisite for transferable potentials), we must somehow be able to separate $g(\theta_i, \phi_i, \theta_j, \phi_j)$ into monomer contributions. Unfortunately, many of the \bar{S} -functions depend on the relative orientation of the dimer itself, and thus must be excluded in the development of *transferable* potentials. Thus as a second ansatz (empirically validated by us in Section 4.5 and by others²⁴⁷) we neglect all contributions from \bar{S} -functions that depend on both local coordinate systems. This leaves us with two sets of \bar{S} -functions, namely

$$\bar{S}_{l0l}^{k0} = C_{lk}(\theta_i, \phi_i) \quad (4.24)$$

and

$$\bar{S}_{0ll}^{0k} = C_{lk}(\theta_j, \phi_j) \quad (4.25)$$

which are simply the renormalized spherical harmonics (Eq. (4.8)) expressed in each of the two local coordinate systems.

Given our truncated expressions for the \bar{S} -functions, we now need only extend our functional form for $f(r_{ij})$ to incorporate these anisotropic contributions. We

choose, in a manner analogous to literature precedent,^{51,115,118,178,179,200,203,222,223} to expand the A_i^{exch} and A_j^{exch} parameters of Eq. (4.7) in terms of a truncated expansion of \bar{S} -functions. (In principle, we could also account for anisotropy in the B_{ij} parameters of our model for $f(r_{ij})$. However, previous literature suggests that in practice this ‘hardness’ parameter can often be treated as constant, and we also neglect its possible anisotropy in this Chapter.) Consequently, all short-range anisotropies are modeled in this Chapter by the expressions given in Eq. (4.9) and Eq. (4.10).

4.B Local Axis Definitions

For each molecule in the 91 dimer test set, listed below are any atom types which have been treated anisotropically. For each anisotropic atom type, the approximate symmetry and all terms included in the spherical harmonic expansion are listed to the right of the atom type. Additionally, the local axis reference frame for each anisotropic atom type is defined in the Axes subsection using the z-then-x convention employed by AMOEBA and other potentials. The first column of the axes subsection denotes the index of the anisotropic atom (atom ordering as in Section A.1), and the second column denotes whether the z or x axis is being defined. For certain local symmetries, the choice of x-axis is unimportant, and so not every anisotropic atom type has a defined x-axis. The remaining columns define the direction vector for the axis in terms of atomic indices. The first index (often the anisotropic atom itself) lists the start of the vector, and the endpoint of the vector is defined as the midpoint of all subsequently listed atoms.

To use water as an example, the oxygen atom is treated anisotropically using a spherical harmonic expansion that includes y10, y20, and y22c terms (notation as in Ref. 48). The z-axis points from the oxygen to the midpoint between the two hydrogens, and the xz plane (and subsequently the x-axis) is defined by one of the O–H bonds.

4.B.1 Acetone

OC c2v y10 y20 y22c

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

1 z 1 0

1 x 0 2

4.B.2 Ar

Ar

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

4.B.3 Chloromethane

Chloromethane

Cl c3v y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

1 z 1 0

4.B.4 Carbon Dioxide

CO2

OCO cinfv y10 y20

CCO2 dinfh y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
0 z 0 1
1 z 1 0
2 z 2 0

4.B.5 Dimethyl Ether

Dimethyl Ether

0 c2v y10 y20 y22c

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
0 z 0 1 2
0 x 0 1

4.B.6 Ethane

Ethane**Axes**

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

4.B.7 Ethanol

Ethanol

OH c2v y10 y20 y22c
HO cinfv y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
2 z 2 1 3
2 x 2 1
3 z 3 2

4.B.8 Ethene**Ethene**

CM c2v y22c
HM cinfv y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
0 z 0 1
0 x 0 2
1 z 1 0
1 x 1 4
2 z 2 0
3 z 3 0
4 z 4 1
5 z 5 1

4.B.9 Water

H2O
OH2 c2v y10 y20 y22c
H2O cinfv y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

0 z 0 1 2

0 x 0 1

1 z 1 0

2 z 2 0

4.B.10 Methane

Methane

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

4.B.11 Methanol

Methanol

OH1 c2v y10 y20 y22c

H01 cinfv y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)

1 z 1 0 5

1 x 1 0

5 z 5 1

4.B.12 Methyl Amine

Methyl Amine

N1 c2v y10 y20 y22c
HN1 cinfv y10 y20

Axes

ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
1 z 1 0 5 6
1 x 1 0
5 z 5 1
6 z 6 1

4.B.13 Ammonia

Ammonia

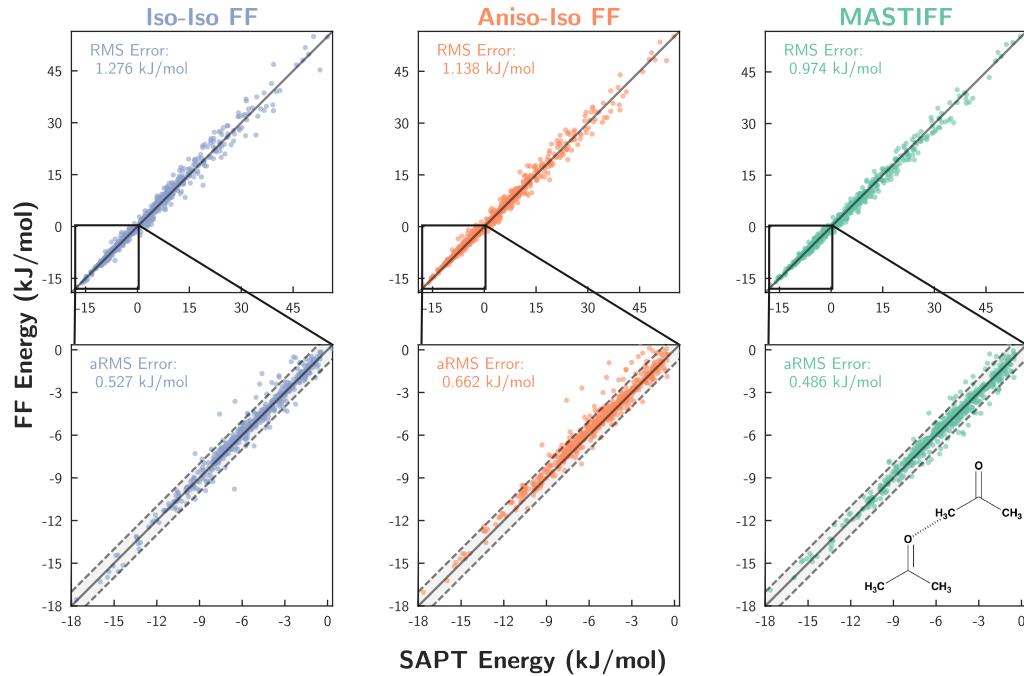
N c3v y10 y20
HN cinfv y10 y20

Axes

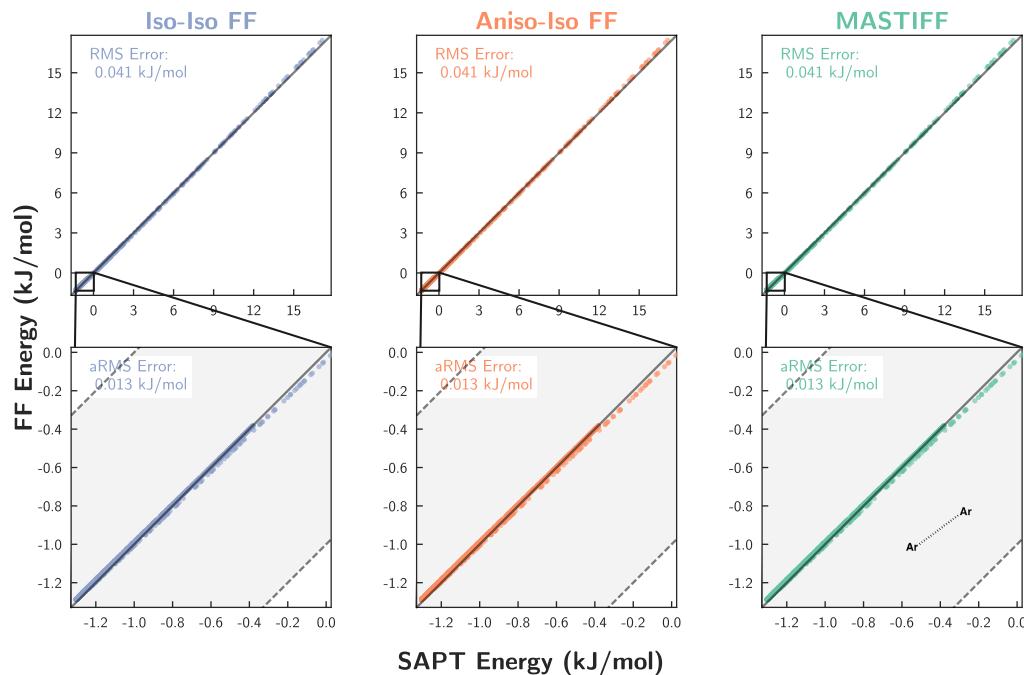
ATOM# AXIS (z or x) Atomic Indices defining vector (either 2 or more integers)
0 z 0 1 2 3
0 x 0 1
1 z 1 0
2 z 2 0
3 z 3 0

