WASHINGTON UNIVERSITY
SCHOOL OF ENGINEERING AND APPLIED SCIENCES
DEPARTMENT OF BIOMEDICAL ENGINEERING

THE THEORY AND EFFECT OF SOLVENT ENVIRONMENT ON BIOMOLECULES

By
Michael J. Schnieders
Prepared under the direction of Professor Jay W. Ponder

A dissertation presented to the School of Engineering and Applied Sciences at
Washington University in partial fulfillment of the
requirements for the degree of
DOCTOR OF SCIENCE

December 2007
St. Louis, Missouri

WASHINGTON UNIVERSITY
SCHOOL OF ENGINEERING AND APPLIED SCIENCES
DEPARTMENT OF BIOMEDICAL ENGINEERING

ABSTRACT

THE THEORY AND EFFECT OF SOLVENT
ENVIRONMENT ON BIOMOLECULES

By

Michael J. Schnieders

ADVISOR: Professor Jay W. Ponder

December 2007

St. Louis, Missouri

This dissertation describes the theory and effect of solvent environment on biomolecules using a computational model known as a force field. Force fields are based on formulating an efficient, empirical function of atomic coordinates designed to reproduce the potential energy surface predicted by the more rigorous, but also intractably expensive Schrödinger equation. In particular, this work is novel due to use of an Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field that represents charge density using polarizable atomic multipoles. Polarizable Multipole Poisson-Boltzmann (PMPB) and generalized Kirkwood (GK) continuum electrostatics models are described that interact self-consistently with AMOEBA biomolecules. In conjunction with a novel apolar estimator, the PMPB and GK models are used to construct two implicit solvents for solutes represented by the AMOEBA force field. The effect of solvent environment on the electrostatic moments of a large set of folded proteins is examined.

Dedicated to my grandparents, parents and sister

Ralph and Lillian Schnieders,
Gerald and Isobel Strathman,
Jerome and Susan Schnieders
and
Laura Maureen

Contents

List of Tables ..... **vii**

List of Figures ..... **xiii**

Acknowledgements ..... **xvii**

1 Introduction **1**

1.1 The Theory of Biomolecular Solvation 2

1.1.1 Numerical Continuum Electrostatics 6

1.1.2 Analytic Continuum Electrostatics 10

1.2 The Effect of Solvation on Biomolecules 13

2 Theoretical Background **15**

2.1 AMOEBA Vacuum Electrostatic Energy 15

2.2 Fixed Charge Linearized Poisson-Boltzmann Energy and Gradient 19

2.3 The Generalized Born Model 22

2.3.1 Effective Radii and the Self-Energy 22

2.3.2 Cross-term Energy 24

3 Polarizable Multipole Poisson-Boltzmann **26**

3.1 Atomic Multipoles as the Source Charge Density 27

3.2 Permittivity and Modified Debye-Hückel Screening Factor 36

3.3 Boundary Conditions 40

3.4 Permanent Multipole Energy and Gradient 45

3.5 Self-Consistent Reaction Field 46

3.6 PMPB Electrostatic Solvation Free Energy 49

3.7 Polarization Energy Gradient 50

3.7.1 Direct Polarization Energy Gradient 53

3.7.2 Mutual Polarization Energy Gradient 55

3.8 PMPB Validation and Application 57

3.8.1 Energy 57

3.8.2 Energy Gradient 61

3.8.3 The Electrostatic Response of Solvated Proteins 67

4 Generalized Kirkwood 73

4.1 Effective Radii and the Multipole Self-Energy 73

4.1.1 The Solvent Field Approximation 76

4.1.2 The Reaction Potential Approximation 86

4.1.3 Self-energy accuracy 89

4.2 Multipole Cross-Term Energy 91

4.2.1 Generalized Kirkwood Auxiliary Reaction Potential 91

4.2.2 Generalized Kirkwood Cross-Term 94

4.3 Factoring of Generalized Kirkwood Tensors 98

4.4 AMOEBA Solutes in a Generalized Kirkwood Continuum 105

4.4.1 Electrostatic Solvation Free Energy 105

4.4.2 Permanent Multipole Energy Gradient 112

4.4.3 Polarization Energy Gradient 113

4.5 Validation and Application 118

4.5.1 Electrostatic Solvation Free Energy of Proteins 119

4.5.2 Dipole Moment of Solvated Proteins..... 122

5 Implicit Solvents for the AMOEBA Force Field 125

5.1 Cavitation Free Energy 126

5.1.1 Cavitation Measurements..... 128

5.1.2 Cavitation Model and Parameterization..... 133

5.2 Dispersion Free Energy..... 137

5.2.1 Dispersion Measurements..... 138

5.2.2 Dispersion Model and Parameterization..... 140

5.3 Solvation Free Energy of Small Molecules 144

6 Spherical Solvent Boundary Potential for Multipoles 147

6.1 Pairwise Electrostatic Solvation Free Energy..... 148

6.2 Electrostatic Solvation Self-Energy..... 150

7 Conclusions 152

7.1 Polarizable Multipole Poisson-Boltzmann 153

7.2 Generalized Kirkwood..... 154

Appendix A Finite-Difference Representation of the LPBE 156

Appendix B Representation of the Delta-Functional Using B-splines..... 157

Appendix C Permanent and Polarization PMPB Forces..... 159

C.1 Permanent Reaction Field Force and Torque..... 159

C.2 Direct Polarization Reaction Field Force and Torque 160

C.3 Mutual Polarization Reaction Field Force 161

C.4 Permanent Dielectric Boundary Force 161

C.5 Direct and Mutual Polarization Dielectric Boundary Forces 163

C.6 Permanent Ionic Boundary Force 164

C.7 Direct and Mutual Polarization Ionic Boundary Force 165

Appendix D Gradients of the Generalized Kirkwood Tensors 166

References 176

Curriculum Vitae 189

List of Tables

Table 3.1. The norm of the gradient sum over all atoms (kcal/mole/Å) for three different solutes is shown for cubic, quintic and heptic characteristic functions at two grid spacings. A norm of zero, indicating perfect conservation of energy, is nearly achieved for acetamide and ethanol at 0.11 Å grid spacing using a heptic characteristic function. Conservation of energy is improved by reducing grid spacing and also by increasing the continuity of the solute-solvent boundary via the characteristic function. 39

Table 3.3. Explicit values for the functions $\alpha_n(x)$ and $k_n(x)$ up to quadrupole order. 44

Table 3.4. Explicit values of the coefficients used to calculate the potential at the grid boundary of LPBE and PE calculations, respectively, under the SDH or MDH approximation. The LPBE coefficients reduce to the PE coefficients as salt concentration goes to zero. 44

Table 3.5. As grid spacing decreases, the numerical solution to the PE approaches the analytic solution for four canonical test cases including a charge, dipole, polarizable dipole and quadrupole. Each test case involved a 3

Å sphere of dielectric 1 and solvent dielectric of 78.3 with a step-function transition between solute and solvent (kcal/mole). 60

Table 3.6. The tests from Table 4 are repeated using 129 grid points (0.078 Å spacing), however, the transition between solute and solvent is defined by a 7th order polynomial, which acts over a total window width of 0.6 Å. Increasing the radius of the low dielectric sphere by approximately 0.2 Å raises the energies to mimic the step function transition results (kcal/mole). 61

Table 3.7. Synopsis of the protein systems studied in explicit and continuum solvent. 68

Table 3.8. The energy (kcal/mole) and dipole moment (debye) of each protein system was studied using a range of grid spacings under the direct polarization model, mutual polarization model, and mutual polarization model with 150 mM salt. The cavity was defined using AMOEBA R_{min} values for each atom and smooth dielectric and ionic boundaries via a total window width of 0.6 Å. 69

Table 3.9. The dipole moment (debye) of each protein in vacuum μ_v, under the direct and mutual polarization models interacting with a continuum of permittivity 78.3, and in explicit water. Ensemble averages were taken over 100 psec trajectories and each has a std. err. of less than ± 0.3. The ratio of the solvated to vacuum dipole moment is given in each case. The cavity was defined using AMOEBA R_{min} values for

each atom and smooth dielectric and ionic boundaries via a total window width of 0.6 Å. 71

Table 3.10. Memory requirements and wall clock timings for each protein system are shown. All calculations were run on a 2.4 Ghz Opteron. 72

Table 4.1. Multipole moment conversions. 80

Table 4.2. Unit vacuum potentials. 81

Table 4.3. Unit vacuum fields. 82

Table 4.4. Selected scalar products of unit magnitude vacuum spherical harmonic fields. 83

Table 4.5. Indefinite integrals for the pairwise descreening of multipoles. 84

Table 4.6. Indefinite integrals for the pairwise descreening of multipoles when $\xi_{ij} = \pi$. 85

Table 4.7. Shown is a comparison of the performance of the SFA and RPA in determining the perfect self-energy (kcal/mole) for a series of five folded proteins. Optimization of a single HCT scale factor for each method removes systematic error as shown by the mean signed percent differences. However, the mean RPA unsigned percent difference of 0.5 is smaller than that of the SFA. 90

Table 4.8. The electrostatic solvation free energy (kcal/mole) for 55 proteins within the PMPB and GK continuum models. The number of atoms and total charge of each protein is listed along with the signed and unsigned relative difference of the GK model to PMPB. 120

Table 4.9. The total dipole moment (Debye) for 55 proteins in vacuum and within the PMPB and GK continuum models are presented. The signed and unsigned percent error of the GK model relative to PMPB is given along with the reaction field factor under both models. 122

Table 5.1. The solvent assessable surface area (SASA) and solvent excluded volume (SEV) for the 39 small molecules used to parameterize PMPB and GK based implicit solvents. The solvent assessable surface area (SASA) and solvent excluded volume (SEV) were defined used AMOEBA R_{min} values and solvent probe radius of 1.4 Å. 125

Table 5.2. Calculated surface tension and solvent pressure are used to determine self-consistent cavitation free energies. The computed standard errors on the ST were all below 0.001 for the ST measurements and below 0.0005 for the SP. 132

Table 5.3. The average solute-solvent enthalpy was calculated from two sets of explicit water simulations as described in the text. Taking their difference gives an estimate for the dispersion free energy. The value of the implicit solvent dispersion term is shown in the 4th column, along with its error relative to the explicit water estimate. All values are in kcal/mol. 139

Table 5.4. Solvation free energy of AMOEBA solutes in both PMPB and GK based implicit solvents compared to experiment. The PMPB and GK values include the same apolar term. All values are in kcal/mol..... 145

Table 6.1. Closed form expressions for the pairwise electrostatic solvation free energies between two off-center multipole components within a sphere of radius *a* up to quadrupole order are given. The vectors r_1 and r_2 are relative to the center of the sphere. When $r_1 = r_2$ the formulas are reduced to self-energies, which are given in Table 24. Kong and Ponder have previously reported infinite series solutions in terms of Legendre polynomials in Appendix B of their work.[28] The convention for repeated summation over Greek subscripts is assumed and $\hat{\mathbf{r}}$ is a unit vector in the direction r. 149

Table 6.2. Here we present closed form expressions for the self-energy for two off-center multipole components at the same site within a spherical solute of radius *a*. As the multipole approaches the center of the sphere $r \to 0$, the formulas simplify to well-known solutions. Kong and Ponder have previously reported infinite series solutions in Appendix B of their work.[28] The convention for repeated summation over Greek subscripts is assumed and $\hat{\mathbf{r}}$ is a unit vector in the direction r. 151

Table 7.1. Gradients of $A^{(0)}_{\{0,0,0\}}$. 166

Table 7.2. Gradients of $A^{(1)}_{1,0,0}$. 167

Table 7.3. Gradients of $A^{(1)}_{\{0,1,0\}}$ 168

Table 7.4. Gradients of $A^{(1)}_{\{0,0,1\}}$ 169

Table 7.5. Gradients of $A^{(2)}_{\{2,0,0\}}$ 170

Table 7.6. Gradients of $A^{(2)}_{\{1,1,0\}}$ 171

Table 7.7. Gradients of $A^{(2)}_{\{1,0,1\}}$ 172

Table 7.8. Gradients of $A^{(2)}_{\{0,2,0\}}$ 173

Table 7.9. Gradients of $A^{(2)}_{\{0,1,1\}}$ 174

Table 7.10. Gradients of $A^{(2)}_{\{0,0,2\}}$ 175

List of Figures

Figure 1.1. This diagram shows the thermodynamic cycle used to motivate the terms of an implicit solvent model. 5

Figure 1.2. This diagram is intended to show the evolution of numeric Poisson-Boltzmann solvers based on classical force fields toward accurate treatment of large biomolecular systems. 8

Figure 1.3. This diagram presents a brief history of analytic continuum electrostatics. 11

Figure 3.1. Normalized 5th order B-spline on the interval [0, 5]. 30

Figure 3.2. The sum of two 4th order B-splines (dashed) are equal to the first derivative of a normalized 5th order B-spline (solid)..... 32

Figure 3.3. The sum of three 3rd order B-splines (dashed) equal the second derivative of a normalized 5th order B-spline (solid)..... 33

Figure 3.4. Comparison of cubic, quintic and heptic characteristic functions for an atom with radius 3 Å using a total window width of 0.6 Å..... 39

Figure 3.5. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with unit positive charge vs. separation. Both spheres have a radius of 3.0 Å and the solvent dielectric is 78.3. The

gradient of the neutral cavity is due entirely to the dielectric boundary force and cancels exactly the force on the charged sphere. 62

Figure 3.6. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with dipole moment components of (2.54, 2.54, 2.54) debye vs. separation. Both spheres have a radius of 3.0 Å and movement of the dipole is along the x-axis. The gradient of the neutral cavity is due entirely to the dielectric boundary force and cancels exactly the sum of the forces on the dipole and a third site (that has no charge density or dielectric properties) that defines the local coordinate system of the dipole. 63

Figure 3.7. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with quadrupole moment components of (5.38, 2.69, 2.69, 2.69, -2.69, 2.69, 2.69, 2.69, -2.69) Buckinghams vs. separation. Both spheres have a radius of 3.0 Å and movement of the quadrupole is along the x-axis. The gradient of the neutral cavity cancels exactly the sum of the forces on the quadrupole and a third site (that has no charge density or dielectric properties) that defines the local coordinate system of the quadrupole. 64

Figure 3.8. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin and a sphere with unit positive charge vs. separation using the *direct* polarization model. Both spheres have a radius of 3.0 Å. The gradient can be seen to approach zero at a

number of points, notably when the spheres are separated by approximately 1.5 Å leading to a maximum in the reaction field produced by the charge at the polarizable site, and again when the spheres are superimposed and the reaction field is zero at the polarizable site. 65

Figure 3.9. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin and a polarizable sphere with unit positive charge vs. separation using the *mutual* polarization model. Both spheres have a radius of 3.0 Å and a polarizability of 1.0 $Å^{-3}$. Note that the mutual polarization gradients are smaller than those in Fig 8. for the otherwise equivalent *direct* polarization model. 66

Figure 3.10. The dielectric of the solvent and test spheres are both set to 1 in this case, while a salt concentration of 150 mM is used to isolate the ionic boundary gradients. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin (3.0 Å radius) and a polarizable sphere with a unit positive charge (1.0 Å radius) vs. separation using the mutual polarization model. Both spheres have a polarizability of 1.0 $Å^{-3}$, and the ionic radius is set to 0.0 Å. 67

Figure 4.1. The solvation energy for a system composed two spheres, each with a radius of 3 Å and permittivity of 1, and a variety of multipole combinations are computed as a function of separation along the x-axis using numerical Poisson solutions (solid lines) and generalized

Kirkwood (dashed lines). The solvent permittivity was 78.3. The limiting cases of wide separation and superimposition are exact in all cases, while intermediate separations are seen to be a reasonable approximation. 97

Figure 5.1. Cavitation free energy for AMOEBA small molecules via SP. 136

Figure 5.2. Cavitation free energy for AMOEBA small molecules via ST. 137

Figure 5.3. A comparison of the analytic continuum dispersion free energy with results from explicit water simulations show good agreement over a range of small molecule sizes. 144

Acknowledgements

First I would like to thank my advisor, Jay Ponder, for his guidance and support during completion of this work. Jay's commitment to developing a molecular mechanics force field with chemical accuracy called AMOEBA will have a profound impact on the quality of biomolecular simulations.