4.C Homodimer Fits

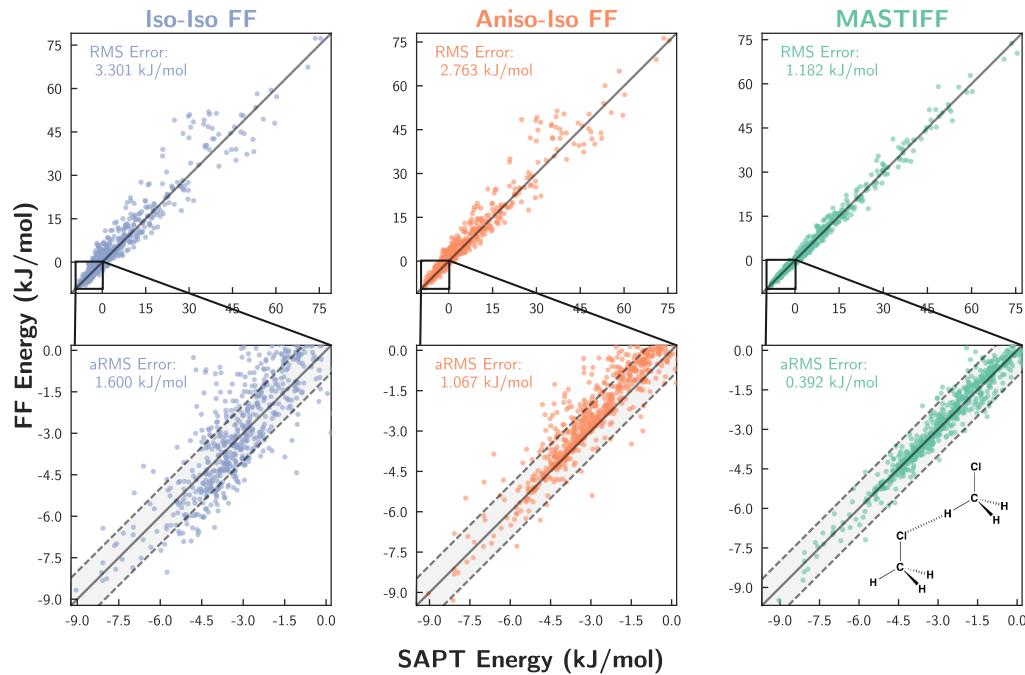
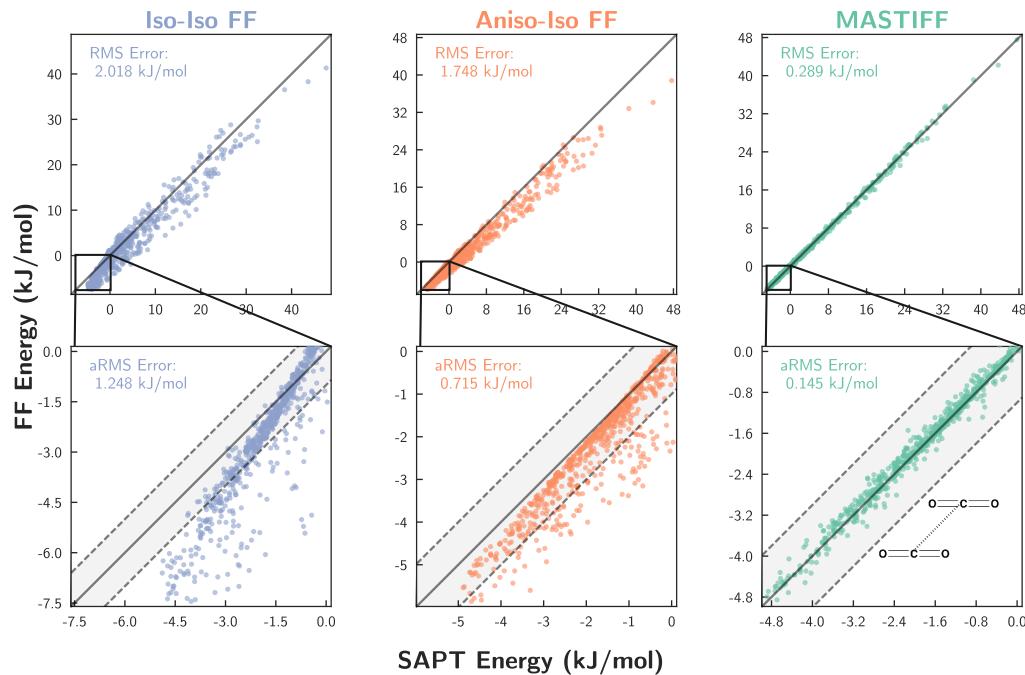
(a) Acetone Dimer