I am indebted to Alan Grossfield and Pengyu Ren who were post-doctoral fellows in the Ponder lab at the beginning of my time in graduate school. Both were always ready and willing to be of service. Specifically, the complexity of the interaction of the AMOEBA model with a continuum solvent would have been beyond my grasp without lots of tutoring from Alan and Pengyu. I wish them continued success in the future. More recently, Sergio Urahata, Chuanji Wu and Justin Xiang have joined the Ponder lab and immediately became encouraging colleagues and friends.

Our collaboration with Nathan Baker and his lab has been productive and rewarding. I would like to express my thanks to Nathan and Todd Dolinksy for their help in integrating polarizable multipole methods into the Baker lab Adaptive Poisson-Boltzmann Solver (APBS). I hope for continued interaction in the future. I also appreciate Nathan's role as a member of my thesis committee. David Gohara, although not a member of the Baker lab, has also been very generous in lending his time and expertise to further development of both Jay's TINKER package and APBS.

The Department of Biomedical Engineering (BME) was experiencing rapid growth as I began graduate school, thanks in part to funding from the Whitaker foundation. I appreciate the guidance of my advisor David Sept and many thoughtful discussions with Rohit Pappu. Their interest in atomic resolution modeling helps to legitimate it in the context of biomedical engineering and I appreciate their feedback as thesis committee members. I also thank Radhakrishna Sureshkumar of the Department of Chemical Engineering for being a thesis committee member and hope the resulting work may be of some use with regard to his interest in protein adsorption.

For funding I express thanks to the BME department, to the NIH for a Computational Biology Training Grant awarded to Washington University and acknowledge a Grace Norman scholarship.

Prior to living in St. Louis I had spent my entire life in and around the small town of West Branch, Iowa, the birthplace of the 31st President of the United States, Herbert Hoover. My parents Jerome and Susan settled outside West Branch after being educated nearby at the University of Iowa, which has become a family tradition. Both have served the community for decades, dad at the National Park Service working to preserve buildings from Hoover's time and mom as a Speech/Language Pathologist with the Grant Wood Area Education Agency. The love, support and values of my family are my greatest treasures.

During my undergraduate years studying biomedical engineering at the University of Iowa, I was fortunate to work in three well-established labs and remain grateful for these formative research experiences. Specifically, I would like to thank Kenneth Moore,

Randy Nessler, Tom Moninger, Kathy Walters and Jean Ross from the Central Microscopy Research Facility; Joseph Buckwalter, James Martin, Jeff Stevens and Louis Lembke of the Ignacio V. Ponseti Biochemistry and Cell Biology Laboratory; and finally Thomas Brown, Douglas Pedersen and Anneliesa Heiner from the Orthopaedic Biomechanics Laboratory. Currently all scientists mentioned above are still active in their respective appointments, most a decade after I met them, which is outstanding.

Michael J. Schnieders

Washington University in St. Louis

December 2007

1 Introduction

The solvent environment influences the structure and behavior of biomolecules within it. For example, the scaling of the radius of gyration of a polymer with chain length in dilute aqueous solution can be predicted by considering whether solvent molecules prefer interactions with themselves to those with the polymer.[1] This scaling law, which describes whether or not a polymer adopts a compact fold, serves to emphasize that rigorous a result can be obtained without treating solvent in explicit atomic detail.

In this work we present numerical and analytic models of the electrostatic interactions between a biomolecule represented by a polarizable atomic multipole force field and a continuum environment characterized by its permittivity, dispensing with the expense of representing explicit solvent molecules. Also presented is a novel formulation of the apolar contribution to solvation, which when combined with either the numerical or analytic continuum electrostatic model forms a complete implicit solvent.

The remainder of this introduction presents the concept of an implicit solvent from the perspective of statistical mechanics and subsequently based on a thermodynamics cycle. It is emphasized why a force field combined with an implicit solvent is a useful computational tool for answering biomolecular questions relevant to the biomedical engineers.

1.1 The Theory of Biomolecular Solvation

Although the theory of biomolecular solvation is a broad topic, for our purposes it will be defined as theories and models that facilitate prediction of solute behavior within a solvent. The theories that will be presented have relevance to many solvents and to a wide variety of systems, however, biomolecules in water will be our focus because of their importance to biomedical engineering. Approaches to capturing the effect of solvent on a solute can be categorized based on the amount of solvent represented explicitly:

1. **Only explicit water**: A periodic box where the solvent and solute molecules are allowed to exit one side of the box and reenter the opposite side.
2. **Explicit water within an implicit solvent**: The spherical solvent boundary potential (SSBP) solvates a solute within a sphere of explicit water molecules surrounded by a continuum.
3. **No explicit water**: A purely implicit solvent.

In general, the more explicit water that is used, the more expensive it is to evaluate the underlying energy function. This motivates the development of the SSBP and implicit solvent approaches.

Parameterization of an explicit water model for the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field was completed prior to the beginning of this work.[2, 3] This allows use of explicit water simulations of individual small molecule solutes in order to collect data that can then be used to parameterize models with implicit components. This approach is taken in Chapter 5 (p. 125). We also present initial work on developing an SSBP for AMOEBA in Chapter 6 (p. 147). However, the major contributions of this work are two continuum electrostatics theories for AMOEBA. These models form the basis for purely implicit models of solvation.

It is important to emphasize that an accurate implicit solvation model does not necessitate any loss of solute thermodynamic information compared to explicit representation of solvent degrees of freedom.[4] This can be demonstrated using statistical mechanics beginning with the probability distribution $\mathrm{P}(\mathbf{X},\mathbf{Y})$ for an explicit solvent simulation

$$\mathrm{P}(\mathbf{X},\mathbf{Y}) = \frac{e^{-\mathrm{U}(\mathbf{X},\mathbf{Y})/kT}}{\int e^{-\mathrm{U}(\mathbf{X},\mathbf{Y})/kT} d\mathbf{X} d\mathbf{Y}}. \tag{1.1.1}$$

Here $\mathbf{X}$ are the solute coordinates, $\mathbf{Y}$ are the solvent coordinates, k is Boltzmann's constant, T the absolute temperature, $\mathrm{U}(\mathbf{X},\mathbf{Y})$ is the potential energy of the system and the integral in the denominator is referred to as the partition function. By integrating out

the solvent degrees of freedom, a reduced probability distribution $\overline{\mathrm{P}}(\mathbf{X})$ that depends on only the solute coordinates can be defined

$$\begin{aligned}\overline{\mathrm{P}}(\mathbf{X}) &= \int \mathrm{P}(\mathbf{X},\mathbf{Y})\,d\mathbf{Y} \\ &= \frac{e^{-\mathrm{W}(\mathbf{X})_{\mathrm{PMF}}/kT}}{\int e^{-\mathrm{W}(\mathbf{X})_{\mathrm{PMF}}/kT}\,d\mathbf{X}}\end{aligned} \tag{1.1.2}$$

where $W(\mathbf{X})_{\mathrm{PMF}}$ is a potential of mean force (PMF) given by the sum of the solute's potential energy in vacuum $\mathrm{U}(\mathbf{X})^{\mathrm{v}}$ and the hydration energy of a rigid solute conformation

$$\mathrm{W}(\mathbf{X})_{\mathrm{PMF}} = \mathrm{U}(\mathbf{X})^{\mathrm{v}} + \Delta\,\mathrm{W}(\mathbf{X})_{\mathrm{hydration}}\,. \tag{1.1.3}$$

In this work, the vacuum potential is calculated using the recently developed AMOEBA force field. The main deliverable of this thesis is the description of two hydration free energy functions $\Delta\,\mathrm{W}(\mathbf{X})_{\mathrm{hydration}}$ that are consistent with AMOEBA, one based on numerical solutions of the linearized Poisson-Boltzmann equation (LPBE) and a second analytic model called Generalized Kirkwood (GK). Both the LPBE and GK based implicit solvent models will be explained in detail and compared.

Although an implicit solvent can be defined via statistical mechanics, a thermodynamic cycle like that shown in Figure 1.1 is useful for motivating an efficient and convenient functional form. The free energy change to move a solute from vacuum (upper left) into solvent (lower left) is path independent, such that infinitely many routes are possible. We will initially describe the path advocated in this work at a qualitative

level to introduce central concepts. Further quantitative detail and more rigorous justifications are presented later.

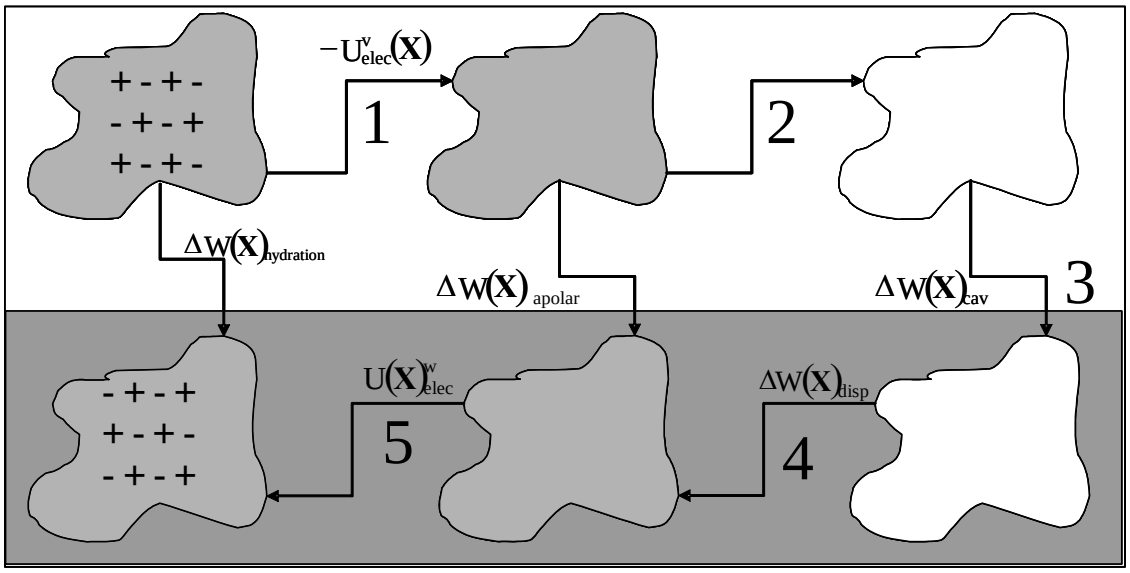

Figure 1.1. This diagram shows the thermodynamic cycle used to motivate the terms of an implicit solvent model.

Beginning in the upper left corner of Figure 1.1 and moving clockwise around the diagram, the steps in the thermodynamic cycle include:

1. Turning off the solute electrostatics
2. Turning off solute-solvent dispersion interactions
3. Forming a solute-shaped cavity in solvent
4. Restoring solute-solvent dispersion interactions
5. Turning on the solute electrostatics

The sum of these five steps gives the solvation free energy for a rigid solute conformation

$$\Delta W(\mathbf{X})_{hydration} = \Delta W(\mathbf{X})_{cav} + \Delta W(\mathbf{X})_{disp} + U(\mathbf{X})^{w}_{elec} - U(\mathbf{X})^{v}_{elec} \qquad (1.1.4)$$

For a solute in vacuum, the second step entails no energetic change, although transfer between two solvents typically would. As suggested above, alternative paths may be taken, including the combination of steps 2-4 into a single apolar term.

1.1.1 Numerical Continuum Electrostatics

Modeling the change in the electrostatic moments of organic molecules upon moving from vacuum to solvent has a long history, with an important initial contribution from Onsager, who in 1936 identified the difference between the cavity field and reaction field.[5] The approach used was to treat the solvent as a high dielectric continuum surrounding a spherical, low dielectric solute with a dipole moment, which was considered to be a sum of permanent and induced contributions. Using the vacuum dipole moment, molecular polarizability, and an estimate of molecular size, a prediction of the experimentally observable liquid permittivity was achieved for a range of molecules. Through the use of computers, this approach has been extended in order to treat solutes with arbitrary geometry and charge distributions by numerically solving the Poisson-Boltzmann equation using finite-difference, finite element or boundary element methods.[6] An advantage of using a continuum solvent over explicit representation of solvent molecules is alleviation of the need to sample over water degrees of freedom in order to determine the mean solvent response. Applications that have benefited from using continuum solvent approaches include predictions of pK_as, redox potentials, binding energies, molecular design, and conformational preferences.

This work concentrates on the electrostatic contribution to solvation, motivated by recent work on improving the accuracy of force field electrostatic models through the incorporation of polarizable multipoles, although novel contributions to the apolar model will also be presented in Chapter 5 (p. 125).[2, 3, 7-10] A consistent interaction between an AMOEBA solute and continuum solvent requires revisiting the theory underlying the electrostatic component of implicit solvent models, including those based on solving the linearized Poisson-Boltzmann equation (LPBE)

$$\nabla \bullet \left[\varepsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) \right] - \bar{\kappa}^2(\mathbf{r}) \Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \tag{1.1.5}$$

where the coefficients are a function of position $\mathbf{r}$, $\Phi(\mathbf{r})$ is the potential, $\varepsilon(\mathbf{r})$ the permittivity, $\bar{\kappa}^2(\mathbf{r})$ the modified Debye-Hückel screening factor and $\rho(\mathbf{r})$ is the solute charge density.

Shown below in Figure 1.2 is a diagram that is intended to present a high level overview of the evolution of numerical solutions to the PB equation using charge distributions from classical force fields. This approach began with the pioneering work of Warwicker and Watson in 1983 and was based on a fixed charge force field using a single CPU.[11] In 2001, the parallel focusing technique introduced by Baker *et al*. dramatically increased tractable system size to millions of atoms using massively parallel computing.[12] Here focusing denotes the use of a coarse solution to the PB equation to define the boundary conditions of a smaller domain, which is subdivided among available processors using a spatial decomposition. The advance described in this dissertation pushes the envelop of numerical continuum electrostatics technology for biomolecular

systems toward higher accuracy by using a charge distribution based on polarizable multipoles rather than fixed point charges.[13] In the future, it should be possible to apply parallel focusing to the Polarizable Multipole Poisson-Boltzmann (PMPB) model in order to study the cooperative electrostatics of large biomolecular assemblies. However, this is a nontrivial next step that will require significant effort.