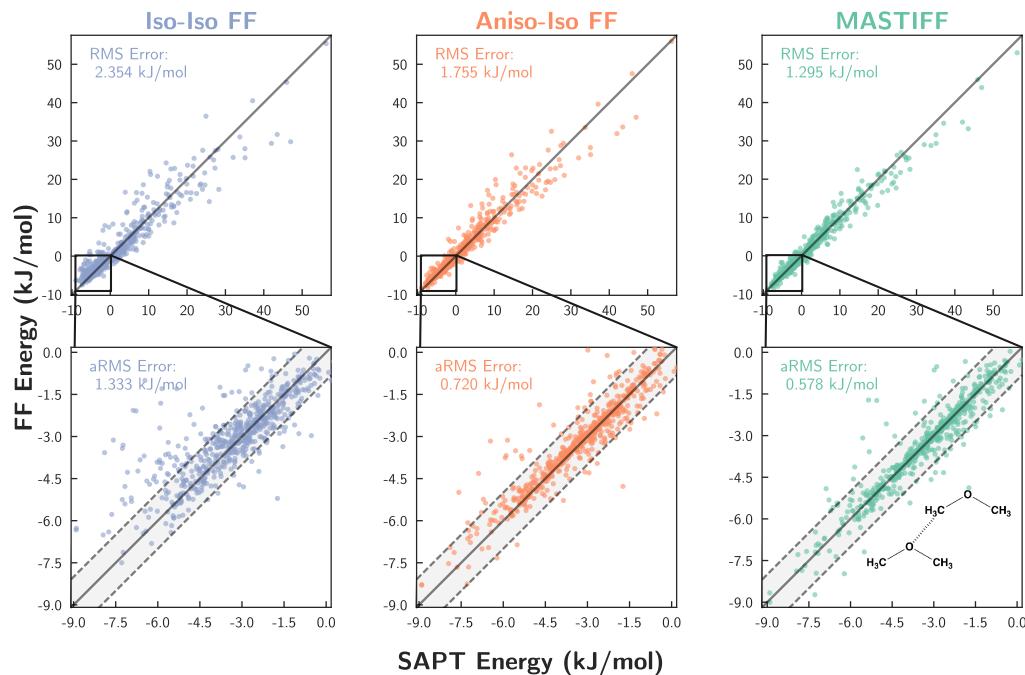
(b) Ar Dimer



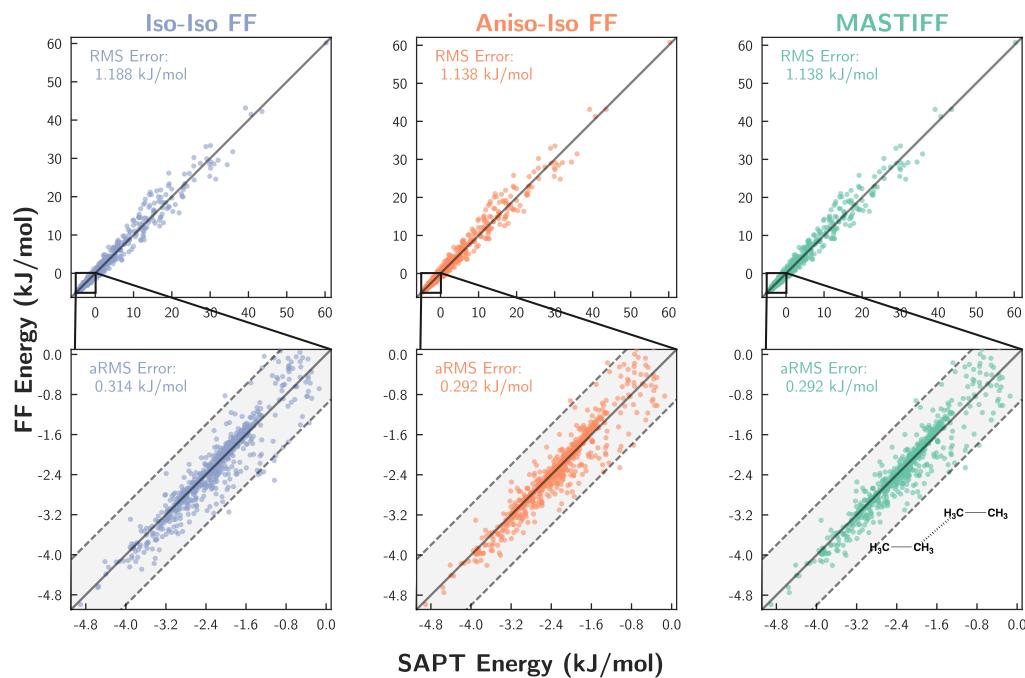
(c) Chloromethane Dimer

(d) CO₂ Dimer

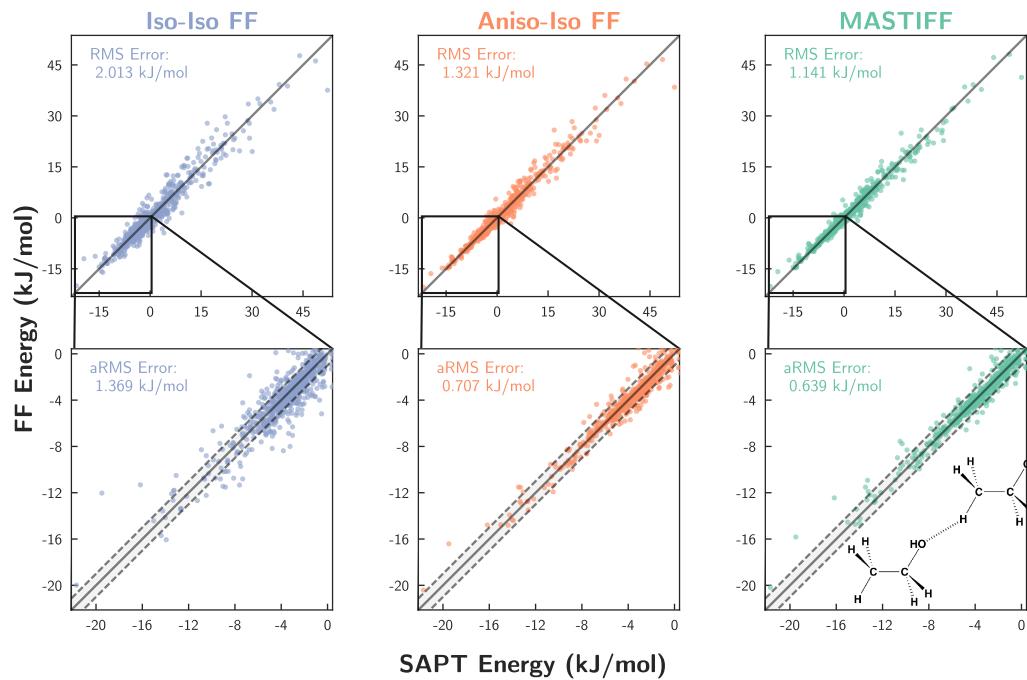
(e) Dimethyl Ether Dimer



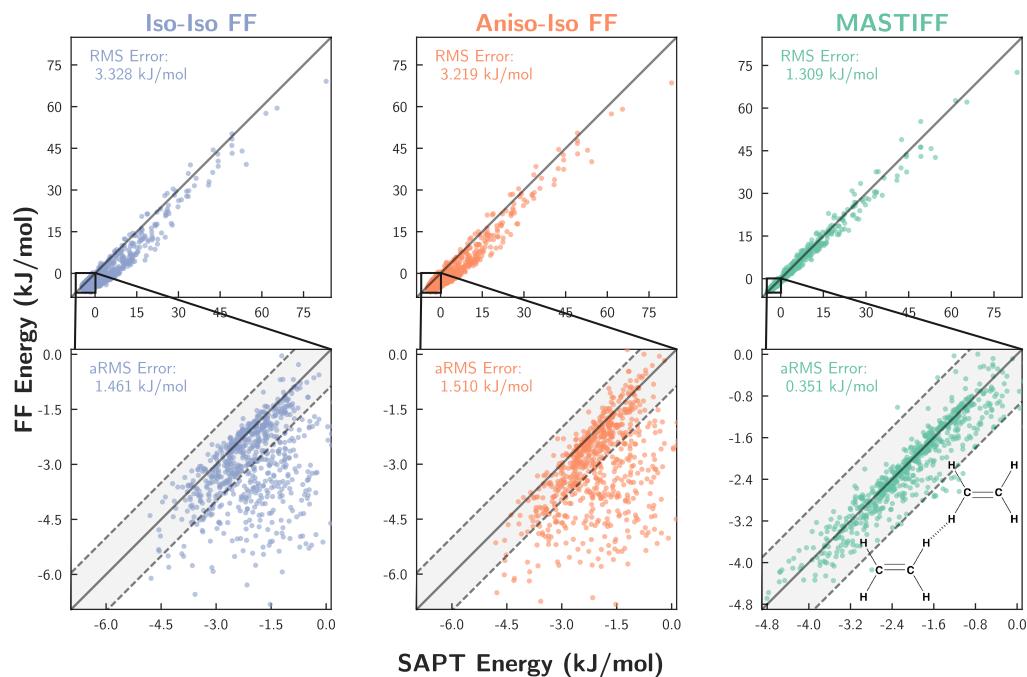
(f) Ethane Dimer

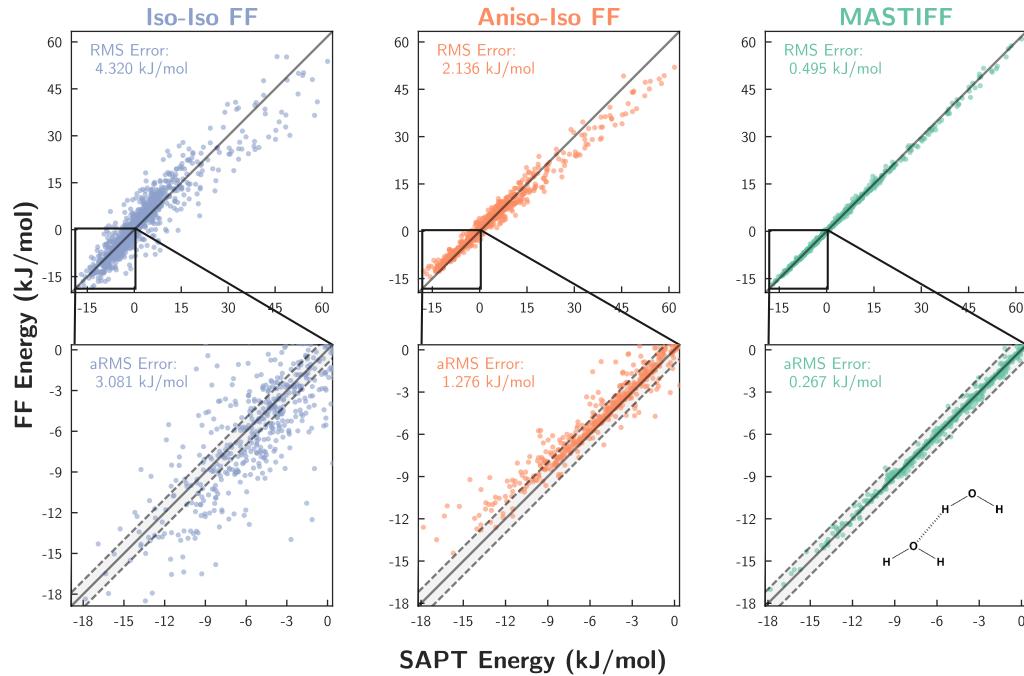


(g) Ethanol Dimer

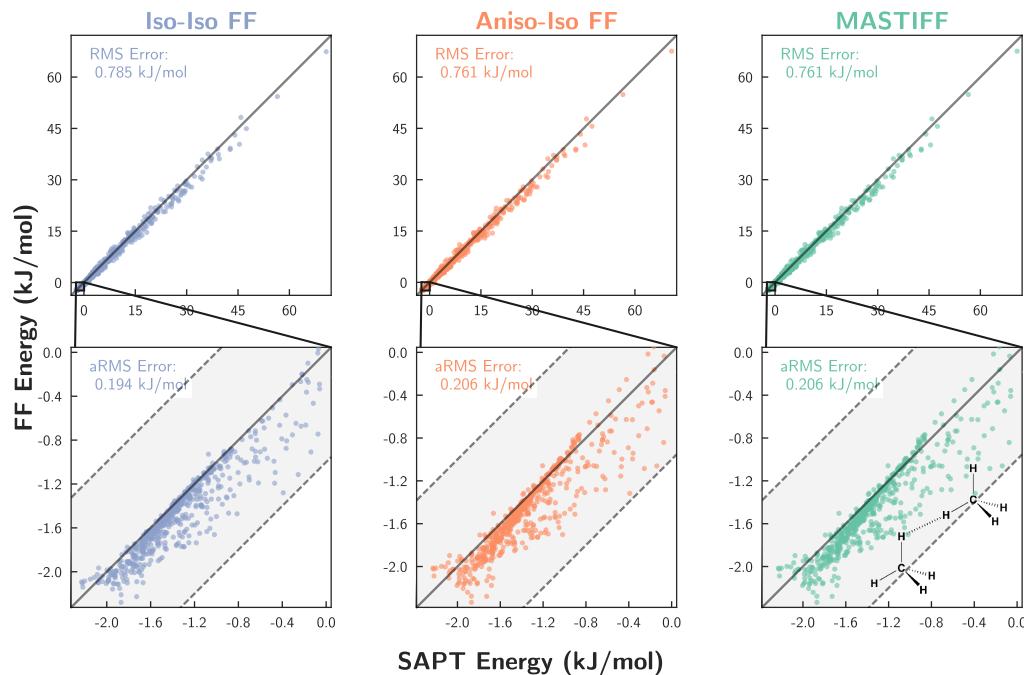


(h) Ethene Dimer

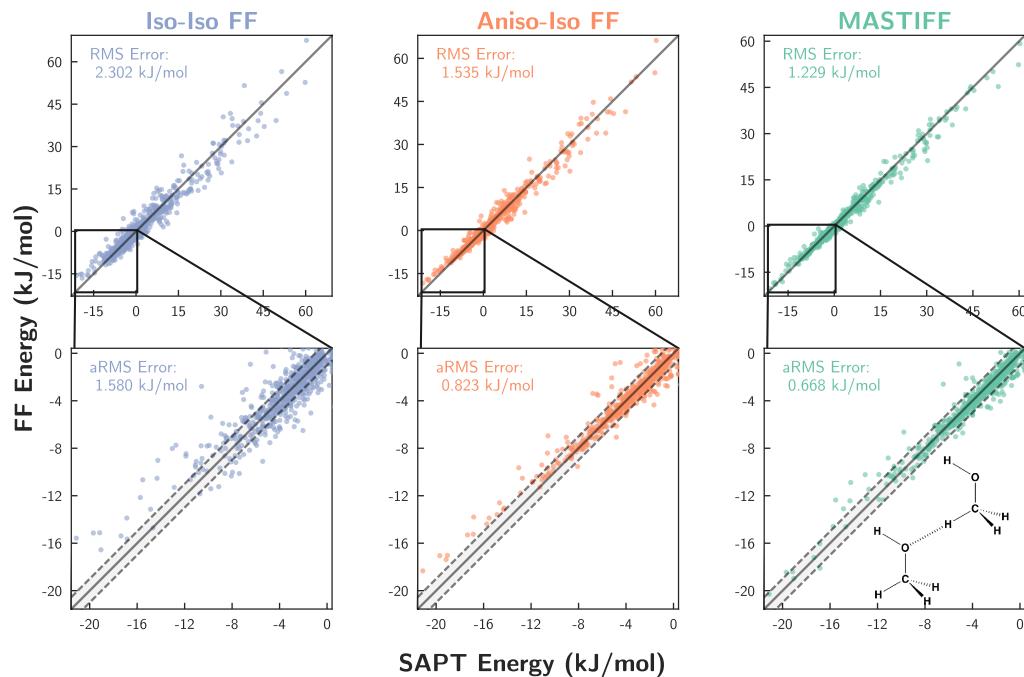


(i) H₂O Dimer

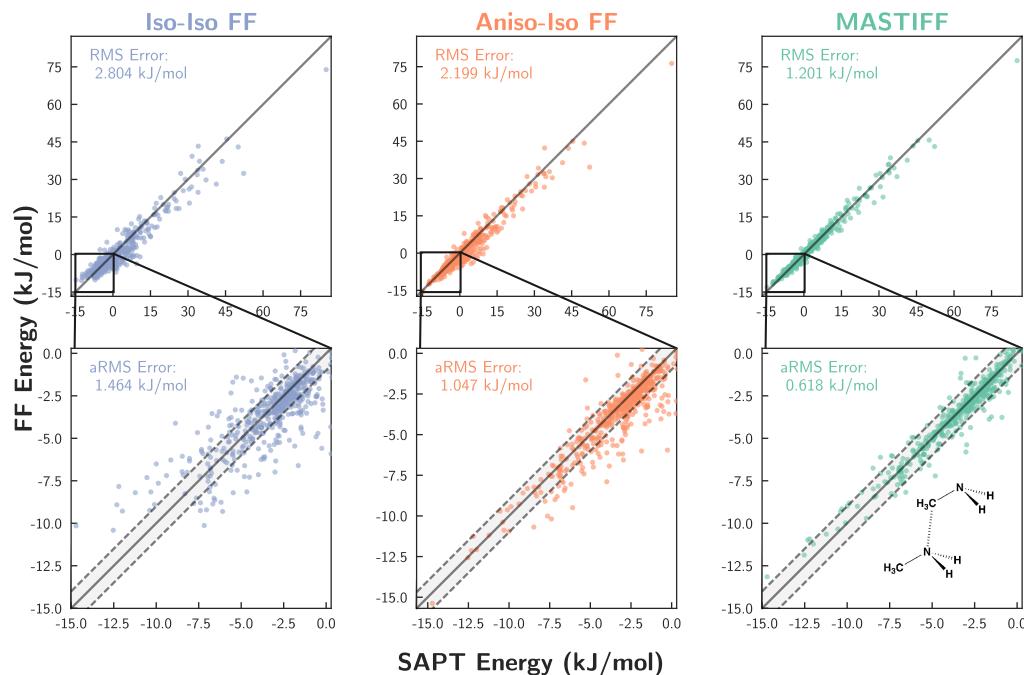
(j) Methane Dimer



(k) Methanol Dimer



(l) Methyl Amine Dimer



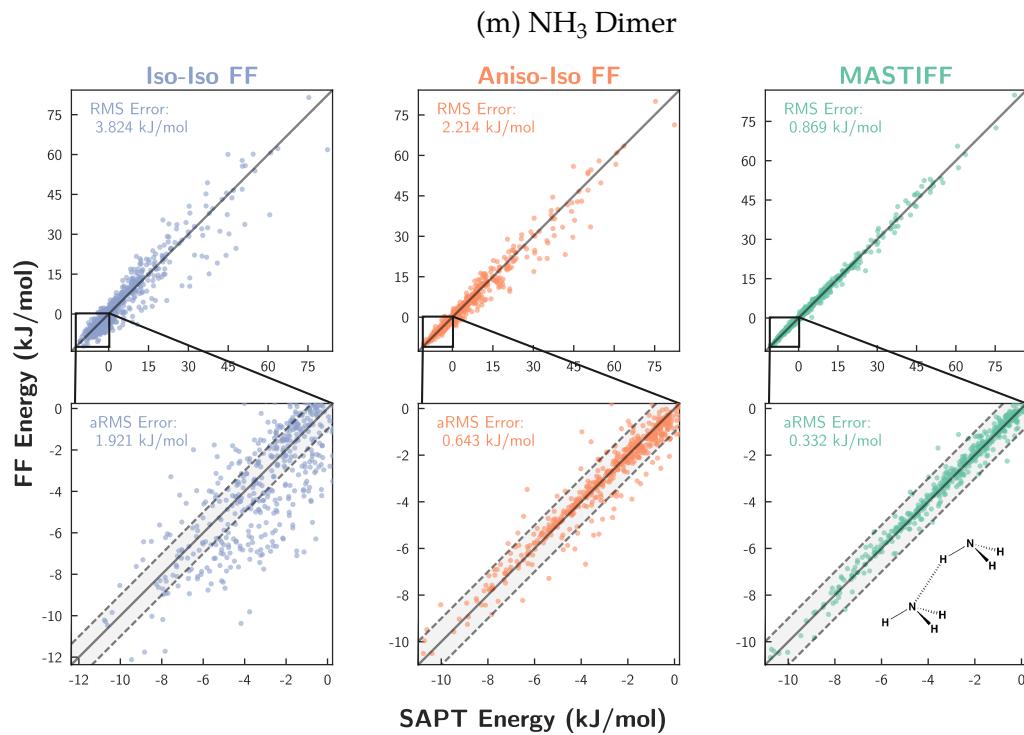


Figure 4.7: Force field fits for each homomeric systems using the Iso-Iso FF (purple), Aniso-Iso FF (orange), and MASTIFF (purple). Two views of the fit to the total energy are displayed along with corresponding RMSE (aRMSE for the inset showing attractive configurations). The $y = x$ line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within ± 1 kJ/mol agreement of the benchmark.

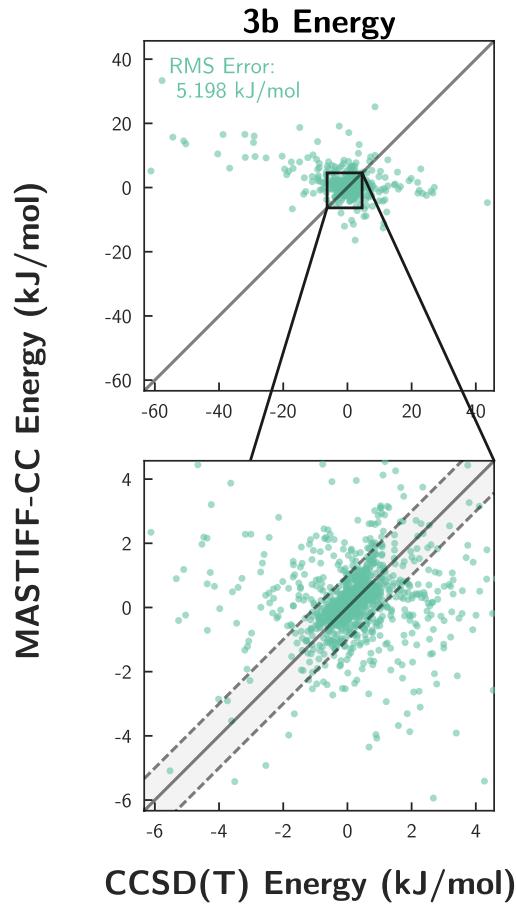


Figure 4.8: Three-body interaction energies for CO_2 as compared to the Hellmann¹² database of 9401 reference trimer configurations computed at a CCSD(T) level of theory.