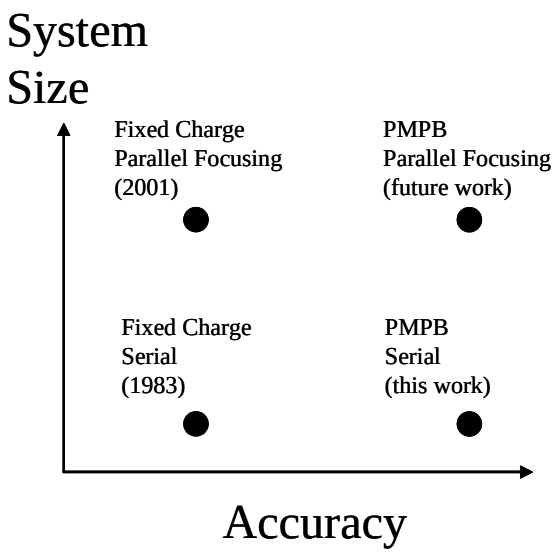

Figure 1.2. This diagram is intended to show the evolution of numeric Poisson-Boltzmann solvers based on classical force fields toward accurate treatment of large biomolecular systems.

Further introduction to Poisson-Boltzmann based methodology is given in Chapter 2 (p. 10), but we also recommend the review of Honig and Nichols[14] or those of Baker.[6, 15] Our novel PMPB electrostatics model follows in Chapter 3 (p. 26). Alternatively, an analytic continuum electrostatic model called GK that is similar in spirit to generalized Born (GB), but is capable of treating polarizable atomic multipoles, is presented in Chapter 4 (p. 73).

Along with the AMOEBA force field, other efforts toward developing polarizable force fields for biomolecular modeling are also under way, for example see the review of Ponder and Case.[8] Here we comment more thoroughly on the Polarizable Force Field (PFF) of Maple and coworkers, since it has recently been incorporated into a continuum environment.[16-18] Specifically, there are a number of salient differences between AMOEBA and PFF, including facets of the underlying polarization model and the use of permanent quadrupoles in AMOEBA. Significantly, AMOEBA allows mutual polarization between atoms with 1-2, 1-3 and 1-4 bonding arrangements, which is crucial for reproducing molecular polarizabilities.

The current work addresses a number of issues raised in the description of the PFF solvation model. First, discretization procedures previously reported for mapping partial charges onto a source grid are inadequate for higher order moments. Instead, we present a multipole discretization procedure using B-splines that leads to essentially exact energy gradients. Furthermore, we provide a rigorous demonstration of the numerical precision of our approach, similar in spirit to the work of Im *et al.* with respect to partial charge models.[19] We also show that divergence of the polarization energy is not possible due to use in AMOEBA of Thole-style damping at short range.[20]

Formulation of consistent energies and gradients based on the LPBE has been reported previously for partial charge force fields.[19, 21-25] Gilson *et al.* pointed out limitations in previous approaches using variational differentiation of an electrostatic free energy density functional.[22] Later Im *et al.* showed that it is possible to begin the derivation based upon the underlying finite-difference calculation used to solve the LPBE.[19] This approach leads to a formulation with optimal numerical consistency and will be adopted here.

1.1.2 Analytic Continuum Electrostatics

Our approach to analytic continuum electrostatics can be traced to work presented by Born in 1920 to describe the electrostatic solvation energy of a charged, spherical ion in terms of macroscopic continuum theory.[26] In 1934, Kirkwood extended this approach to a spherical particle with arbitrary electrostatic multipole moments with application to the study of zwitterions, which have a large dipole moment.[27] More recently, Kong and Ponder revisited Kirkwood's theory to allow analytic treatment of off-center point multipoles.[28] For a single spherical particle in isolation, therefore, the theoretical foundations to enable use of macroscopic continuum theory have already been established.

However, a general analytic solution to the Poisson equation for an arbitrarily spaced collection of spherical dielectric particles embedded in solvent is tenable only via approximations. For example, the generalization of Born's method to a collection of

monopoles began to be considered in the 1990's by a number of groups including Schaefer *et al*.[29-31], Hawkins *et al*.[32, 33], Still *et al.*[34-36], Feig *et al.*[37-40] and Onufriev *et al.*[41-44]. This GB approach is intended to approximate the numerical solution of the Poisson equation for realistic molecular geometries and monopole charge distributions. Given highly accurate self-energies, GB has been shown to be remarkably quantitative.[37, 41, 42, 45] A goal of the present work is to extend the ideas underlying GB to more accurate charge distributions, specifically to the treatment of polarizable atomic multipoles, which is termed Generalized Kirkwood, or GK, by analogy.[46] This progression in model complexity is illustrated by Figure 1.3 below.

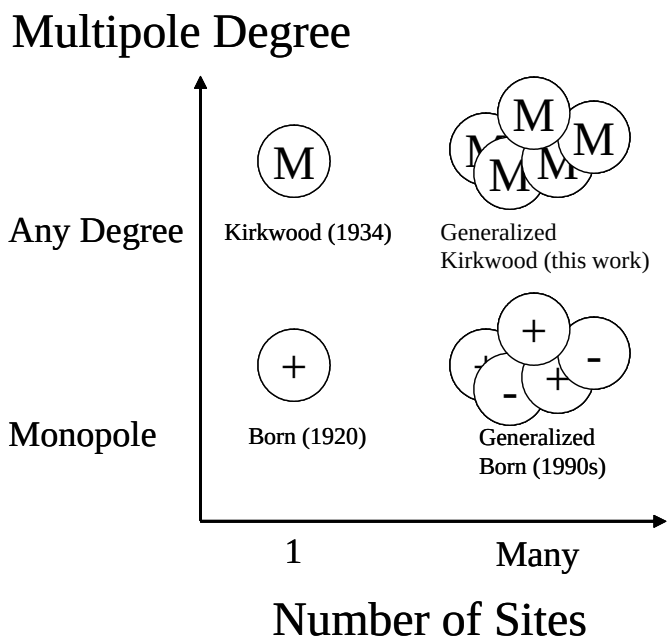

Figure 1.3. This diagram presents a brief history of analytic continuum electrostatics.

In order to further motivate the present work, we recall the electrostatic solvation energy is a key component of an implicit solvent model, which typically also includes apolar contributions due to cavitation and dispersion.[4, 47, 48] Given a solute potential and implicit solvent, a broad range of physical properties can be predicted, including conformational preferences such as radius of gyration, binding energies and pK_as.[38] Recent work by a number of groups to explicitly include higher order permanent moments and polarization within the functional form of empirical force field electrostatics may improve the quality of theoretical predictions based on implicit solvent approaches.[8, 16, 49-53] However, this step forward can only be realized if the improved detail of the molecular mechanics electrostatic model is propagated through to the reaction potential.

For an excellent introduction to the fundamentals of GB theory, including treatment of salt effects, we recommend the review by Bashford and Case.[54] Feig and Brooks present a review of recent improvements in GB methodology as well as novel applications.[38] Assuming this level of familiarity, we outline the key components of GB that need to be further generalized in Chapter 4 in order to incorporate polarizable atomic multipoles.

1.2 The Effect of Solvation on Biomolecules

The relevance of polarization to the effect of solvation on biomolecules is suggested by the fact that the dipole moment of a polar solute can increase by 30% or more during transfer from gas to aqueous phase. However, empirical potentials have typically neglected explicit treatment of polarization for reasons of computational efficiency. On the other hand, *ab initio* implicit solvent models, including the Polarizable Continuum Model (PCM) introduced in 1981 by Miertus, Scrocco and Tomasi[55-58], the Conductor-Like Screening Model (COSMO) of Klamt[59, 60], the distributed multipole approach of Rinaldi *et al.*[61, 62] and the SM*x* series of models introduced in the early 90s by Cramer and Truhlar[63-72] have long incorporated self-consistent reaction fields (SCRF). These models allow the solute, described using a range of *ab initio* or semi-empirical levels of quantum theory, and continuum solvent to relax self-consistently based on their mutual interaction. The principle advantage of the present PMPB model over these existing formulations is computational savings resulting from a purely classical representation of the solute Hamiltonian, which facilitates the study of large biomolecular systems. However, each of these models demonstrates that there continues to be broad interest in coupling highly accurate solute potentials with a continuum treatment of the environment.

For example, there is growing evidence that current goals of computational protein design, including incorporation of catalytic activity and protein-protein recognition, may require a more accurate description of electrostatics than has been achieved by fixed partial charge force fields used in conjunction with implicit solvents.[51] For example, it has been shown that both Poisson-Boltzmann and Generalized Born models used with the CHARMM22[73] potential tend to favor burial of polar residues over non-polar ones.[50] It is important to note that this behavior may not be directly due to the treatment of solvation electrostatics, but could result from inaccuracies in the underlying protein force field or the apolar component of the solvation model. However, the fixed charge nature of traditional protein potentials may also contribute to such discrepancies. Polar residues elicit a solvent field that increases their dipole moment via polarization, relative to an apolar environment, which has a favorable energetic consequence that cannot be captured by a fixed charge force field even when using explicit water.

2 Theoretical Background

In order to describe the total electrostatic energy of the PMPB and GK models, specifically of an AMOEBA solute in a LPBE or GK continuum solvent, it is first necessary to present the energy in vacuum using a convenient notation. We will also summarize the previously developed procedure for determining robust energies and gradients for fixed charge force fields in a LPBE continuum.[19, 22] Given this background, the theoretical infrastructure required to implement the PMPB and GK models can be motivated and presented in a self-contained fashion.

2.1 AMOEBA Vacuum Electrostatic Energy

Following Ren and Ponder[3], each permanent atomic multipole site can be considered as a vector of coefficients including charge, dipole and quadrupole components

$$\mathbf{M}_i = \left[q_i, d_{i,x}, d_{i,y}, d_{i,z}, \Theta_{i,xx}, \Theta_{i,xy}, \Theta_{i,xz}, ..., \Theta_{i,zz} \right]^t, \tag{2.1.1}$$

where the superscript t denotes the transpose. The interaction energy between two sites i and j separated by distance s_{ij} can then be represented in tensor notation as

$$
\begin{aligned}
\mathrm{U}\left(\mathbf{s}_{ij}\right) &= \mathbf{M}_i^t\,\mathbf{T}_{ij}\,\mathbf{M}_j \\
&= \begin{bmatrix} q_i \\ d_{i,x} \\ d_{i,y} \\ d_{i,z} \\ \Theta_{i,xx} \\ \vdots \end{bmatrix}^t
\begin{bmatrix}
1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\
\frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\
\frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\
\frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\frac{1}{s_{ij}}
\begin{bmatrix} q_j \\ d_{j,x} \\ d_{j,y} \\ d_{j,z} \\ \Theta_{j,xx} \\ \vdots \end{bmatrix}.
\end{aligned}
\tag{2.1.2}
$$

Each site may also be polarizable, such that an induced dipole $\boldsymbol{\mu}_i$ proportional to the strength of the local field is present

$$
\begin{aligned}
\boldsymbol{\mu}_i &= \alpha_i \mathbf{E}_i \\
&= \alpha_i \left(\sum_{j \neq i} \mathbf{T}_{\mathrm{d},ij}^{(1)} \mathbf{M}_j + \sum_{k \neq i} \mathbf{T}_{ik}^{(11)} \boldsymbol{\mu}_k \right).
\end{aligned}
\tag{2.1.3}
$$

Here α_i is an isotropic atomic polarizability and $\mathbf{E}_i$ is the total field, which can be decomposed into contributions from permanent multipole sites and induced dipoles, and the summations are over $\mathrm{N_s}$ multipole sites. Later, this expression will be modified to include the solvent reaction field. The interaction tensors $\mathbf{T}_{\mathrm{d},ij}^{(1)}$ and $\mathbf{T}_{ik}^{(11)}$ are, respectively,

$$
\mathbf{T}_{\mathrm{d},ij}^{(1)} = \begin{bmatrix}
\frac{\partial}{\partial x_j} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\
\frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\
\frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots
\end{bmatrix} \frac{1}{s_{ij}}
\tag{2.1.4}
$$

and

$$\mathbf{T}_{ik}^{(11)} = \begin{bmatrix} \frac{\partial^2}{\partial x_i \partial x_k} & \frac{\partial^2}{\partial x_i \partial y_k} & \frac{\partial^2}{\partial x_i \partial z_k} \\ \frac{\partial^2}{\partial y_i \partial x_k} & \frac{\partial^2}{\partial y_i \partial y_k} & \frac{\partial^2}{\partial y_i \partial z_k} \\ \frac{\partial^2}{\partial z_i \partial x_k} & \frac{\partial^2}{\partial z_i \partial y_k} & \frac{\partial^2}{\partial z_i \partial z_k} \end{bmatrix} \frac{1}{s_{ik}} . \tag{2.1.5}$$

where the subscript d in $\mathbf{T}_{\mathrm{d},ij}^{(1)}$ indicates that masking rules for the AMOEBA group-based polarization model are applied.[2, 3, 7] This linear system of equations can be solved via a number of approaches, including direct matrix inversion or iterative schemes such as successive over-relaxation (SOR). Note at short range the field is damped via the Thole model, which is not included above for clarity and is discussed elsewhere.[3] The total vacuum electrostatic energy $\mathrm{U}_{\mathrm{elec}}^{\mathrm{v}}$ includes pairwise permanent multipole interactions and many-body polarization

$$\mathrm{U}_{\mathrm{elec}}^{\mathrm{v}} = \frac{1}{2}\left[\mathbf{M}^t \mathbf{T} - \left(\boldsymbol{\mu}^{\mathrm{v}} \right)^t \mathbf{T}_{\mathrm{p}}^{(1)} \right] \mathbf{M} , \tag{2.1.6}$$

where the factor of one-half avoids double-counting of permanent multipole interactions in the first term and accounts for the cost of polarizing the system in the second term. Furthermore, $\mathbf{M}$ is a column vector of $13\mathrm{N_s}$ multipole components

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_{\mathrm{N_s}} \end{bmatrix} , \tag{2.1.7}$$

$\mathbf{T}$ is a N_s x N_s supermatrix with $\mathbf{T}_{ij}$ as the off-diagonal elements

$$\mathbf{T} = \begin{bmatrix} 0 & \mathbf{T}_{12} & \mathbf{T}_{13} & \cdots \\ \mathbf{T}_{21} & 0 & \mathbf{T}_{23} & \cdots \\ \mathbf{T}_{31} & \mathbf{T}_{32} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2.1.8}$$