4.D 2- and 3-body MASTIFF-CC CO_2 energies

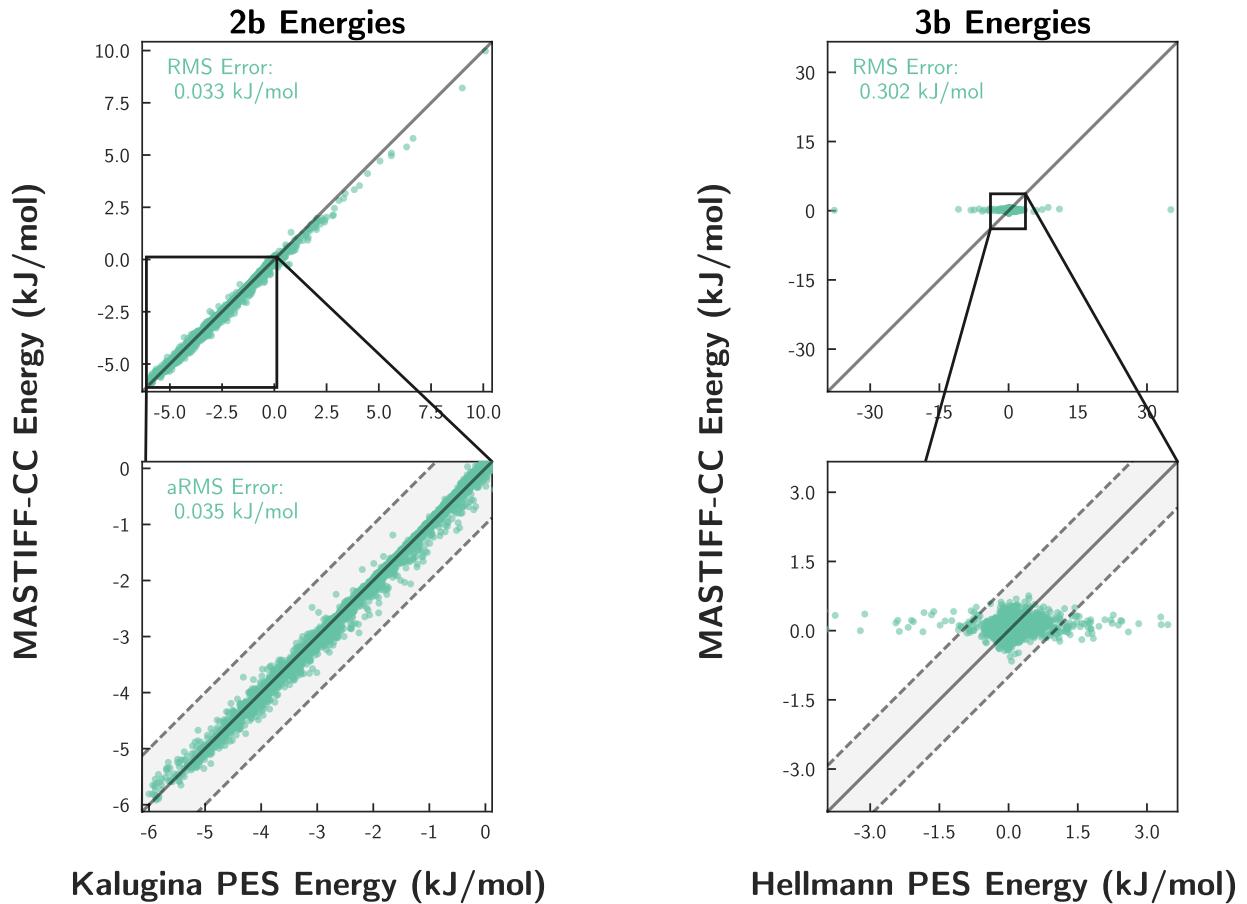


Figure 4.9: Force field quality for MASTIFF-CC in reproducing (left) two-body and (right) three-body CO_2 interaction energies. The $y = x$ line (solid) and $\pm 1 \text{ kJ/mol}$ boundaries (dashed lines) are shown for reference. All dimer/trimer configurations were taken from a snapshot of CO_2 liquid simulated using MASTIFF-CC at 273.15 K and 100 bar. Reference energies are taken from the Kalugina et al.¹³ PES for the two-body energies, and the Hellmann¹² PES for the three-body energies. A total of 62,583 dimer configurations and 43,784 trimer configurations are represented in the two plots.

Part III

Unpublished Work

Part IV

Practical Matters

Part V

Conclusions and Future Work

Part VI

Codes

A FORCE FIELD DEVELOPMENT WORKFLOW

Purpose

Derive a first-principles, SAPT-based force field.

Relevant Literature

- VanVleet2016: 10.1021/acs.jctc.6b00209
- VanVleet2017: TBA
- McDaniel2013: 10.1021/jp3108182
- Schmidt2015: 10.1021/ar500272n
- Yu2011: 10.1021/jp204563n

Overview

To generate a SAPT-based force field, the following inputs are required: 1. Benchmark dimer energies from SAPT, computed for a variety of dimer configurations 2. Long-range multipole moments, induced dipoles, and dispersion parameters, computed from monomer properties (and BS-ISA in particular) 3. Short-range exponents computed from monomer properties (and BS-ISA in particular) 4. Short-range pre-factors fit to dimer energies

The following scripts are designed to simplify (as much as is possible) the workflow for force field generation.

Method

1. Generate the necessary input files upon which the scripts in step #2 depend. The following files must be manually created/edited, and can all be found in the templates subdirectory (with an example set of input files given for the pyridine dimer):
 1. `dimer_info.dat`
 - For each monomer, list the monomer's name and the charge on the monomer. The appropriate file format should be clear from the pyridine example.
 - In the manner described in `dimer_info.dat`, list all midbonds that should be added between monomers. Midbonds are important for running accurate SAPT calculations; see Yu2011 for details.
 2. `generate_grid_settings.inp`
 - This is the input file for `GenerateGridPoints`, which generates the dimer configurations for running SAPT calculations. The input file is commented so as to be self-explanatory; you will need to change (at the very least) the 1st, 3rd, and 4th input sections based on the identities of the two monomers
 3. `MONA_MONB.inp` (where MONA and MONB are replaced by the monomer names listed in `dimer_info.dat`)
 - This file contains a title line (line 1), and (for each monomer) the number of atoms followed by a list of coordinates in .xyz format. See `pyridine_pyridine.inp` for an example.

4. MONA.atomtypes, MONB.atomtypes

- Each .atomtypes file has the format of a .xyz file, where the element names have been replaced by atomtypes. This file will be used to generate the CamCASP input files needed for ISA calculations, and is also necessary for pre-processing the input files for force field fitting.

2. To generate all files necessary to run force field calculations, run the following pre-processing scripts (from this main directory).

```
./scripts/make_geometries.sh
./scripts/get_global_coordinates.py
./scripts/submit_ip_calcs.py
```

(wait until IP calculation is finished)

```
./scripts/make_sapt_ifiles.py
./scripts/make_isa_files.py
./scripts/make_dispersion_files.py
```

3. Submit all SAPT and ISA calculations to relevant locations. At the time of this writing, SAPT calculations should preferably be run on HCTC (Condor). ISA and dispersion calculations should be run on Phoenix using Camcasp 5.8. Copy all output files back to Pople.

4. Workup the results of the SAPT and ISA calculations by running the following post-processing scripts:

```
./scripts/workup_sapt_energies.py
./scripts/workup_dispersion_files.sh
```

(Depending on the force field, dynamic polarizabilities may need to be added to templates/dispersion_base_constraints.index before running this script. See Jesse McDaniel's thesis and \cite{McDaniel2013} for a full description of the parameterization process for dispersion coefficients.)

```
./scripts/workup_drude_files.sh
```

(Depending on the force field, static polarizabilities may need to be added to templates/drude_base_constraints.index before running this script. See Jesse McDaniel's thesis and \cite{McDaniel2013} for a full description of the parameterization process for drude oscillator charges.)

```
./scripts/workup_isa_charges.py  
./scripts/workup_isa_exponents.py
```

After running these scripts, you should have the SAPT energies, long-range coefficients, and short-range exponents required to run the force fitting code (which is needed to generate short-range pre-factors, see \cite{VanVleet2016}). The proper running of this code is described in the POInter documentation, see

https://git.chem.wisc.edu/schmidt/force_fields/wikis/home

Overview of Important Files

- dimer_info.dat <- monomer names and midbond positions
- dispersion_template.clt <- CamCASP input file for getting induction and
- dispersion paramters
- generate_grid_settings.inp <- geometry configuration settings
- isa_template.clt <- CamCASP input file for getting ISA exponents
- pbe0_template.com <- DF-DFT-SAPT template for the PBE0 functional
- pyridine.atomtypes <- change elements to atomtypes; only matters for
- dispersion
- pyridine_pyridine.inp <- monomer geometries

For most systems, only dimer_info.dat, the .inp files, and the .atomtypes file will need to be changed. The examples provided for these files should hopefully make the format self-explanatory.

System Requirements

Python dependencies: * numpy * scipy * chemistry (mvanvleet package; not standard, so this needs to be * downloaded and added to your \$PYTHONPATH)

Figure A.1: An overview of the semi-automated force field development process. The full workflow and required scripts can be found at <https://github.com/mvanvleet/workflow-for-force-fields>.

A.1 Monomer Geometries

BIBLIOGRAPHY

- [2] McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.
- [3] Witt, R. K.; Kemp, J. D. *J. Am. Chem. Soc.* **1937**, *59*, 273–276.
- [4] Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Techniques of Chemistry, Vol. II: Organic Solvents: Physical Properties and Methods of Purification*, 4th ed.; Wiley-Interscience: New York, 1986.
- [5] Span, R.; Wagner, W. A new EOS for CO₂ covering the fluid region from the triple point temperature to 1100K at pressures up to 800MPa.pdf. 1996.
- [6] Dymond, J. H.; Smith, E. B. *The Virial Coefficients of Pure Gases and Mixtures*, 2nd ed.; Clarendon Press: Berlin Heidelberg, 1980.
- [7] Rosen, N. *Phys. Rev. Lett.* **1931**, *38*, 255–276.
- [8] Tai, H. *Phys. Rev. A* **1986**, *33*, 3657–3666.
- [9] Duška, M.; Hrubý, J. *EPJ Web Conf.* **2013**, *45*, 01024.
- [10] Tillner-Roth, R.; Harms-Watzenberg, F.; Baehr, H. D. *Dkv Tragungsbericht* **1993**, *20*, 67.
- [11] Massucci, M.; Wormald, C. *J. Chem. Thermodyn.* **1998**, *30*, 919–927.
- [12] Hellmann, R. *J. Chem. Phys.* **2017**, *146*, 054302.
- [13] Kalugina, Y. N.; Buryak, I. A.; Ajili, Y.; Vigasin, A. A.; Jaidane, N. E.; Hochlaf, M. *J. Chem. Phys.* **2014**, *140*, 234310.
- [14] van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chemie Int. Ed. English* **1990**, *29*, 992–1023.
- [15] Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. *Adv. Appl. Bioinforma. Chem.* **2015**, *8*, 37–47.

- [16] Chen, L.-Q.; Chen, L.-D.; Kalinin, S. V.; Klimeck, G.; Kumar, S. K.; Neugebauer, J.; Terasaki, I. *npj Comput. Mater.* **2015**, 1–2.
- [17] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, 9, 646–652.
- [18] Ciccotti, G.; Ferrario, M.; Schuette, C. *Molecular Dynamics Simulation*; 2014; pp 1–617.
- [19] Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, 32, 425–443.
- [20] Levitt, M.; Warshel, A. *Nature* **1975**, 253, 694–698.
- [21] Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2013**, 23, 58–65.
- [22] Piana, S.; Klepeis, J. L.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2014**, 24, 98–105.
- [23] Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. *Curr. Opin. Struct. Biol.* **2016**, 36, 25–31.
- [24] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. *J. Med. Chem.* **2016**, 59, 4035–4061.
- [25] Jiang, J.; Babarao, R.; Hu, Z. *Chem. Soc. Rev.* **2011**, 40, 3599–3612.
- [26] Maurin, G. *Chem. Met. Fram. Synth. Charact. Appl.* **2016**, 765.
- [27] Bereau, T.; Andrienko, D.; Kremer, K. *APL Mater.* **2016**, 4.
- [28] Kalikmanov, V. I. *Nucleation theory*; 2013; p 316.
- [29] Wilding, N. *Am. J. Phys.* **2001**, 69, 1147.
- [30] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science Publ; Clarendon Press, 1989.
- [31] Boltzmann, L. *Vorlesungen über Gastheorie: Th. Theorie van der Waals'; Gase mit zusammengesetzten Molekülen; Gasdissociation; Schlussbemerkungen; Vorlesungen über Gastheorie*; J. A. Barth, 1898.

- [32] Karplus, M. *Angew. Chemie Int. Ed.* **2014**, *53*, 9992–10005.
- [33] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- [34] Harrison, R. L.; Granja, C.; Leroy, C. Introduction to Monte Carlo Simulation. AIP Conf. Proc. 2010; pp 17–21.
- [35] Schneider, G.; Fechner, U. *Nat. Rev. Drug Discov.* **2005**, *4*, 649–663.
- [36] Jorgensen, W. L. *Science* **2004**, *303*, 1813–8.
- [37] Ballone, P. *Entropy* **2014**, *16*, 322–349.
- [38] Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. *Methods*; 2015; Vol. 1215; pp 47–71.
- [39] Saunders, M. G.; Voth, G. a. *Annu. Rev. Biophys.* **2013**, *42*, 73–93.
- [40] De Carvalho, F. F.; Bouduban, M. E. F.; Curchod, B. F. E.; Tavernelli, I. *Entropy* **2014**, *16*, 62–85.
- [41] Lei, H.; Duan, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187–191.
- [42] Grossfield, A.; Zuckerman, D. M. *Annu Rep Comput Chem* **2009**, *1400*, 23–48.
- [43] Theodorou, D. N. *Ind. Eng. Chem. Res.* **2010**, *49*, 3047–3058.
- [44] E, W.; Vanden-Eijnden, E. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391–420.
- [45] Pande, V. S.; Beauchamp, K.; Bowman, G. R. *Methods* **2010**, *52*, 99–105.
- [46] Rohrdanz, M. a.; Zheng, W.; Clementi, C. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295–316.
- [47] Chaplin, M. Protein Folding and Denaturation. 2017; http://www1.lsbu.ac.uk/water/protein{_}denatured.html.
- [48] Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; OUP Oxford, 2013.

- [49] Margenau, H.; Kestner, N. R. *Theory of Intermolecular Forces*; International series of monographs in natural philosophy; Pergamon Press: Oxford, 1969.
- [50] Riley, K. E.; Pitončák, M.; Jurecčka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.
- [51] Stone, A. J.; Misquitta, A. J. *Int. Rev. Phys. Chem.*; 2007; Vol. 26; pp 193–222.
- [52] Dykstra, C. E.; Lisy, J. M. *J. Mol. Struct. THEOCHEM* **2000**, *500*, 375–390.
- [53] Stone, A. J.; Tough, R. *Chem. Phys. Lett.* **1984**, *110*, 123–129.
- [54] Williams, G. J.; Stone, A. J. *J. Chem. Phys.* **2003**, *119*, 4620–4628.
- [55] Misquitta, A. J.; Stone, A. J. *J. Chem. Phys.* **2006**, *124*, 024111.
- [56] Dehez, F.; Chipot, C.; Millot, C.; Ángyán, J. G. *Chem. Phys. Lett.* **2001**, *338*, 180–188.
- [57] Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.
- [58] Misquitta, A. J.; Stone, A. J.; Fazeli, F. *J. Chem. Theory Comput.* **2014**, *10*, 5405–5418.
- [59] Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
- [60] Szalewicz, K. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 254–272.
- [61] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- [62] Grimme, S.; Djukic, J. P. *Inorg. Chem.* **2011**, *50*, 2619–2628.
- [63] Lennard-Jones, J. *Proc. Phys. Soc.* **1931**, *43*, 461–482.
- [64] Born, M.; Mayer, J. E. *Zeitschrift für Phys.* **1932**, *75*, 1–18.
- [65] Buckingham, R. A. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1938**, *168*, 264–283.

- [66] Nezbeda, I. *Mol. Phys.* **2005**, *103*, 59–76.
- [67] Galliero, G.; Boned, C. *J. Chem. Phys.* **2008**, *129*.
- [68] Gordon, P. A. *J. Chem. Phys.* **2006**, *125*.
- [69] Ruckenstein, E.; Liu, H. *Society* **1997**, 3927–3936.
- [70] Galliero, G.; Boned, C.; Baylaucq, A.; Montel, F. *Chem. Phys.* **2007**, *333*, 219–228.
- [71] Wu, G.-W.; Sadus, R. J. *Fluid Phase Equilib.* **2000**, *170*, 269–284.
- [72] Errington, J. R.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **1998**, *109*, 1093–1100.
- [73] McGrath, M. J.; Ghogomu, J. N.; Tsона, N. T.; Siepmann, J. I.; Chen, B.; Napani, I.; Vehkamaki, H. *J. Chem. Phys.* **2010**, *133*.
- [74] Parker, T. M.; Sherrill, C. D. *J. Chem. Theory Comput.* **2015**, *11*, 4197–4204.
- [75] Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187–2193.
- [76] Zgarbová, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476–10493.
- [77] Bastea, S. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **2003**, *68*, 031204.
- [78] Errington, J. R.; Panagiotopoulos, A. Z. *J. Phys. Chem. B* **1999**, *103*, 6314–6322.
- [79] Ross, M.; Ree, F. H. *J. Chem. Phys.* **1980**, *73*, 6146.
- [80] Schmidt, J. R.; Yu, K.; McDaniel, J. G. *Acc. Chem. Res.* **2015**, *48*, 548–556.
- [81] Abrahamson, A. A. *Phys. Rev.* **1963**, *130*, 693–707.
- [82] Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- [83] Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

- [84] Kim, Y. S.; Kim, S. K.; Lee, W. D. *Chem. Phys. Lett.* **1981**, *80*, 574–575.
- [85] Misquitta, A. J.; Stone, A. J. Ab initio atom-atom potentials using CamCASP: Theory. 2015; <https://arxiv.org/abs/1512.06150v2>.
- [86] Nyeland, C.; Toennies, J. P. *Chem. Phys. Lett.* **1986**, *127*, 3–8.
- [87] Ihm, G.; Cole, M. W.; Toigo, F.; Klein, J. R. *Phys. Rev. A* **1990**, *42*, 5244–5252.
- [88] Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-P.; Cisneros, G. A. *J. Chem. Theory Comput.* **2014**, *10*, 1361–1365.
- [89] Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190–202.
- [90] Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*.
- [91] Chaudret, R.; Gresh, N.; Narth, C.; Lagardere, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J. P. *J. Phys. Chem. A* **2014**, *118*, 7598–7612.
- [92] Chaudret, R.; Gresh, N.; Cisneros, G. A.; Scemama, A.; Piquemal, J.-p. *Can. J. Chem.* **2013**, *91*, 804–810.
- [93] Öhrn, A.; Hermida-Ramon, J. M.; Karlström, G. *J. Chem. Theory Comput.* **2016**, *12*, 2298–2311.
- [94] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- [95] Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- [96] Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
- [97] Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459–467.

- [98] Patil, S. H.; Tang, K. T. In *Asymptotic methods in quantum mechanics*; Scäfer, F. P., Toennies, J. P., Zinth, W., Eds.; Springer, 2000.
- [99] Hoffmann-Ostenhof, M.; Hoffmann-Ostenhof, T. *Phys. Rev. A* **1977**, *16*, 1782–1785.
- [100] Amovilli, C.; March, N. H. *J. Phys. A. Math. Gen.* **2006**, *39*, 7349–7357.
- [101] Bunge, A. V.; Esquivel, R. O. *Phys. Rev. A* **1986**, *34*, 853.
- [102] Rappe, A. K.; Casewit, C.; Colwell, K.; Goddard, W. A. I.; Skiff, W. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- [103] Waldman, M.; Hagler, A. T. *J. Comput. Chem.* **1993**, *14*, 1077–1084.
- [104] Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.
- [105] McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. C* **2012**, *116*, 14031–14039.
- [106] Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.
- [107] Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A. *J. Chem. Phys.* **2006**, *125*, 044301.
- [108] Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2007**, *127*, 124303.
- [109] Sum, A. K.; Sandler, S. I.; Bukowski, R.; Szalewicz, K. *J. Chem. Phys.* **2002**, *116*, 7637.
- [110] Konieczny, J. K.; Sokalski, W. A. *J. Mol. Model.* **2015**, *21*, 197.
- [111] Kita, S.; Noda, K.; Inouye, H. *J. Chem. Phys.* **1976**, *64*, 3446–3449.
- [112] Kuechler, E. R.; Giese, T. J.; York, D. M. *J. Chem. Phys.* **2015**, *143*.