$\boldsymbol{\mu}^{\mathrm{v}}$ is a $3N_s$ column vector of converged induced dipole components in vacuum

$$\boldsymbol{\mu}^{\mathrm{v}} = \begin{bmatrix} \mu_{1,x} \\ \mu_{1,y} \\ \mu_{1,z} \\ \vdots \\ \mu_{N_s,z} \end{bmatrix}, \tag{2.1.9}$$

and $\mathbf{T}_{\mathrm{p}}^{(1)}$ is a $3N_s$ x $13N_s$ supermatrix with $\mathbf{T}_{\mathrm{p},ij}^{(1)}$ as off-diagonal elements

$$\mathbf{T}_{\mathrm{p}}^{(1)} = \begin{bmatrix} 0 & \mathbf{T}_{\mathrm{p},12}^{(1)} & \mathbf{T}_{\mathrm{p},13}^{(1)} & \cdots \\ \mathbf{T}_{\mathrm{p},21}^{(1)} & 0 & \mathbf{T}_{\mathrm{p},23}^{(1)} & \cdots \\ \mathbf{T}_{\mathrm{p},31}^{(1)} & \mathbf{T}_{\mathrm{p},32}^{(1)} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.1.10}$$

The subscript p denotes a tensor matrix that operates on the permanent multipoles to produce the electric field in which the polarization energy is evaluated, while the subscript d was used above to specify an analogous tensor matrix that produces the field that induces dipoles. The differences between the two are masking rules that scale short-range through-bond interactions in the former case and use the AMOEBA group-based polarization scheme for the later.[3, 7]

2.2 Fixed Charge Linearized Poisson-Boltzmann Energy and Gradient

LPBE solvation energies and gradients have been determined previously by a number of groups for fixed partial charge force fields.[21-24] We will briefly restate the results of Im *et al.* to introduce the approach extended here for use with the PMPB model.[19] The solvation free energy ΔG of a permanent charge distribution is

$$\Delta G_{\text{elec}} = \frac{1}{2}\left(\boldsymbol{\Phi}_{\text{s}}^{t} - \boldsymbol{\Phi}_{\text{v}}^{t}\right)\boldsymbol{q}\,, \tag{2.2.1}$$

where $\boldsymbol{q}$ is a column vector of fractional charges, $\boldsymbol{\Phi}_{\text{s}}^{t}$ is the transpose of a column vector containing the electrostatic potential of the solvated system and $\boldsymbol{\Phi}_{\text{v}}^{t}$ is the corresponding vacuum potential. The number of components in each vector is equal to the number of grid points used to represent the system. In the present work the grid will always be cubic, and as discussed in the section on multipole discretization *must* use equal grid spacing in each dimension, although the number of grid points along each axis can vary. The potential can be determined numerically using a finite-difference representation of the LPBE[11, 19, 74-76], which can be formally defined as a linear system of equations

$$\mathbf{A}\,\boldsymbol{\Phi} = -4\pi\boldsymbol{q}\,, \tag{2.2.2}$$

where $\mathbf{A}$ is a symmetric matrix that represents the linear operator (differential and linear term). An equivalent, but more cumbersome representation that makes clear the

underlying finite-difference formalism is given in the Appendix A (p. 156). Solving Eq. (2.2.2) for the potential

$$\boldsymbol{\Phi} = -4\pi \mathbf{A}^{-1}\boldsymbol{q} \tag{2.2.3}$$

highlights that $\mathbf{A}^{-1}$ is the Green's function with dimensions N_r x N_r, where N_r is the number of grid points. By defining the Green's function for the solvated $\mathbf{A}_s^{-1}$ and $\mathbf{A}_v^{-1}$ homogeneous cases, the electrostatic hydration free energy is

$$\Delta G_{elec} = \frac{1}{2}\left(-4\pi \boldsymbol{q}^t\right)\left(\mathbf{A}_s^{-1} - \mathbf{A}_v^{-1}\right)\boldsymbol{q} \tag{2.2.4}$$

The derivative with respect to movement of the γ coordinate of atom j is

$$\begin{aligned}\frac{\partial \Delta G_{elec}}{\partial s_{j,\gamma}} = -2\pi \Bigg[& \frac{\partial \boldsymbol{q}^t}{\partial s_{j,\gamma}}\left(\mathbf{A}_s^{-1} - \mathbf{A}_v^{-1}\right)\boldsymbol{q} + \boldsymbol{q}^t \frac{\partial\left(\mathbf{A}_s^{-1} - \mathbf{A}_v^{-1}\right)}{\partial s_{j,\gamma}}\boldsymbol{q} \\ & + \boldsymbol{q}^t\left(\mathbf{A}_s^{-1} - \mathbf{A}_v^{-1}\right)\frac{\partial \boldsymbol{q}}{\partial s_{j,\gamma}}\Bigg]\end{aligned} . \tag{2.2.5}$$

This expression can be simplified by noting that the derivative of the homogeneous Green's function is zero everywhere because the permittivity is constant and there is no salt concentration

$$\frac{\partial \mathbf{A}_v^{-1}}{\partial s_{j,\gamma}} = 0 , \tag{2.2.6}$$

and the derivative of the solvated Green's function can be substituted for using a standard relationship of matrix algebra.

$$\begin{aligned} \mathbf{A}_{\mathrm{s}}^{-1}\mathbf{A}_{\mathrm{s}} &= \mathbf{I} \\ \frac{\partial \mathbf{A}_{\mathrm{s}}^{-1}}{\partial \mathrm{s}_{j,\gamma}}\mathbf{A}_{\mathrm{s}} + \mathbf{A}_{\mathrm{s}}^{-1}\frac{\partial \mathbf{A}_{\mathrm{s}}}{\partial \mathrm{s}_{j,\gamma}} &= 0 \\ \frac{\partial \mathbf{A}_{\mathrm{s}}^{-1}}{\partial \mathrm{s}_{j,\gamma}} &= -\mathbf{A}_{\mathrm{s}}^{-1}\frac{\partial \mathbf{A}_{\mathrm{s}}}{\partial \mathrm{s}_{j,\gamma}}\mathbf{A}_{\mathrm{s}}^{-1} \end{aligned} \tag{2.2.7}$$

Finally, due to the symmetry of the Green's function, the first and third terms of Eq. (2.2.5) are equivalent

$$\frac{\partial \boldsymbol{q}^{t}}{\partial s_{j,\gamma}}\left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1}\right)\boldsymbol{q} = \boldsymbol{q}^{t}\left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1}\right)\frac{\partial \boldsymbol{q}}{\partial s_{j,\gamma}}. \tag{2.2.8}$$

Using the relationships in Eqs. (2.2.6) through (2.2.8), Eq. (2.2.5) becomes

$$\frac{\partial \Delta \mathrm{G}_{\mathrm{elec}}}{\partial s_{j,\gamma}} = -4\pi \boldsymbol{q}^{t}\left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1}\right)\frac{\partial \boldsymbol{q}}{\partial s_{j,\gamma}} + \frac{1}{8\pi}\left(4\pi \boldsymbol{q}^{t}\mathbf{A}_{\mathrm{s}}^{-1}\right)\frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}}\left(4\pi \mathbf{A}_{\mathrm{s}}^{-1}\boldsymbol{q}\right). \tag{2.2.9}$$

Finally, the two products of Green's functions with source charges in the second term on the right-hand side can be replaced by the resulting potentials using Eq. (2.2.3) to give

$$\frac{\partial \Delta \mathrm{G}_{\mathrm{elec}}}{\partial s_{j,\gamma}} = \left(\boldsymbol{\Phi}_{\mathrm{s}} - \boldsymbol{\Phi}_{\mathrm{v}}\right)^{t}\frac{\partial \boldsymbol{q}}{\partial s_{j,\gamma}} + \frac{1}{8\pi}\boldsymbol{\Phi}_{\mathrm{s}}^{t}\frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}_{\mathrm{s}}. \tag{2.2.10}$$

In the limit of infinitesimal grid spacing and infinite grid size, it was shown that this solution is equivalent to the forces derived by Gilson *et al.*[19, 22] To generalize this result for permanent atomic multipoles, the derivative $\partial \boldsymbol{q} / \partial s_{j,\gamma}$ remains to be defined and is discussed below. Additionally, all moments except the monopole are subject to torques, which are equivalent to forces on the local multipole frame defining sites.

2.3 The Generalized Born Model

2.3.1 Effective Radii and the Self-Energy

The electrostatic solvation free energy for a single charge or multipole site of a solute with all other charges or multipoles set to zero is called that site's self energy. Definition of the "perfect" effective radius a_i for site i under the GB approximation[41] guarantees an exact self-energy. It is based on the following equality

$$a_i = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\frac{q_i^{\,2}}{\Delta \mathrm{W}_{self,i}^{Poisson}} \tag{2.3.1}$$

where the factor of ½ accounts for the cost of polarizing the continuum, q_i is a partial charge, ε_h is the permittivity of a homogeneous reference state and ε_s is the permittivity of the solvent. The self-energy $\Delta \mathrm{W}_{self,i}^{Poisson}$ can be determined to high precision numerically. In this manner, the self-energy for each fixed partial charge of a solute is mapped onto the Born equation.[26] Alternatively, an analytic solution for the self-energy in terms of an energy density is possible after making the Coulomb field approximation

$$\Delta \mathrm{W}_{self,i}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\frac{q_i^2}{4\pi}\int_{solvent}\frac{1}{r^4}dV \tag{2.3.2}$$

which is explained along with other methods in Section 4.1. Substituting for $\Delta \mathrm{W}_{self,i}^{Poisson}$ in Eq. (2.3.1) with $\Delta \mathrm{W}_{self,i}^{GB}$ from Eq. (2.3.2) and changing the limits of integration for convenience shows that each effective Born radius is[54]

$$a_i = \left(\frac{1}{r_i} - \frac{1}{4\pi}\int_{solute, r>r_i}\frac{1}{r^4}dV\right)^{-1} \tag{2.3.3}$$

where the integration over the solute does not include the region within the atomic radius r_i. A number of analytic methods have been developed for determining this integral, notably the pairwise descreening method of Hawkins, Cramer and Truhlar that we will refer to as HCT[32, 33], a method by Qiu *et al.* that assumes constant energy density within each descreening atom[36], and more recently a parameter free approach by Gallicchio *et al*.[48] Although effective radii determine the reaction potential, we note that the electrostatic solvation energy of a polarizable atomic multipole also depends on its higher order gradients.

After computing effective radii, the total self-energy of a solute within GB is

$$\Delta \mathrm{W}_{self}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\sum_i \frac{q_i^2}{a_i}. \tag{2.3.4}$$

For permanent multipoles, the self-energy of higher order components must be considered. Furthermore, if the solute is polarizable, self-consistent induced moments elicit a reaction potential that leads to an additional contribution to the electrostatic

solvation free energy. We will avoid decomposing the polarization energy into self-energy and cross-term contributions, since it is inherently many-body and therefore any partitioning is somewhat artificial.

2.3.2 Cross-term Energy

An analytic continuum electrostatics model designed to match results from the Poisson equation must also include an estimate of the pairwise cross-term energy between all multipole pairs. Given effective radii, the GB cross-term energy for fixed partial charges is given by

$$\Delta \mathrm{W}_{cross}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\sum_i \sum_{j \neq i} \frac{q_i q_j}{f} \tag{2.3.5}$$

where the empirical generalizing function f usually takes the form[35]

$$f = \sqrt{r_{ij}^2 + a_i a_j e^{-r_{ij}^2 / c_f a_i a_j}} \tag{2.3.6}$$

and r_{ij} is the distance between sites i and j and the tuning parameter c_f is chosen in the range 2-8. As r_{ij} goes to zero, the Born formula is recovered, such that the self-energy is simply a special case of the cross-term energy. Derivation of a general form for the pairwise cross-term energy between two multipole components will be presented, which is similar in spirit to GB in that the limiting cases of superimposition and wide separation for a pair of solvated multipoles are reproduced. The accuracy of the proposed interpolation at intermediate separations will be investigated via a series of tests ranging

from simple systems consisting of only two sites up to the electrostatic solvation energy and dipole moment for a series of 55 proteins.

Our tests of GK rely on the PMPB model[13] as a standard of accuracy, which has been implemented for solutes described by the AMOEBA force field and will be described in Chapter 3. Excellent agreement will be seen in the electrostatic response of proteins solvated by the PMPB continuum when compared to ensemble average explicit water simulations, indicating that at the length scale of proteins treatment of solvent as a continuum is valid. As an alternative to numerical PMPB electrostatics, the analytic GK formulation for the AMOEBA force field is orders of magnitude more efficient.

3 Polarizable Multipole Poisson-Boltzmann

Based on the AMOEBA electrostatic energy in vacuum and previous work in obtaining the energy and gradients for a solute represented by a fixed partial charge force field described above[19], we now derive the formulation needed to describe the electrostatic solvation energy within the PMPB model. First, we consider the steps necessary to express the LPBE on a grid, including discretization of the source multipoles or induced dipoles, assignment of the permittivity, assignment of the modified Debye-Hückel screening factor and estimation of the potential at the grid boundary. A variety of techniques are available to solve the algebraic system of equations that result, although this work uses an efficient multigrid approach implemented in PMG[75] and used via the APBS software package.[12] Second, given the electrostatic potential solution to the LPBE, we describe how to determine the electrostatic solvation energy and its gradient. In fact, at least four LPBE solutions are required to determine the PMPB electrostatic solvation energy, and at least six to determine energy gradients. For comparison, fixed charged models typically require at most two LPBE solutions, the vacuum and solvated states, as outlined in the previous background section, although formulations that eliminate the

self-energy exist.[77] The reasons and implications for the increased number of solutions of the LPBE required for the PMPB model will be discussed below.