- [113] Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.
- [114] Giese, T. J.; Chen, H.; Dissanayake, T.; GiambaÅYu, G. M.; Hedenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K.; York, D. M. *J. Chem. Theory Comput.* **2013**, *9*, 1417–1427.
- [115] Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2003**, *125*, 16434–16443.
- [116] Nobel, I.; Price, S. L.; Wheatley, R. *J. Mol. Phys.* **1998**, *95*, 525–537.
- [117] McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. A* **2013**, *117*, 2053–2066.
- [118] Totton, T. S.; Misquitta, A. J.; Kraft, M. *J. Chem. Theory Comput.* **2010**, *6*, 683–695.
- [119] Misquitta, A. J. *J. Chem. Theory Comput.* **2013**, *9*, 5313–5326.
- [120] Misquitta, A. J.; Stone, A. J. Ab initio atom-atom potentials using CamCASP: Application to Pyridine. 2015; <https://arxiv.org/abs/1512.06155v2>.
- [121] Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726–3741.
- [122] Tang, K. T.; Toennies, J. P. *Surf. Sci.* **1992**, *279*, L203–L206.
- [123] Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *69*, 507–533.
- [124] Mitchell, J. B. O.; Price, S. L. *J. Phys. Chem. A* **2000**, *104*, 10958–10971.
- [125] Söderhjelm, P.; Karlström, G.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 244101.
- [126] Tkatchenko, A.; Distasio, R. A.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, 1–5.
- [127] Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 6–9.
- [128] Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.
- [129] Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.

- [130] Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2012**, *8*, 2844–2867.
- [131] Yu, K.; McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. B* **2011**, *115*, 10054–10063.
- [132] Levy, M.; Perdew, J. P.; Sahni, V. *Phys. Rev. A* **1984**, *30*, 2745–2748.
- [133] Kitaigorodsky, A. *Molecular crystals and Molecules*; Physical Chemistry; Elsevier Science: New York, 2012.
- [134] Lillestolen, T. C.; Wheatley, R. *J. Chem. Commun.* **2008**, *7345*, 5909–5911.
- [135] Lillestolen, T. C.; Wheatley, R. *J. J. Chem. Phys.* **2009**, *131*, 144101.
- [136] Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129–138.
- [137] Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301–306.
- [138] Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*, 033201.
- [139] Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*.
- [140] Heßelmann, A.; Jansen, G.; Schulz, M. *J. Chem. Phys.* **2005**, *122*, 014103.
- [141] Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Chem. Theory Comput.* **2006**, *2*, 400–412.
- [142] Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319–325.
- [143] Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
- [144] Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
- [145] Jansen, G.; Hesselmann, A.; Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 11156–11158.
- [146] Podeszwa, R.; Szalewicz, K. Accurate interaction energies from perturbation theory based on Kohn-Sham model. 2005; <http://arxiv.org/abs/physics/0501023>.

- [147] Jeziorska, M.; Jeziorski, B.; Čížek, J. *Int. J. Quantum Chem.* **1987**, *32*, 149–164.
- [148] Drude, P.; Riborg, C.; Millikan, R. A. *The Theory of Optics... Translated from the German by CR Mann and RA Millikan*; London; New York [printed], 1902.
- [149] Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.
- [150] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. 2015; <http://cccbdb.nist.gov/>.
- [151] Shoemake, K. *Graph. Gems 3*; 1992; Chapter 6, pp 124–132.
- [152] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *WIREs Comput Mol Sci* **2012**, *2*, 242–253.
- [153] Misquitta, A. J.; Stone, A. J. CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form, version 5.8. University of Cambridge, 2015.
- [154] Aidas, K. et al. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.
- [155] Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. ORIENT: a program for studying interactions between molecules, version 4.8. 2015; <http://www-stone.ch.cam.ac.uk/programs/orient.html>.
- [156] Ferenczy, G. G.; Winn, P. J.; Reynolds, C. a. *J. Phys. Chem. A* **1997**, *101*, 5446–5455.
- [157] Eastman, P. et al. *J. Chem. Theory Comput.* **2013**, *9*, 461–469.
- [158] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [159] Sebetci, A.; Beran, G. J. O. *J. Chem. Theory Comput.* **2010**, *6*, 155–167.
- [160] Misquitta, A.; Welch, G.; Stone, A.; Price, S. **2008**, *456*, 105–109.

- [161] Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478–8490.
- [162] Totton, T. S.; Misquitta, A. J.; Kraft, M. *J. Chem. Theory Comput.* **2010**, *6*, 683–695.
- [163] Hermida-Ramón, J. M.; Ríos, M. A. *Chem. Phys.* **2000**, *262*, 423–436.
- [164] Nyeland, C. *Chem. Phys.* **1990**, *147*, 229–240.
- [165] Vogel, E.; Jäger, B.; Hellmann, R.; Bich, E. *Mol. Phys.* **2010**, *108*, 3335–3352.
- [166] McDaniel, J. G.; Yu, K.; Schmidt, J. R. *J. Phys. Chem. C* **2012**, *116*, 1892–1903.
- [167] Mayo, S. L.; Olafson, B. D.; Goddard, W. A. I. *J. Phys. Chem.* **1990**, *101*, 8897–8909.
- [168] Lim, T. *Zeitschrift für Naturforschung-A* **2009**, *64*, 200–204.
- [169] Van Duin, a. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. a. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- [170] Eramian, H.; Tian, Y.-H.; Fox, Z.; Beneberu, H. Z.; Kertesz, M. *J. Phys. Chem. A* **2013**, *117*, 14184–14190.
- [171] Badenhoop, J. K.; Weinhold, F. *J. Chem. Phys.* **1997**, *107*, 5422.
- [172] Kim, H.; Doan, V. D.; Cho, W. J.; Madhav, M. V.; Kim, K. S. *Sci. Rep.* **2014**, *4*, 1–8.
- [173] Wheatley, R. J.; Gopal, A. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2087–2091.
- [174] Rezáč, J.; Hobza, P. *Chem. Rev.* **2016**, *116*, 5038–5071.
- [175] Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **2000**, *100*, 4227–4252.
- [176] Jan, S.; Martin, M. L.; Chem, P.; Phys, C. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905–20925.

- [177] Hassanali, A. A.; Cuny, J.; Verdolino, V.; Parrinello, M. **2014**,
- [178] Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325–3335.
- [179] Price, S. L. *Rev. Comput. Chem.* **2000**, *14*, 225–289.
- [180] Coppens, P.; Guru Row, T. N.; Leung, P.; Stevens, E. D.; Becker, P. J.; Yang, Y. W. *Acta Crystallogr. Sect. A* **1979**, *35*, 63–72.
- [181] Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.
- [182] Nyburg, S. C.; Faerman, C. H. *Acta Crystallogr. Sect. B Struct. Sci.* **1985**, *B41*, 274–279.
- [183] Batsanov, S. S. *Inorg. Mater. Transl. from Neorg. Mater. Orig. Russ. Text* **2001**, *37*, 871–885.
- [184] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16789–94.
- [185] Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. *J. Am. Chem. Soc.* **1996**, *118*, 3108–3116.
- [186] Kramer, C.; Spinn, A.; Liedl, K. R. *J. Chem. Theory Comput.* **2014**, *10*, 4488–4496.
- [187] Badenhoop, J. K.; Weinhold, F. *J. Chem. Phys.* **1997**, *107*, 5422.
- [188] Bankiewicz, B.; Palusiak, M. *Struct. Chem.* **2012**, *24*, 1297–1306.
- [189] Chessari, G.; Hunter, C. A.; Low, C. M. R.; Packer, M. J.; Vinter, J. G.; Zonta, C. *Chem. - A Eur. J.* **2002**, *8*, 2860–2867.
- [190] Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. *Biopolymers* **2013**, *99*, 978–988.
- [191] Bartocci, A.; Belpassi, L.; Cappelletti, D.; Falcinelli, S.; Grandinetti, F.; Tarantelli, F.; Pirani, F. *J. Chem. Phys.* **2015**, *142*, 184304.

- [192] Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508–19516.
- [193] Politzer, P.; Murray, J. S.; Concha, M. C. *J. Mol. Model.* **2008**, *14*, 659–665.
- [194] Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. *Chem. Rev.* **2016**, *116*, 7501–7528.
- [195] Cardamone, S.; Hughes, T. J.; Popelier, P. L. a. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10367.
- [196] Hagler, A. T. *J. Chem. Theory Comput.* **2015**, *11*, 5555–5572.
- [197] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- [198] Ponder, J. W.; Wu, C.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- [199] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.
- [200] Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478.
- [201] Day, G. M.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2005**, *5*, 1023–1033.
- [202] Price, S. L. *Int. Rev. Phys. Chem.* **2008**, *27*, 541–568.
- [203] Misquitta, A. J.; Stone, A. J. *J. Chem. Theory Comput.* **2016**, *12*, 4184–4208.
- [204] Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.
- [205] Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975–1997.