3.1 Atomic Multipoles as the Source Charge Density

An important first step to expressing the LPBE in finite-difference form is discretization of an ideal point multipole onto the source charge grid. We begin by recalling that an ideal multipole arises from a Taylor series expansion of the potential in vacuum at a location $\mathbf{R}$ due to n charges near the origin, each with a magnitude and position denoted by c_i and $\mathbf{r}_i$, respectively.[78]

$$\mathrm{V}(\mathbf{R}) = \sum_{i=1}^{n} \frac{c_i}{|\mathbf{R} - \mathbf{r}_i|} \tag{3.1.1}$$

In performing the expansion, the convention for repeated summation over subscripts is utilized and we truncate after second order,

$$\mathrm{V}(\mathbf{R}) = \sum_{i=1}^{n} c_i \left[\frac{1}{R} - r_{i,\alpha} \nabla_\alpha \frac{1}{R} + \frac{1}{2} r_{i,\alpha} r_{i,\beta} \nabla_\alpha \nabla_\beta \frac{1}{R} \right], \tag{3.1.2}$$

where the α and β subscripts each denote an x-, y- or z-component of a position vector or differentiation with respect to that coordinate. Based on this expansion, the monopole q, dipole d, and traceless quadrupole Θ moments are defined as

$$
\begin{aligned}
q &= \sum_{i=1}^{n} c_i \\
d_\alpha &= \sum_{i=1}^{n} r_{i,\alpha} c_i \\
\Theta_{\alpha\beta} &= \sum_{i=1}^{n} \frac{3}{2} r_{i,\alpha} r_{i,\beta} c_i - \frac{1}{2} r_i^2 \delta_{\alpha\beta}
\end{aligned}
\quad . \tag{3.1.3}
$$

There are various ways to define the quadrupole moment because only five quadrupole components are independent. This particular formulation ensures that it is traceless, which simplifies many formulae because summations over the trace vanish.[78] Substitution into the potential gives

$$
\mathrm{V}(\mathbf{R}) = q\left(\frac{1}{R}\right) - d_\alpha \nabla_\alpha \left(\frac{1}{R}\right) + \frac{1}{3}\Theta_{\alpha\beta}\nabla_\alpha \nabla_\beta \left(\frac{1}{R}\right). \tag{3.1.4}
$$

The reverse operation, representation of an ideal multipole by partial charges at grid sites (or charge density over finite volumes), is degenerate. However, some necessary properties reduce the space of practical solutions. These include local support (region of non-zero values on the grid) and smooth derivatives for the change in charge magnitude due to movement of a multipole site with respect to the grid. For fixed partial charges a normalized cubic basis spline, or B-spline, has been used successfully for discretizing monopole charge distributions (delta functions) on finite difference grids. For quadrupole moments at least 4th order continuity is required such that a normalized 5th order B-spline $\mathrm{N}_5(x)$, which is a piecewise polynomial (Figure 3.1), is appropriate.[79]

$$\mathrm{N}_5(x)=\begin{cases} \frac{1}{24}x^4, & 0\le x\le 1 \\ -\frac{1}{8}+\frac{1}{6}x+\frac{1}{4}(x-1)^2+\frac{1}{6}(x-1)^3-\frac{1}{6}(x-1)^4, & 1\le x\le 2 \\ -\frac{13}{24}+\frac{1}{2}x-\frac{1}{4}(x-2)^2-\frac{1}{2}(x-2)^3+\frac{1}{4}(x-2)^4, & 2\le x\le 3 \\ \frac{47}{24}-\frac{1}{2}x-\frac{1}{4}(x-3)^2+\frac{1}{2}(x-3)^3-\frac{1}{6}(x-3)^4 & 3\le x\le 4 \\ \frac{17}{24}-\frac{1}{6}x+\frac{1}{4}(x-4)^2-\frac{1}{6}(x-4)^3+\frac{1}{24}(x-4)^4, & 4\le x\le 5 \\ 0, & \text{otherwise} \end{cases}. \quad (3.1.5)$$

The sum of this function evaluated at any five evenly spaced points between 0 and 5 is unity. The second part of the Appendix gives a rigorous demonstration that B-splines satisfy the properties of the delta functional and therefore can be used to implement a gradient operator.

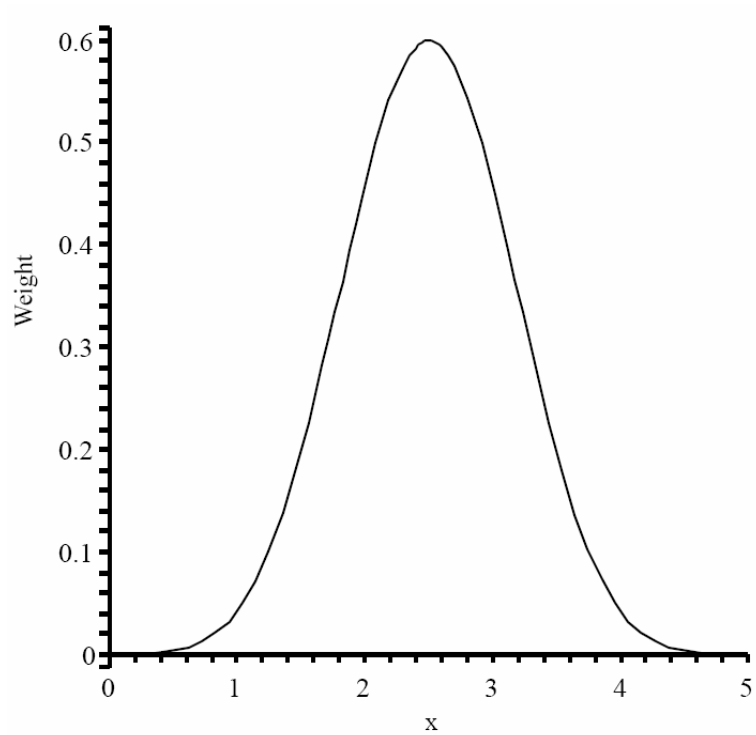

Figure 3.1. Normalized 5th order B-spline on the interval [0, 5].

To illustrate this approach, the fraction of charge that a grid point with coordinates $\mathbf{r}_i$ will receive from a charge site with coordinates $\mathbf{s}_j$ will now be described. We use **r** to denote an N_r x 3 matrix containing all grid coordinates, while **s** is an N_s x 3 matrix containing the coordinates of multipole sites. Elements of both matrices will be specified using two subscripts, the first is an index and the second is a dimension. We first consider the x-dimension, which requires the relative distance of nearby y-z planes from the charge site in dimensionless grid units $\left(r_{i,x} - s_{j,x}\right)/h$, where h is the grid spacing. The B-spline domain is centered over the charge site by shifting its domain from [0,5] to

[-2.5,2.5] by adding 2.5 to its argument. Therefore, the weights of the 5 closest y-z planes to the charge site will be nonzero and sum to 1, where each weight is given by

$$\mathrm{W}\left(r_{i,x}, s_{j,x}\right) = \mathrm{N}_5\left(\frac{r_{i,x} - s_{j,x}}{h} + 2.5\right). \tag{3.1.6}$$

If the charge site is located on a y-z grid plane, then the maximum of the B-spline will be assigned to that plane. Repeated partitioning in the y- and z-dimensions leads to a tensor product description of the charge density

$$\mathrm{B}\left(\mathbf{r}_i, \mathbf{s}_j\right) = \mathrm{W}\left(r_{i,x}, s_{j\,x}\right)\mathrm{W}\left(r_{i,y}, s_{j,y}\right)\mathrm{W}\left(r_{i,z}, s_{j,z}\right). \tag{3.1.7}$$

A further useful property of *n*th order B-splines is that their derivative can be formulated as a linear combination of *n*-1 order B-splines.[79]

$$\frac{\partial \mathrm{N}_n(x)}{\partial x} = \mathrm{N}_{n-1}(x) - \mathrm{N}_{n-1}\left(x-1\right) \tag{3.1.8}$$

For example, the first derivative of the normalized 5th order B-spline can be constructed from two of 4th order, suggesting a dipole basis (or gradient stencil) for determining the electric field from the potential grid as shown in Figure 3.2.

$$\frac{\partial \mathrm{N}_5(x)}{\partial x} = \mathrm{N}_4(x) - \mathrm{N}_4\left(x-1\right) \tag{3.1.9}$$

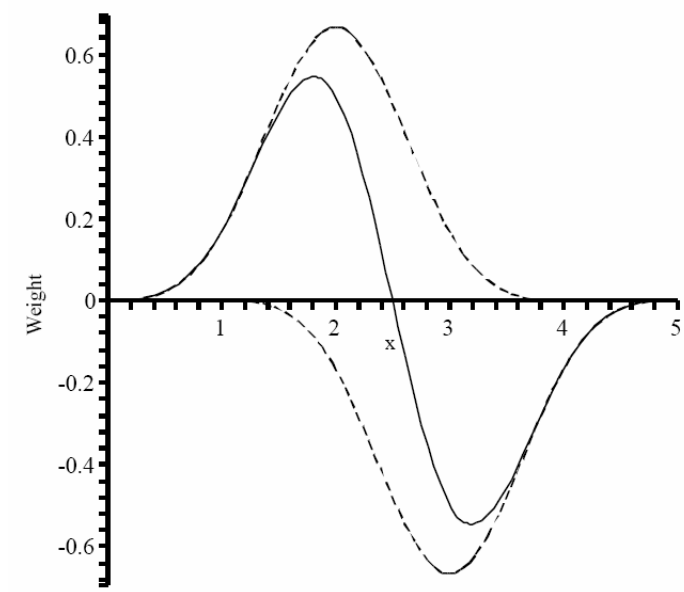

Figure 3.2. The sum of two 4th order B-splines (dashed) are equal to the first derivative of a normalized 5th order B-spline (solid).

Similarly, the 2nd derivative can be constructed from a linear combination of 3rd order B-splines, suggesting an axial quadrupole basis as well as an axial 2nd potential gradient stencil as in Figure 3.3.

$$\frac{\partial^2 \mathrm{N}_5(x)}{\partial x^2} = \mathrm{N}_3(x) - 2\,\mathrm{N}_3(x-1) + \mathrm{N}_3(x-2) \tag{3.1.10}$$

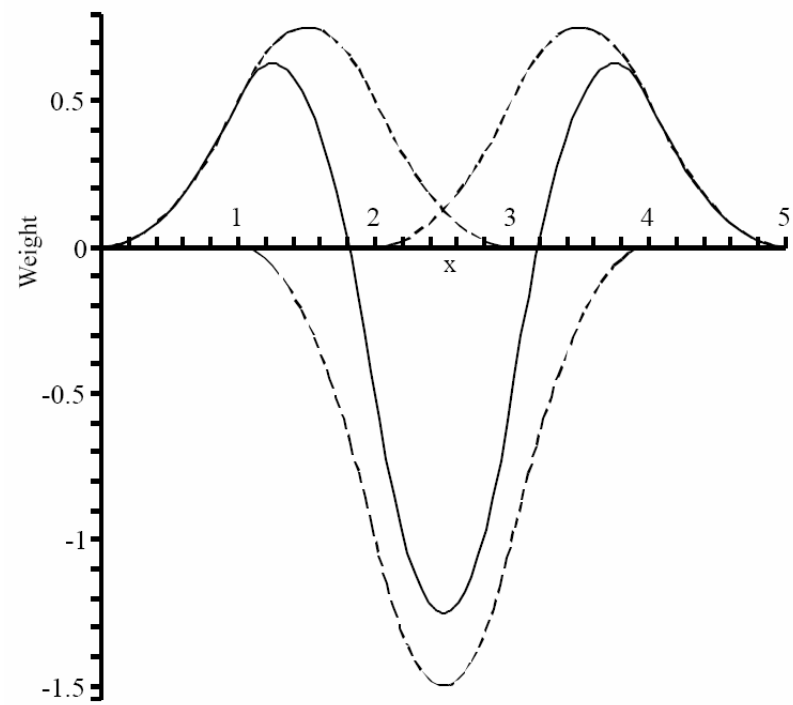

Figure 3.3. The sum of three 3rd order B-splines (dashed) equal the second derivative of a normalized 5th order B-spline (solid).

For notational convenience we define a matrix **B** with dimension N_r x N_s which is used to convert a collection of N_s permanent atomic multipole sites into grid charge density over the N_r grid points

$$\mathbf{B} = \begin{bmatrix} \mathrm{B}(\mathbf{r}_1,\mathbf{s}_1) & \cdots & \mathrm{B}(\mathbf{r}_1,\mathbf{s}_{N_s}) \\ \vdots & \ddots & \vdots \\ \mathrm{B}(\mathbf{r}_{N_r},\mathbf{s}_1) & \cdots & \mathrm{B}(\mathbf{r}_{N_r},\mathbf{s}_{N_s}) \end{bmatrix}. \tag{3.1.11}$$

Only 125 entries per column will have non-zero coefficients, due to each multipole being partitioned locally among 5^3 grid points.

Given the matrix $\mathbf{B}$, the charge density at all grid points due to the permanent multipoles of an AMOEBA solute is

$$\boldsymbol{q}_{\mathrm{M}} = \mathbf{B}\boldsymbol{q} - \frac{1}{h}\nabla_{\alpha}\mathbf{B}\boldsymbol{d}_{\alpha} + \frac{1}{3h^{2}}\nabla_{\alpha}\nabla_{\beta}\mathbf{B}\boldsymbol{\Theta}_{\alpha\beta} \tag{3.1.12}$$

where $\boldsymbol{q}$, $\boldsymbol{d}_x$, $\boldsymbol{d}_y$, $\boldsymbol{d}_z$, $\boldsymbol{\Theta}_{xx}$, $\boldsymbol{\Theta}_{xy}$, $\boldsymbol{\Theta}_{xz}$, …, $\boldsymbol{\Theta}_{zz}$ are column vectors and h normalizes for grid size. This is, in effect, the inverse operation to the original Taylor expansion by which the multipole moments were defined from a collection of point charges. All atomic multipole moments are exactly conserved to numerical precision as long as equal grid spacing is used in each dimension. While most finite difference methods can be generalized to non-uniform Cartesian meshes[75], the nature of traceless multipoles requires uniform Cartesian mesh discretizations. If unequal grid spacing is used, then the trace will be nonzero due to inconsistent coupling between the axial quadrupole components.

The gradient of the charge density at grid sites with respect to an atomic coordinate can be written

$$\frac{\partial \boldsymbol{q}_{\mathrm{M}}}{\partial s_{j,\gamma}} = \frac{\partial \mathbf{B}}{\partial s_{j,\gamma}}\boldsymbol{q} - \frac{1}{h}\nabla_{\alpha}\frac{\partial \mathbf{B}}{\partial s_{j,\gamma}}\boldsymbol{d}_{\alpha} + \frac{1}{3h^{2}}\nabla_{\alpha}\nabla_{\beta}\frac{\partial \mathbf{B}}{\partial s_{j,\gamma}}\boldsymbol{\Theta}_{\alpha\beta}\,. \tag{3.1.13}$$

A more compact notation is needed for derivations presented below, and can be achieved by defining a matrix $\mathbf{T}_{\mathrm{B}}$ of size N_r x $13N_s$

$$\mathbf{T}_B = \begin{bmatrix} \mathrm{B}(\mathbf{r}_1,\mathbf{s}_1) & -\frac{1}{h}\frac{\partial \mathrm{B}(\mathbf{r}_1,\mathbf{s}_1)}{\partial s_{1,x}} & \cdots & \frac{1}{3h^2}\frac{\partial^2 \mathrm{B}(\mathbf{r}_1,\mathbf{s}_1)}{\partial s_{1,z}^2} & \cdots & \frac{1}{3h^2}\frac{\partial^2 \mathrm{B}(\mathbf{r}_1,\mathbf{s}_{N_s})}{\partial s_{N_s,z}^2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{B}(\mathbf{r}_1,\mathbf{s}_1) & -\frac{1}{h}\frac{\partial \mathrm{B}(\mathbf{r}_{N_r},\mathbf{s}_1)}{\partial s_{1,x}} & \cdots & \frac{1}{3h^2}\frac{\partial^2 \mathrm{B}(\mathbf{r}_{N_r},\mathbf{s}_1)}{\partial s_{1,z}^2} & \cdots & \frac{1}{3h^2}\frac{\partial^2 \mathrm{B}(\mathbf{r}_{N_r},\mathbf{s}_{N_s})}{\partial s_{N_s,z}^2} \end{bmatrix}. \quad (3.1.14)$$

The matrix product $\boldsymbol{\Phi}^t\mathbf{T}_\mathrm{B}$, where $\boldsymbol{\Phi}$ is a column vector of length $\mathrm{N_r}$ containing the potential from a numerical solution to the LPBE, produces the same tensor components (i.e., the potential, field and field gradient) as $\mathbf{M}^t\mathbf{T}$ in Eq. (2.1.6) for the AMOEBA vacuum electrostatic energy. Using this notation allows manipulation of reaction potentials and intramolecular potentials to be handled on equal footing. The same approach is appropriate for induced dipoles.