- [206] Williams, D. E. *J. Comput. Chem.* **1988**, *9*, 745–763.
- [207] Demerdash, O.; Yap, E.-H.; Head-Gordon, T. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149–74.
- [208] Chaudret, R.; Gresh, N.; Narth, C.; Lagarde, L.; Darden, T. A.; Andre, G.; Piquemal, J.-p. **2014**,
- [209] Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*.
- [210] Dixon, R. W.; Kollman, P. a. *J. Comput. Chem.* **1997**, *18*, 1632–1646.
- [211] Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; Jr, A. D. M. **2006**, 1587–1597.
- [212] Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. *J. Phys. Chem. B* **2014**, *118*, 6456–6465.
- [213] Wikfeldt, K. T.; Batista, E. R.; Vila, F. D.; Jónsson, H. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16542–56.
- [214] Piquemal, J. P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem. A* **2007**, *111*, 8170–8176.
- [215] Loboda, O.; Ingrosso, F.; Ruiz-López, M. F.; Szalewicz, K.; Millot, C. *J. Chem. Phys.* **2016**, *144*.
- [216] Misquitta, A. J.; Stone, A. J. *Mol. Phys.* **2008**, *106*, 1631–1643.
- [217] Langhoff, P. W. *J. Chem. Phys.* **1971**, *55*, 2126.
- [218] Krishtal, A.; Vannomeslaeghe, K.; Geldof, D.; Van Alsenoy, C.; Geerlings, P. *Phys. Rev. A - At. Mol. Opt. Phys.* **2011**, *83*, 3–6.
- [219] Stone, A. J. *J. Am. Chem. Soc.* **2013**, *135*, 7005–7009.
- [220] Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-p.; Andre, G. **2014**,

- [221] Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344–2353.
- [222] Torheyden, M.; Jansen, G. *Mol. Phys.* **2006**, *104*, 2101–2138.
- [223] Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J. *J. Phys. Chem. A* **2001**, *105*, 9961–9971.
- [224] Misquitta, A. J.; Stone, A. J.; Price, S. L. *J. Chem. Theory Comput.* **2008**, *4*, 19–32.
- [225] Holt, A.; Karlström, G. *J. Comput. Chem.* **2008**, *29*, 2033–2038.
- [226] Holt, A.; Boström, J.; Karlström, G.; Lindh, R. **2010**,
- [227] Parker, T. M.; Burns, L. a.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. *J. Chem. Phys.* **2014**, *140*, 094106.
- [228] Werner, H.-J. et al. MOLPRO, version 2012.1, a package of ab initio programs. 2012.
- [229] Oakley, M. T.; Wheatley, R. J. *J. Chem. Phys.* **2009**, *130*, 034110.
- [230] Korona, T. *Theor. Chem. Acc.* **2011**, *129*, 15–30.
- [231] Giauque, W. F.; Egan, C. J. *J. Chem. Phys.* **1937**, *5*, 45.
- [232] Cervinka, C.; Fulem, M. *J. Chem. Theory Comput.* **2017**, *13*, 2840–2850.
- [233] Heit, Y. N.; Nanda, K. D.; Beran, G. J. O. *Chem. Sci.* **2016**, *7*, 246–255.
- [234] Simon, A.; Peters, K. *Acta Crystallogr. Sect. B* **1980**, *36*, 2750–2751.
- [235] Yu, K.; Schmidt, J. R. *J. Chem. Phys.* **2012**, *136*, 034503.
- [236] Misquitta, A. J.; Stone, A. J. *J. Chem. Theory Comput.* **2007**, *7*–18.
- [237] Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. *J. Chem. Theory Comput.* **2017**, *acs.jctc.7b00225*.
- [238] Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.

- [239] Van Vleet, M. J.; Misquitta, A. J.; Schmidt, J. R. *J. Chem. Phys.* **2017**, invited, manuscript in preparation.
- [240] Hawkins, D. M. **2004**, 1–12.
- [241] Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.
- [242] Bukowski, R.; Sadlej, J.; Jeziorski, B.; Jankowski, P.; Szalewicz, K.; Kucharski, S. A.; Williams, H. L.; Rice, B. M. *J. Chem. Phys. Addit. Inf. J. Chem. Phys. J. Homepage* **1999**, 110.
- [243] Babin, V.; Leforestier, C.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.
- [244] Desgranges, C.; Delhommelle, J. **2015**,
- [245] Pérez-Sánchez, G.; González-Salgado, D.; Piñeiro, M. M.; Vega, C. *J. Chem. Phys.* **2013**, *138*, 084506.
- [246] Stone, A. *J. Mol. Phys.* **1978**, *36*, 241–256.
- [247] Millot, C.; Stone, A. *Mol. Phys.* **1992**, *77*, 439–462.
- [248] Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. *Science (80-.)* **2013**, *341*, 1230444–1230444.
- [249] Millward, A. R.; Yaghi, O. M. *J. Am. Chem. Soc.* **2005**, *127*, 17998–17999.
- [250] Dietzel, P. D. C. et al. *J. Mater. Chem.* **2009**, *19*, 7362.
- [251] Dzubak, A. L.; Lin, L.-C.; Kim, J.; Swisher, J. a.; Poloni, R.; Maximoff, S. N.; Smit, B.; Gagliardi, L. *Nat. Chem.* **2012**, *4*, 810–816.
- [252] Czaja, A. U.; Trukhan, N.; Müller, U. *Chem. Soc. Rev.* **2009**, *38*, 1284.
- [253] Krishna, R.; van Baten, J. M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10593–10616.

- [254] Getman, R. B.; Bae, Y.-s.; Wilmer, C. E.; Snurr, R. Q.; Carlo, M. *Adsorpt. J. Int. Adsorpt. Soc.* **2012**, 703–723.
- [255] Yazaydin, a. O.; Snurr, R. Q.; Park, T.-H.; Koh, K.; Liu, J.; Levan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. *J. Am. Chem. Soc.* **2009**, 131, 18198–9.
- [256] Valenzano, L.; Civalleri, B.; Chavan, S.; Palomino, G. T.; Areán, C. O.; Bordiga, S. *J. Phys. Chem. C* **2010**, 114, 11185–11191.
- [257] Poloni, R.; Lee, K.; Berger, R. F.; Smit, B. **2014**,
- [258] Lin, L.-c.; Lee, K.; Gagliardi, L.; Smit, B. *J. Chem. Theory Comput.* **2014**, 10, 1477–1488.
- [259] Haldoupis, E.; Borycz, J.; Shi, H.; Vogiatzis, K. D.; Bai, P.; Queen, W. L.; Gagliardi, L.; Siepmann, J. I. *J. Phys. Chem. C* **2015**, 119, 16058–16071.
- [260] Mercado, R.; Vlaisavljevich, B.; Lin, L.-C.; Lee, K.; Lee, Y.; Mason, J. A.; Xiao, D. J.; Gonzalez, M. I.; Kapelewski, M. T.; Neaton, J. B.; Smit, B. *J. Phys. Chem. C* **2016**, acs.jpcc.6b03393.
- [261] Becker, T. M.; Heinen, J.; Dubbeldam, D.; Lin, L.-C.; Vlugt, T. J. H. *J. Phys. Chem. C* **2017**, acs.jpcc.6b12052.
- [262] McDaniel, J. G.; Li, S.; Tylianakis, E.; Snurr, R. Q.; Schmidt, J. R. *J. Phys. Chem. C* **2015**, 119, 3143–3152.
- [263] Lao, K. U.; Schaeffer, R.; Jansen, G.; Herbert, J. M. *J. Chem. Theory Comput.* **2015**, 150417132228001.
- [264] Pastorcak, E.; Corminboeuf, C. *J. Chem. Phys.* **2017**, 146, 120901.
- [265] Żuchowski, P. *Chem. Phys. Lett.* **2008**, 450, 203–209.
- [266] Horn, P. R.; Head-gordon, M. *Phys. Chem. Chem. Phys.* **2016**, 18, 23067–23079.

- [267] Su, P.; Li, H. *J. Chem. Phys.* **2009**, *131*, 014102.
- [268] Chen, Y.; Li, H. *J. Phys. Chem. A* **2010**, *114*, 11719–24.
- [269] Su, P.; Jiang, Z.; Chen, Z.; Wu, W. *J. Phys. Chem. A* **2014**, *118*, 2531–42.
- [270] Fedorov, D. G.; Kitaura, K. **2006**,
- [271] McDaniel, J. G. Development and Application of Physically-Motivated First-Principles Force Fields for Complex Chemical Systems. Ph.D. thesis, UW-Madison, 2014.
- [272] Yu, K.; Kiesling, K.; Schmidt, J. R. **2012**,
- [273] Verma, P.; Xu, X.; Truhlar, D. G. **2013**,
- [274] Valenzano, L.; Civalleri, B.; Sillar, K.; Sauer, J. **2011**, 21777–21784.
- [275] Haldoupis, E.; Borycz, J.; Shi, H.; Vogiatzis, K. D.; Bai, P.; Queen, W. L.; Gagliardi, L.; Siepmann, J. I. *J. Phys. Chem. C* **2015**, *74*, 150616135429005.
- [276] Jansen, G.; Scha, R. **2012**,
- [277] Frenkel, D.; Smit, B. *Acad. Press*; 2002; Vol. New York,; p 638.
- [278] Guibas, L. Representing rotations with quaternions. 1992; *graphics.stanford.edu/courses/cs164-09-spring/Handouts/handout12.pdf*.
- [279] Unke, O. T.; Devereux, M.; Meuwly, M. *J. Chem. Phys.* **2017**, *147*, 161712.
- [280] Halgren, T. A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205–10.
- [281] Rick, S. W.; Stuart, S. J. *Potentials and Algorithms for Incorporating Polarizability in Computer Simulations*; 2002; Vol. 18.
- [282] Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc.* **2009**, *124*, 11–28.

- [283] Cisneros, G. A.; Wikfeldt, K. T.; Ojam??e, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bart??k, A. P.; Cs??nyi, G.; Molinero, V.; Paesani, F. *Chem. Rev.* **2016**, *116*, 7501–7528.
- [284] Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. *J. Chem. Theory Comput.* **2008**, *4*, 522–532.
- [285] Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. *J. Phys. Condens. Matter* **2009**, *21*, 333102.
- [286] Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999–1016.
- [287] Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.
- [288] Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M.-J.; Wang, J.; Duan, Y.; Luo, R. *J. Phys. Chem. B* **2012**, *116*, 7999–8008.
- [289] Bistoni, G.; Belpassi, L.; Tarantelli, F. *J. Chem. Theory Comput.* **2016**,
- [290] Lao, K. U.; Herbert, J. M. *J. Chem. Theory Comput.* **2016**, *acs.jctc.6b00155*.