There is a trade-off between higher order B-splines and the goal of maintaining the smallest possible support for the multipoles. As the support grows, the charge density is less representative of the ideal multipole limit. Additionally, placement of solute charge density outside the low-dielectric cavity should be avoided. This restriction of charge density to the solute interior places an upper bound on acceptable grid spacings for use with finite difference discretizations of higher-order B-splines. If, for example, the solute cavity for a hydrogen atom ends approximately 1.2 Å from its center, then the maximum recommended grid spacing when using 5th order B-splines is 0.48 Å (1.2 / 2.5), whereas for third order B-splines a value of 0.80 Å is reasonable (1.2 / 1.5). Therefore,

the use of quintic B-splines requires a smaller upper bound on grid spacing than cubic B-splines.

3.2 Permittivity and Modified Debye-Hückel Screening Factor

The permittivity $\varepsilon(\mathbf{r})$ and modified Debye-Hückel screening factor $\bar{\kappa}^2(\mathbf{r})$ functions are defined through a characteristic function $\mathrm{H}(\mathbf{r}_i, \mathbf{s})$, where $\mathbf{r}_\mathrm{i}$ represents the coordinates of a grid point and **s** the coordinates of all multipole sites. Inside the solute cavity the characteristic function is 0, while in the solvent it is 1. For the homogeneous calculation the permittivity is set to unity over all space, while for the solvated state it takes the value 1 inside the solute, ε_s in solvent, and intermediate values over a transition region

$$\varepsilon(\mathbf{r}_i) = 1 + (\varepsilon_s - 1)\mathrm{H}(\mathbf{r}_i, \mathbf{s}, b, e)\,. \tag{3.2.1}$$

where b is the beginning of the transition and e defined below.

The modified Debye-Hückel screening factor is zero everywhere for the vacuum calculation and for the solvated calculation is defined by

$$\bar{\kappa}^2(\mathbf{r}_i) = \bar{\kappa}_b^2 \mathrm{H}(\mathbf{r}_i, \mathbf{s}, b, e) \tag{3.2.2}$$

where $\bar{\kappa}_b^2 = \varepsilon_s \kappa_b^2$ is the modified bulk screening factor and is related to the ionic strength $I = \frac{1}{2}\sum_i q_i^2 c_i$ via $\kappa_b^2 = 8\pi I / \varepsilon_s k_B T$. Here q_i and c_i are the charge and number

concentration of mobile ion species i, respectively, k_{B} is the Boltzmann constant and T the absolute temperature. The characteristic function itself is sometimes formulated as the product of a radially symmetric function applied to each solute atom

$$\mathrm{H}\left(\mathbf{r}_i, \mathbf{s}\right) = \prod_{j=1,N_s} \mathrm{H}_j\left(\left|\mathbf{r}_i - \mathbf{s}_j\right|, b, e\right) \tag{3.2.3}$$

where $\mathrm{H}_j\left(r, b, e\right)$ must allow for a smooth transition across the solute-solvent boundary to achieve stable Cartesian energy gradients.[19, 22] This is a result of terms that depend on the gradient of the characteristic function with respect to an atomic displacement. A successful approach to defining a differentiable boundary is the use of a polynomial switch $S_n\left(r\right)$ of order n, although other definitions have been suggested based on atom centered Gaussians.[80, 81] For any atom j, $\mathrm{H}_j\left(r, b, e\right)$ takes the form

$$\mathrm{H}_j\left(r, b, e\right) = \begin{cases} 0, & r \le b \\ S_n\left(r, b, e\right), & b < r < e \\ 1, & e \le r \end{cases}, \tag{3.2.4}$$

For the permittivity, the switch begins and ends at

$$\begin{aligned} b &= \sigma_j - w \\ e &= \sigma_j + w \end{aligned}, \tag{3.2.5}$$

where $\mathbf{T}_{\mathrm{d}}^{(1)}$ is the radius of atom j and w indicates how far the smoothing window extends radially inward and outward. For the modified Debye-Hückel screening factor the radius of the largest ionic species σ_{ion} is also taken into account

$$\begin{aligned} b &= \sigma_j + \sigma_{ion} - w \\ e &= \sigma_j + \sigma_{ion} + w \end{aligned} \tag{3.2.4}$$

For fixed partial charge force fields, a cubic switch S_3 has been used with success. However, a characteristic function with higher order continuity has been found to improve energy conservation at a given grid spacing. Table 3.1 reports representative examples of this effect. By using a 7th order polynomial switch S_7,

$$
\begin{aligned}
&S_7(r,b,e) = \frac{c_7 r^7 + c_6 r^6 + c_5 r^5 + c_4 r^4 + c_3 r^3 + c_2 r^2 + c_1 r + c_0}{-b^7 + 7b^6 e - 21b^5 e^2 + 35b^4 e^3 - 35b^3 e^4 + 21b^2 e^5 - 7be^6 + e^7}, \\
&c_0 = b^4\left(-b^3 + 7b^2 e - 21be^2 + 35e^3\right), \\
&c_1 = -140b^3 e^3, \\
&c_2 = 210b^2 e^2 (b+e), \\
&c_3 = -140be\left(b^2 + 3be + e^2\right), \\
&c_4 = 35\left(b^3 + 9b^2 e + 9be^2 + e^3\right), \\
&c_5 = -84\left(b^2 + 3be + e^2\right), \\
&c_6 = 70(b+e), \\
&c_7 = -20
\end{aligned}
\quad , \quad (3.2.5)
$$

the first three derivatives of the characteristic function can be constrained to zero at the beginning and end of the switching region. The cubic, quintic and heptic volume exclusions functions are shown in Figure 3.4.

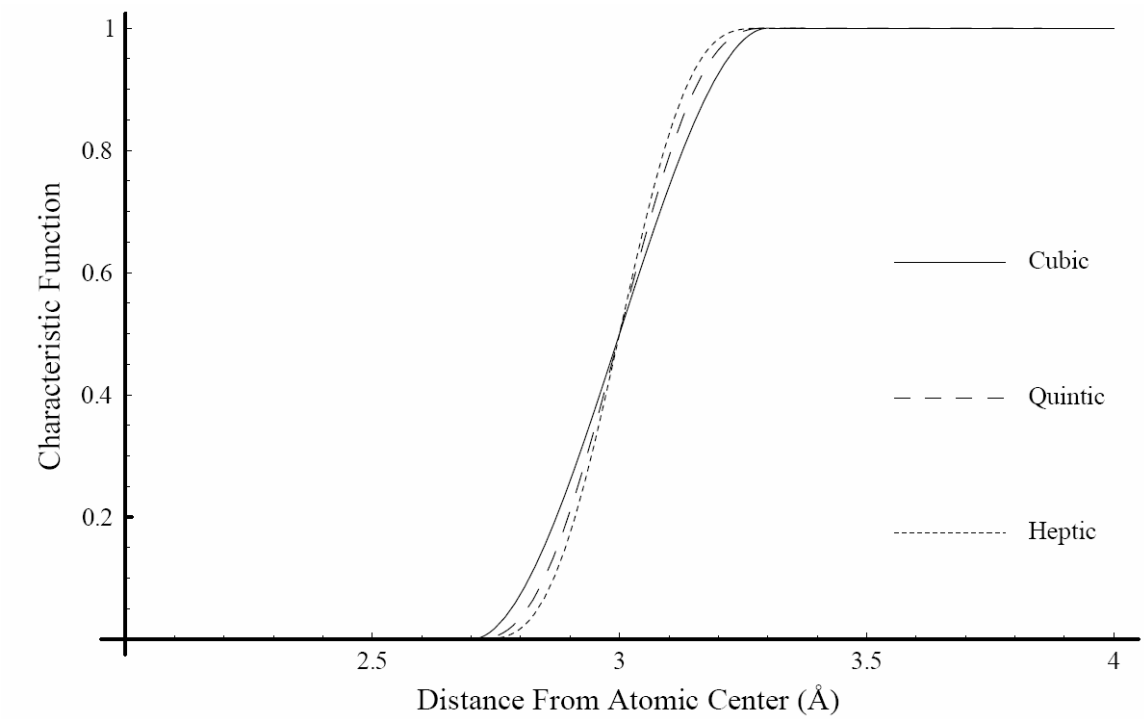

Figure 3.4. Comparison of cubic, quintic and heptic characteristic functions for an atom with radius 3 Å using a total window width of 0.6 Å.

Table 3.1. The norm of the gradient sum over all atoms (kcal/mole/Å) for three different solutes is shown for cubic, quintic and heptic characteristic functions at two grid spacings. A norm of zero, indicating perfect conservation of energy, is nearly achieved for acetamide and ethanol at 0.11 Å grid spacing using a heptic characteristic function. Conservation of energy is improved by reducing grid spacing and also by increasing the continuity of the solute-solvent boundary via the characteristic function.

	Acetamide		**Ethanol**		**CRN**	
Grid Spacing	0.21	0.11	0.22	0.11	0.32	0.18
Cubic	2.48	0.46	0.77	0.29	18.82	2.58
Quintic	1.93	0.23	0.32	0.15	9.38	2.42
Heptic	0.88	0.06	0.21	0.06	6.27	1.26

3.3 Boundary Conditions

Single Debye-Hückel (SDH) and multiple Debye-Hückel (MDH) boundary conditions for a solute are two common approximations to the true potential used to specify Dirichlet boundary conditions for non-spherical solutes described by a collection of atomic multipoles.[6] SDH assumes that all atomic multipole sites are collected into a single multipole at the center of the solute, which is approximated by a sphere. MDH assumes the superposition of the contribution of each atomic multipole considered in the absence of all other sites that displace solvent. Therefore, to construct the Dirichlet problem for a solute described by an arbitrary number of atomic multipole sites, the potential outside a solvated multipole located at the center of a sphere is required.

For solvent described by the LPBE

$$\nabla^2 \Phi(\mathbf{r}) = \kappa_b^2 \Phi(\mathbf{r}) \tag{3.3.1}$$

a solution was first formulated by Kirkwood.[27] This form of the LPBE is simplified relative to Eq. (1.1.5) since there is no fixed charge distribution and no spatial variation in either the permittivity or Debye-Hückel screening factor. Inside the cavity, the Poisson equation is obeyed

$$\nabla^2 \Phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon}, \tag{3.3.2}$$

where the charge density $\rho(\mathbf{r})$ may contain moments of arbitrary order. The boundary conditions are enforced by requiring the potential and dielectric displacement to be continuous across the interface between the solute and solvent via

$$\Phi(\mathbf{r})_{\text{in}} = \Phi(\mathbf{r})_{\text{out}}, \tag{3.3.3}$$

and

$$\varepsilon \frac{\partial \Phi(\mathbf{r})_{\text{in}}}{\partial r} = \varepsilon_{\text{out}} \frac{\partial \Phi(\mathbf{r})_{\text{out}}}{\partial r}, \tag{3.3.4}$$

respectively. Additional requirements on the solution are that it be bounded at the origin and approach an arbitrary constant at infinity, usually chosen to be zero.

In presenting the solution, it is convenient for our purposes to continue using Cartesian multipoles, rather than switching to spherical harmonics, although there is a well-known equivalence between the two approaches.[82, 83] The potential at **r** due to a symmetric, traceless multipole in a homogeneous dielectric ε is

$$\begin{aligned}\Phi_{\varepsilon}(\mathbf{r}_{ij}) &= (\mathbf{T}_{\varepsilon})^{t}\,\mathbf{M}_{j} \\ &= \left(\begin{bmatrix} 1 \\ -\frac{\partial}{\partial x} \\ -\frac{\partial}{\partial y} \\ -\frac{\partial}{\partial z} \\ \frac{1}{3}\frac{\partial^2}{\partial x \partial x} \\ \vdots \end{bmatrix} \left(\frac{1}{\varepsilon r_{ij}} \right) \right)^{t} \begin{bmatrix} q_j \\ \mu_{x,j} \\ \mu_{y,j} \\ \mu_{z,j} \\ \Theta_{xx,j} \\ \vdots \end{bmatrix},\end{aligned} \tag{3.3.5}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{s}_j$ might be the difference between a grid location and a multipole site. The potential inside the spherical cavity is the superposition of the homogeneous potential and the reaction potential

$$\Phi_{\text{in}}\left(\mathbf{r}_{ij}\right) = \left[\left(\mathbf{I} + \mathbf{R}_{\text{in}}\right)\mathbf{T}_{\varepsilon}\right]^{t}\mathbf{M}_{j}, \tag{3.3.6}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{R}_{\text{in}}$ is a diagonal matrix with diagonal elements

$$\left[c_{\text{in}}(0), c_{\text{in}}(1), c_{\text{in}}(1), c_{\text{in}}(1), c_{\text{in}}(2), \cdots\right] \tag{3.3.7}$$

that are based on coefficients for multipoles of order n to be determined by the boundary conditions

$$c_{\text{in}}(n) = \beta_n \left(\frac{r_{ij}}{a}\right)^{2n+1}. \tag{3.3.8}$$

Similarly, the potential outside the cavity is

$$\Phi_{\text{out}}\left(\mathbf{r}_{ij}\right) = \left(\mathbf{R}_{\text{out}}\mathbf{T}_{\varepsilon}\right)^{t}\mathbf{M}_{j}, \tag{3.3.9}$$

where $\mathbf{R}_{\text{out}}$ is a diagonal matrix with diagonal elements

$$\left[c_{\text{out}}(0), c_{\text{out}}(1), c_{\text{out}}(1), c_{\text{out}}(1), c_{\text{out}}(2), \cdots\right] \tag{3.3.10}$$

based on a second set of coefficients for multipoles of order n also determined by the boundary conditions

$$c_{\text{out}}(n) = \frac{\varepsilon}{\varepsilon_{\text{out}}}\kappa r_{ij}\alpha_n k_n\left(\kappa r_{ij}\right)\left(\frac{r_{ij}}{a}\right)^{n}, \tag{3.3.11}$$

where $k_n(x)$ is the modified spherical Bessel function of the third kind

$$k_n(x) = \frac{\pi \mathrm{e}^{-x}}{2x} \sum_{i=0}^{n} \frac{(n+i)!}{i!(n-i)!(2x)^i} . \tag{3.3.12}$$

Kirkwood first solved Eqs. (3.3.1) and (3.3.2) subject to the boundary conditions in Eqs. (3.3.3) and (3.3.4) to determine α_n and β_n as

$$\alpha_n = \frac{(2n+1)/\kappa a}{n k_n(\kappa a)/\hat{\varepsilon} - \kappa a k_n'(\kappa a)} \tag{3.3.13}$$

and

$$\beta_n = \frac{(n+1) k_n(\kappa a)/\hat{\varepsilon} + \kappa a k_n'(\kappa a)}{k_n(\kappa a)/\hat{\varepsilon} - \kappa a k_n'(\kappa a)} , \tag{3.3.14}$$

where $k_n'(x)$ is the derivative of $k_n(x)$ and $\hat{\varepsilon}$ is the ratio of the permittivity in solvent to that inside the sphere $\varepsilon_{out}/\varepsilon$.[27] We only require the potential outside the cavity to construct SDH and MDH boundary conditions and therefore we provide specific values of α_n and k_n through quadrupole order as shown in Table 3.2. As the ionic strength goes to zero, the Laplace equation is obeyed in solvent. For multipoles through quadrupole order, the difference between the LPBE and Laplace potentials outside the cavity are summarized in Table 3.3.

Table 3.2. Explicit values for the functions $\alpha_n(x)$ and $k_n(x)$ up to quadrupole order.

n	$\alpha_n(x)/\left(\frac{2\exp(x)}{\pi}\right)$	$k_n(x)/\left(\frac{\pi}{2\exp(x)}\right)$
0	$\frac{1}{(1+x)}$	$\frac{1}{x}$
1	$\frac{3\hat{\varepsilon}x}{1+x+\hat{\varepsilon}\left(2+2x+x^2\right)}$	$\frac{1+x}{x^2}$
2	$\frac{5\hat{\varepsilon}x^2}{2\left(3+3x+x^2\right)+\hat{\varepsilon}\left(9+9x+4x^2+x^3\right)}$	$\frac{\left(3+3x+x^2\right)}{x^3}$

Table 3.3. Explicit values of the coefficients used to calculate the potential at the grid boundary of LPBE and PE calculations, respectively, under the SDH or MDH approximation. The LPBE coefficients reduce to the PE coefficients as salt concentration goes to zero.

n	$\kappa r\alpha_n(\kappa a)k_n(\kappa r)\left(\frac{r}{a}\right)^n$	$\lim_{\kappa\to 0}\left[\kappa r\alpha_n(\kappa a)k_n(\kappa r)\left(\frac{r}{a}\right)^n\right]$
0	$\frac{\exp(\kappa(a-r))}{1+\kappa a}$	1
1	$\frac{3\hat{\varepsilon}\exp(\kappa(a-r))(1+\kappa r)}{1+\kappa a+\hat{\varepsilon}\left(2+2\kappa a+(\kappa a)^2\right)}$	$\frac{3\hat{\varepsilon}}{1+2\hat{\varepsilon}}$
2	$\frac{5\hat{\varepsilon}\exp(\kappa(a-r))\left(3+3\kappa r+(\kappa r)^2\right)}{2\left(3+3\kappa a+(\kappa a)^2\right)+\hat{\varepsilon}\left(9+9\kappa a+4(\kappa a)^2+(\kappa a)^3\right)}$	$\frac{5\hat{\varepsilon}}{2+3\hat{\varepsilon}}$

3.4 Permanent Multipole Energy and Gradient

The PMPB permanent atomic multipole (PAM) solvation energy and gradient are very similar to those for fixed partial charge force fields. Based on Eqs. (2.2.3) and (3.1.12) the PAM vacuum, solvated and reaction potentials are, respectively,

$$\begin{aligned} \boldsymbol{\Phi}_{\mathrm{v}}^{\mathrm{M}} &= -4\pi \mathbf{A}_{\mathrm{v}}^{-1} \boldsymbol{q}_{\mathrm{M}} \\ \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} &= -4\pi \mathbf{A}_{\mathrm{s}}^{-1} \boldsymbol{q}_{\mathrm{M}} \\ \boldsymbol{\Phi}^{\mathrm{M}} &= \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} - \boldsymbol{\Phi}_{\mathrm{v}}^{\mathrm{M}} \end{aligned} . \tag{3.4.1}$$

The expression for the permanent electrostatic solvation energy is then identical to that for a fixed partial charge force field given in Eqs. (2.2.1) and (2.2.4), except the source charge density is based on PAM via $\boldsymbol{q}_{\mathrm{M}}$

$$\begin{aligned} \Delta \mathrm{G}_{\mathrm{M}} &= \frac{1}{2}\left(\boldsymbol{\Phi}^{\mathrm{M}}\right)^{t} \boldsymbol{q}_{\mathrm{M}} \\ &= \frac{1}{2}\left(-4\pi \boldsymbol{q}_{\mathrm{M}}\right)^{t}\left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1}\right)\boldsymbol{q}_{\mathrm{M}} \end{aligned} . \tag{3.4.2}$$

Derivation of the energy gradient is identical to Eqs. (2.2.5) through (2.2.10) and yields

$$\frac{\partial \Delta \mathrm{G}_{\mathrm{M}}}{\partial s_{j,\gamma}} = \left(\boldsymbol{\Phi}^{\mathrm{M}}\right)^{t} \frac{\partial \boldsymbol{q}_{\mathrm{M}}}{\partial s_{j,\gamma}} + \frac{1}{8\pi}\left(\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}}\right)^{t} \frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} . \tag{3.4.3}$$

There are, however, some important differences between achieving smooth gradients for a fixed partial charge force field and one based on PAM. First, as discussed in the section on multipole discretization, quadrupoles require at least 5th order B-splines to guarantee continuous derivatives of the source charge density with respect to movement of a multipole site

$$\frac{\partial \boldsymbol{q}_{\mathrm{M}}}{\partial \mathrm{s}_{j,\gamma}} = \frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial \mathrm{s}_{j,\gamma}} \mathbf{M} . \tag{3.4.4}$$

Second, we have found that if a third order polynomial is used to define the transition between solute and solvent for purposes of assigning the permittivity and the modified Debye-Hückel screening factor, energy conservation is achieved only for very fine grid spacing. As discussed earlier, use of a 7th order polynomial improves energy conservation for coarser grids. Details of the numerical realization of Eq. (3.4.3), including torques, are presented in Appendix C (p. 159).

3.5 Self-Consistent Reaction Field

An SCRF protocol is used to achieve numerical convergence of the coupling between a polarizable solute and continuum solvent. The starting point of the iterative convergence is the total "direct" field $\mathbf{E}_{\mathrm{d}}$ at each polarizable site. This is defined by the sum of the PAM intramolecular field

$$\mathbf{E}_{\mathrm{d}} = \mathbf{T}_{\mathrm{d}}^{(1)} \mathbf{M} , \tag{3.5.1}$$

where $\mathbf{T}_{\mathrm{d}}^{(1)}$ is analogous to the tensor matrix used in deriving the AMOEBA vacuum energy in Eq. (2.1.6), and the PAM reaction field

$$\mathbf{E}_{\mathrm{RF}}^{\mathrm{M}} = -\mathbf{D}_{\mathrm{B}}^{t}\boldsymbol{\Phi}^{\mathrm{M}} , \tag{3.5.2}$$

where $\mathbf{D}_{\mathrm{B}}$ is a matrix of B-spline derivatives of size $\mathrm{N_r}$ by $3\mathrm{N_s}$

$$\mathbf{D}_{\mathrm{B}} = -\frac{1}{h}\begin{bmatrix} \frac{\partial \mathrm{B}\left(\mathbf{r}_1,\mathbf{s}_1\right)}{\partial s_{1,x}} & \frac{\partial \mathrm{B}\left(\mathbf{r}_1,\mathbf{s}_1\right)}{\partial s_{1,y}} & \frac{\partial \mathrm{B}\left(\mathbf{r}_1,\mathbf{s}_1\right)}{\partial s_{1,z}} & \cdots & \frac{\partial \mathrm{B}\left(\mathbf{r}_1,\mathbf{s}_{\mathrm{N_s}}\right)}{\partial s_{\mathrm{N_s},z}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{B}\left(\mathbf{r}_{\mathrm{N_r}},\mathbf{s}_1\right)}{\partial s_{1,x}} & \frac{\partial \mathrm{B}\left(\mathbf{r}_{\mathrm{N_r}},\mathbf{s}_1\right)}{\partial s_{1,y}} & \frac{\partial \mathrm{B}\left(\mathbf{r}_{\mathrm{N_r}},\mathbf{s}_1\right)}{\partial s_{1,z}} & \cdots & \frac{\partial \mathrm{B}\left(\mathbf{r}_{\mathrm{N_r}},\mathbf{s}_{\mathrm{N_s}}\right)}{\partial s_{\mathrm{N_s},z}} \end{bmatrix}, \tag{3.5.3}$$

that produces the reaction field at induced dipoles sites given a potential grid. The induced dipoles are determined as the product of the direct field $\mathbf{E}_{\mathrm{d}}$ with a vector of isotropic atomic polarizabilities $\boldsymbol{\alpha}$:

$$\begin{aligned} \boldsymbol{\mu}_{\mathrm{d}} &= \boldsymbol{\alpha}\mathbf{E}_{\mathrm{d}} \\ &= \boldsymbol{\alpha}\left(\mathbf{T}_{\mathrm{d}}^{(1)}\mathbf{M} - \mathbf{D}_{\mathrm{B}}^{t}\boldsymbol{\Phi}^{\mathrm{M}}\right) . \end{aligned} \tag{3.5.4}$$

We define the direct model of polarization to consist of induced dipoles not acted upon by each other or their own reaction field. Although this is a nontrivial approximation, the direct model requires little more work to compute energies than a fixed partial charge force field since the limiting factor in both cases is two numerical LPBE solutions. Energy gradients under the direct model require three pairs of LPBE solutions and are therefore a factor of 3 more expensive than for a fixed charge solute. The direct model is expected to be quite useful for many applications. For example, a geometry optimization

might utilize the direct polarization model initially, then switch to the more expensive mutual polarization model described below as the minimum is approached.

In contrast to the direct model, the total solvated field $\mathbf{E}$ has two additional contributions due to the induced dipoles and their reaction field,

$$\mathbf{E} = \mathbf{T}_{\mathrm{d}}^{(1)}\mathbf{M} + \mathbf{T}^{(11)}\boldsymbol{\mu} - \mathbf{D}_{\mathrm{B}}^{t}\left(\boldsymbol{\Phi}^{\mathrm{M}} + \boldsymbol{\Phi}^{\mu}\right), \tag{3.5.5}$$

for a sum of 4 contributions. The procedure for determining the vacuum, solvated and reaction potential, respectively, due to the induced dipoles is identical to that of the PAM

$$\begin{aligned} \boldsymbol{\Phi}_{\mathrm{v}}^{\mu} &= -4\pi\mathbf{A}_{\mathrm{v}}^{-1}\boldsymbol{q}_{\mu} \\ \boldsymbol{\Phi}_{\mathrm{s}}^{\mu} &= -4\pi\mathbf{A}_{\mathrm{s}}^{-1}\boldsymbol{q}_{\mu}\,, \\ \boldsymbol{\Phi}^{\mu} &= \boldsymbol{\Phi}_{\mathrm{s}}^{\mu} - \boldsymbol{\Phi}_{\mathrm{v}}^{\mu} \end{aligned} \tag{3.5.6}$$

except the source charge density is

$$\boldsymbol{q}_{\mu} = \mathbf{D}_{\mathrm{B}}\boldsymbol{\mu}\,. \tag{3.5.7}$$

The induced dipoles

$$\boldsymbol{\mu} = \boldsymbol{\alpha}\left[\mathbf{T}_{\mathrm{d}}^{(1)}\mathbf{M} + \mathbf{T}^{(11)}\boldsymbol{\mu} - \mathbf{D}_{\mathrm{B}}^{t}\left(\boldsymbol{\Phi}^{\mathrm{M}} + \boldsymbol{\Phi}^{\mu}\right)\right], \tag{3.5.8}$$

can be determined in an iterative fashion using successive over-relaxation (SOR) to accelerate convergence.[84] The SCRF is usually deemed to have converged when the change in the induced dipoles is less than 10^{-2} RMS debye between steps. This generally requires 4-5 cycles and therefore the mutual polarization model necessitates 8-10 additional numerical solutions of the LPBE to determine the PMPB solvation energy. Although calculation of the direct polarization energy is no more expensive than that for

fixed multipoles, mutual polarization energies that depend on SCRF convergence are approximately a factor of 5 more costly.

3.6 PMPB Electrostatic Solvation Free Energy

Having described the PAM solvation energy and gradients and our approach for determining the induced dipoles, it is now possible to discuss the total solvated electrostatic energy for the PMPB model,

$$\mathrm{U}_{\mathrm{elec}} = \frac{1}{2}\left[\mathbf{M}^t\mathbf{T} - \boldsymbol{\mu}^t\mathbf{T}_{\mathrm{p}}^{(1)} + \boldsymbol{\Phi}^t\mathbf{T}_{\mathrm{B}}\right]\mathbf{M}\,, \tag{3.6.1}$$

where $\boldsymbol{\Phi}$ is the LPBE reaction potential for the converged solute charge distribution

$$\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\mathrm{M}} + \boldsymbol{\Phi}^{\mu}\,. \tag{3.6.2}$$

The total electrostatic energy in solvent is similar to the vacuum electrostatic energy of Eq. (2.1.6), with an important difference. The vacuum induced dipoles $\boldsymbol{\mu}^{\mathrm{v}}$ change in the presence of a continuum solvent by an amount represented by $\boldsymbol{\mu}^{\Delta}$, such that the SCRF induced moments $\boldsymbol{\mu}$ can be decomposed into a sum

$$\boldsymbol{\mu} = \boldsymbol{\mu}^{\mathrm{v}} + \boldsymbol{\mu}^{\Delta}\,. \tag{3.6.3}$$

The change in the potential, field, etc. within the solute is not only a result of the solvent response, but also due to changes in intramolecular polarization. By definition, the electrostatic solvation energy $\Delta\mathrm{G}_{\mathrm{elec}}$ is the change in total electrostatic energy due to moving from vacuum to solvent

$$\begin{aligned}\Delta G_{elec} &= U_{elec} - U^{v}_{elec} \\ &= \frac{1}{2}\left[-\left(\boldsymbol{\mu}^{\Delta}\right)^{t}\mathbf{T}^{(1)}_{p} + \boldsymbol{\Phi}^{t}\mathbf{T}_{B}\right]\mathbf{M}\end{aligned} . \tag{3.6.4}$$

In practice, it is convenient to compute the total solvated electrostatic energy U_{elec} and vacuum electrostatic energy U^{v}_{elec} using the SCRF $\boldsymbol{\mu}$ and vacuum $\boldsymbol{\mu}^{v}$ induced dipole moments, respectively. The electrostatic solvation energy ΔG_{elec} is then determined as the difference.

3.7 Polarization Energy Gradient

As described in Section 2.1, the induced dipoles are determined using an iterative SOR procedure until a predetermined convergence criterion is achieved. Since this is a linear system, it is possible to solve for the induced dipoles directly, which facilitates derivation of the polarization energy gradient with respect to atomic displacements. Substitution of the reaction potential due to the induced dipoles $\boldsymbol{\mu}^{v}$ from Eq. (3.5.6) into the expression for the induced dipoles in Eq. (3.5.8) makes clear all dependencies on $\boldsymbol{\mu}$.

$$\boldsymbol{\mu} = \boldsymbol{\alpha}\left[\mathbf{T}^{(1)}_{d}\mathbf{M} - \mathbf{D}^{t}_{B}\boldsymbol{\Phi}^{M} + \mathbf{T}^{(11)}\boldsymbol{\mu} - 4\pi\mathbf{D}^{t}_{B}\left(\mathbf{A}^{-1}_{s} - \mathbf{A}^{-1}_{v}\right)\mathbf{D}_{B}\right]. \tag{3.7.1}$$

Collecting all terms containing the induced dipoles on the left hand side gives

$$\left[\boldsymbol{\alpha}^{-1} - \mathbf{T}^{(11)} + 4\pi\mathbf{D}^{t}_{B}\left(\mathbf{A}^{-1}_{s} - \mathbf{A}^{-1}_{v}\right)\mathbf{D}_{B}\right]\boldsymbol{\mu} = \mathbf{T}^{(1)}_{d}\mathbf{M} - \mathbf{D}^{t}_{B}\boldsymbol{\Phi}^{M} . \tag{3.7.2}$$

For convenience, a matrix **C** is defined as

$$\mathbf{C} = \left[\boldsymbol{\alpha}^{-1} - \mathbf{T}^{(11)} + 4\pi \mathbf{D}_{\mathrm{B}}^{t} \left(\mathbf{A}_{s}^{-1} - \mathbf{A}_{v}^{-1} \right) \mathbf{D}_{\mathrm{B}} \right], \tag{3.7.3}$$

which is substituted into Eq. (3.7.2) to show the induced dipoles are a linear function of the PAM **M**

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{C}^{-1} \left(\mathbf{T}_{\mathrm{d}}^{(1)} \mathbf{M} - \mathbf{D}_{\mathrm{B}}^{t} \boldsymbol{\Phi}^{\mathrm{M}} \right) \\ &= \mathbf{C}^{-1} \left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}} \right) \end{aligned} . \tag{3.7.4}$$

The first term results from the intramolecular interaction tensor $\mathbf{T}_{\mathrm{d}}^{(1)}$ that implicitly contains the AMOEBA group based polarization scheme, and the second term is the permanent reaction field.

The polarization energy can now be described in terms of the permanent reaction field and permanent intramolecular solute field $\mathbf{E}_{\mathrm{p}}$

$$\mathrm{U}_{\boldsymbol{\mu}} = -\frac{1}{2} \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}} \right)^{t} \boldsymbol{\mu} . \tag{3.7.5}$$

To find the polarization energy gradient, we wish to avoid terms that rely on the change in induced dipoles with respect to atomic displacement. Therefore, the induced dipoles in Eq. (3.7.5) are replaced using Eq. (3.7.4) to yield

$$\mathrm{U}_{\boldsymbol{\mu}} = -\frac{1}{2}\left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right)^{t} \mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right). \tag{3.7.6}$$

By the chain rule, the polarization energy gradient is

$$\begin{aligned}\frac{\partial \mathrm{U}_{\boldsymbol{\mu}}}{\partial s_{j,\gamma}} = -\frac{1}{2}\Bigg[&\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial s_{j,\gamma}} + \frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}}\right)^{t} \mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right) + \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right)^{t} \frac{\partial \mathbf{C}^{-1}}{\partial s_{j,\gamma}}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right) \\ &+ \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right)^{t} \mathbf{C}^{-1}\left(\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial s_{j,\gamma}} + \frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}}\right)\Bigg]\end{aligned}. \tag{3.7.7}$$

For convenience a mathematical quantity $\boldsymbol{\nu}$ is defined as

$$\boldsymbol{\nu} = \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}\right)\mathbf{C}^{-1}, \tag{3.7.8}$$

which is similar to $\boldsymbol{\mu}$. We can now greatly simplify Eq. (3.7.7) using Eqs. (3.7.4) and (3.7.8) along with the identity $\frac{\partial \mathbf{C}^{-1}}{\partial s_{j,\gamma}} = -\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial s_{j,\gamma}}\mathbf{C}^{-1}$ to give

$$\frac{\partial \mathrm{U}_{\boldsymbol{\mu}}}{\partial s_{j,\gamma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial s_{j,\gamma}}\right)^{t} \boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial s_{j,\gamma}} + \left(\frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}}\right)^{t} \boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}} - \boldsymbol{\nu}^{t}\frac{\partial \mathbf{C}}{\partial s_{j,\gamma}}\boldsymbol{\mu}\right]. \tag{3.7.9}$$

3.7.1 Direct Polarization Energy Gradient

Under the direct polarization model, **C** is an identity matrix whose derivative is zero, and therefore Eq. (3.7.9) simplifies to

$$\frac{\partial \mathrm{U}_{\boldsymbol{\mu}_\mathrm{d}}}{\partial s_{j,\gamma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_\mathrm{p}}{\partial s_{j,\gamma}}\right)^t \boldsymbol{\mu} + \boldsymbol{\nu}^t \frac{\partial \mathbf{E}_\mathrm{d}}{\partial s_{j,\gamma}} + \left(\frac{\partial \mathbf{E}_\mathrm{RF}^\mathrm{M}}{\partial s_{j,\gamma}}\right)^t \boldsymbol{\mu} + \boldsymbol{\nu}^t \frac{\partial \mathbf{E}_\mathrm{RF}^\mathrm{M}}{\partial s_{j,\gamma}}\right]. \tag{3.7.10}$$

The first two terms on the RHS appear in the polarization energy gradient even in the absence of a continuum reaction field and are described elsewhere.[2, 7] The third and fourth terms are specific to LPBE calculations and will now be discussed. The derivative of the LPBE reaction field due to permanent multipoles with respect to movement of any atom has a similar form to the analogous derivative of the potential. Substitution for the field using Eq. (3.5.2) into the third term of Eq. (3.7.10) gives

$$\left(\frac{\partial \mathbf{E}_\mathrm{RF}^\mathrm{M}}{\partial s_{j,\gamma}}\right)^t \boldsymbol{\mu} = -\frac{\partial \mathbf{D}_\mathrm{B}^t \boldsymbol{\Phi}^\mathrm{M}}{\partial s_{j,\gamma}} \boldsymbol{\mu}\,. \tag{3.7.11}$$

Substitution of the permanent multipole potential from Eq. (3.4.1) into Eq. (3.7.11) yields

$$\left(\frac{\partial \mathbf{E}_\mathrm{RF}^\mathrm{M}}{\partial s_{j,\gamma}}\right)^t \boldsymbol{\mu} = 4\pi \frac{\partial}{\partial s_{j,\gamma}}\left[\mathbf{D}_\mathrm{B}^t\left(\mathbf{A}_\mathrm{s}^{-1} - \mathbf{A}_\mathrm{v}^{-1}\right)\mathbf{T}_\mathrm{B}\mathbf{M}\right]\boldsymbol{\mu}\,, \tag{3.7.12}$$

which is differentiated by applying the chain rule

$$\left(\frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}}\right)^{t} \boldsymbol{\mu} = 4\pi \left[\frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}} \left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1} \right) \mathbf{T}_{\mathrm{B}} \mathbf{M} + \mathbf{D}_{\mathrm{B}}^{t} \frac{\partial \mathbf{A}_{\mathrm{s}}^{-1}}{\partial s_{j,\gamma}} \mathbf{T}_{\mathrm{B}} \mathbf{M} + \mathbf{D}_{\mathrm{B}}^{t} \left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1} \right) \frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial s_{j,\gamma}} \mathbf{M} \right] \boldsymbol{\mu} . \tag{3.7.13}$$

The same simplifications described in Eqs. (2.2.6) through (2.2.9) are applied to Eq. (3.7.13), except that in this case the first and third terms are not equivalent and cannot be combined

$$\left(\frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}}\right)^{t} \boldsymbol{\mu} = -\left(\boldsymbol{\Phi}^{\mu}\right)^{t} \frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial s_{j,\gamma}} \mathbf{M} - \frac{1}{4\pi} \left(\boldsymbol{\Phi}_{\mathrm{s}}^{\mu}\right)^{t} \frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} - \boldsymbol{\mu}^{t} \frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}^{\mathrm{M}} . \tag{3.7.14}$$

The fourth term on the RHS of Eq. (3.7.10) leads to a result analogous to Eq. (3.7.14) using similar arguments

$$\boldsymbol{\nu}^{t} \frac{\partial \mathbf{E}_{\mathrm{RF}}^{\mathrm{M}}}{\partial s_{j,\gamma}} = -\left(\boldsymbol{\Phi}^{\nu}\right)^{t} \frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial s_{j,\gamma}} \mathbf{M} - \frac{1}{4\pi} \left(\boldsymbol{\Phi}_{\mathrm{s}}^{\mu}\right)^{t} \frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} - \boldsymbol{\nu}^{t} \frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}^{\mathrm{M}} . \tag{3.7.15}$$

3.7.2 Mutual Polarization Energy Gradient

In addition to the implicit difference due to the induced dipoles being converged self-consistently, the full mutual polarization gradient includes an additional contribution beyond the direct polarization gradient. Specifically, the derivative of the matrix **C** leads to four terms

$$\begin{aligned}\boldsymbol{\nu}^t \frac{\partial \mathbf{C}}{\partial s_{j,\gamma}} \boldsymbol{\mu} = \boldsymbol{\nu}^t \Bigg[& \frac{\partial \mathbf{T}^{(11)}}{\partial s_{j,\gamma}} + 4\pi \frac{\partial \mathbf{D}_{\mathrm{B}}^t}{\partial s_{j,\gamma}} \left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1} \right) \mathbf{D}_{\mathrm{B}} + 4\pi \mathbf{D}_{\mathrm{B}}^t \frac{\partial \mathbf{A}_{\mathrm{s}}^{-1}}{\partial s_{j,\gamma}} \mathbf{D}_{\mathrm{B}} \\ & + 4\pi \mathbf{D}_{\mathrm{B}}^t \left(\mathbf{A}_{\mathrm{s}}^{-1} - \mathbf{A}_{\mathrm{v}}^{-1} \right) \frac{\partial \mathbf{D}_{\mathrm{B}}}{\partial s_{j,\gamma}} \Bigg] \boldsymbol{\mu}\end{aligned} \tag{3.7.16}$$

The first term on the RHS occurs in vacuum and is described elsewhere[2, 7], while the final three terms are specific to LPBE calculations. Using the simplifications described in Eqs. (2.2.6) through (2.2.9) results in

$$\begin{aligned}\boldsymbol{\nu}^t \frac{\partial \mathbf{C}}{\partial s_{j,\gamma}} \boldsymbol{\mu} = & \boldsymbol{\nu}^t \frac{\partial \mathbf{T}^{(11)}}{\partial s_{j,\gamma}} \boldsymbol{\mu} - \boldsymbol{\nu}^t \frac{\partial \mathbf{D}_{\mathrm{B}}^t}{\partial s_{j,\gamma}} \boldsymbol{\Phi}^{\mu} \\ & - \frac{1}{4\pi} \left(\boldsymbol{\Phi}_{\mathrm{s}}^{\nu} \right)^t \frac{\partial \mathbf{A}_{\mathrm{s}}}{\partial s_{j,\gamma}} \boldsymbol{\Phi}_{\mathrm{s}}^{\mu} - \left(\boldsymbol{\Phi}^{\nu} \right)^t \frac{\partial \mathbf{D}_{\mathrm{B}}}{\partial s_{j,\gamma}} \boldsymbol{\mu}\end{aligned} . \tag{3.7.17}$$

Substitution of Eqs. (3.7.14), (3.7.15) and (3.7.17) into Eq. (3.7.9) gives the total mutual polarization energy gradient for an AMOEBA solute interacting self-consistently with the PMPB continuum.

$$
\begin{aligned}
\frac{\partial \mathrm{U}_{\mu}}{\partial s_{j,\gamma}} = & -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial s_{j,\gamma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial s_{j,\gamma}}\right] - \frac{1}{2}\boldsymbol{\nu}^{t}\frac{\partial \mathbf{T}^{(11)}}{\partial s_{j,\gamma}}\boldsymbol{\mu} \\
& + \frac{1}{2}\left[\left(\boldsymbol{\Phi}^{\mu}\right)^{t}\frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial s_{j,\gamma}}\mathbf{M} + \boldsymbol{\mu}^{t}\frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}^{\mathrm{M}}\right] + \frac{1}{8\pi}\left(\boldsymbol{\Phi}_{\mathrm{s}}^{\mu}\right)^{t}\frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} \\
& + \frac{1}{2}\left[\left(\boldsymbol{\Phi}^{\nu}\right)^{t}\frac{\partial \mathbf{T}_{\mathrm{B}}}{\partial s_{j,\gamma}}\mathbf{M} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}^{\mathrm{M}}\right] + \frac{1}{8\pi}\left(\boldsymbol{\Phi}_{\mathrm{s}}^{\nu}\right)^{t}\frac{\partial \mathbf{A}_{s}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}} \\
& + \frac{1}{2}\left[\left(\boldsymbol{\Phi}^{\nu}\right)^{t}\frac{\partial \mathbf{D}_{\mathrm{B}}}{\partial s_{j,\gamma}}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{D}_{\mathrm{B}}^{t}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}^{\mu}\right] + \frac{1}{8\pi}\left(\boldsymbol{\Phi}_{\mathrm{s}}^{\nu}\right)^{t}\frac{\partial \mathbf{A}_{\mathrm{s}}}{\partial s_{j,\gamma}}\boldsymbol{\Phi}_{\mathrm{s}}^{\mu}
\end{aligned}
\tag{3.7.18}
$$

The first two terms (where each set of square brackets will be considered a single term) are evaluated even in the absence of continuum solvent, although in this case $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ have been converged in a self-consistent field that includes continuum contributions. The remaining terms are analogous to those found in Poisson-Boltzmann calculations involving only permanent electrostatics. The number of LPBE calculations required for evaluation of the energy gradient includes two for the permanent multipoles and two each for $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ at each SOR convergence step. Further details on the numerical implementation of Eq. (3.7.18) can be found in the third section of the Appendix.

3.8 PMPB Validation and Application

This section presents useful benchmarks for demonstrating the expected numerical precision of the present work. Our first goal is to compare against analytical results for a source charge distribution described by a single charge, dipole, polarizable dipole or quadrupole located at the center of a low dielectric sphere in high dielectric solvent. The transition between solute and solvent is initially specified using a step function, and then subsequently using a smooth transition described by a heptic polynomial. We then compare analytic gradients to those determined using finite-differences of the energy for a variety of two sphere systems to isolate the reaction field, dielectric boundary and ionic boundary gradients for the permanent multipole solvation energy, the direct polarization model and the mutual polarization model. Finally, the method is applied to a series of proteins and comparisons are made to corresponding simulations in explicit water.

3.8.1 Energy

The numerical accuracy of the multipole discretization procedure was studied by comparison to analytical solutions of the Poisson equation for a monopole, dipole, polarizable dipole and quadrupole located within a spherical cavity of radius 3.0 Å. The monopole case, or Born ion[26], has a well-known analytical solution

$$\mathrm{U}_q = \frac{1}{2}\left(\frac{1}{\varepsilon} - 1\right)\frac{q^2}{a} \tag{3.8.1}$$

where q is the charge magnitude, a the cavity radius and ε is the solvent dielectric. For a permanent dipole, the analogous solution was used by Onsager[5],

$$\mathrm{U}_d = -\frac{1}{2}\left[\frac{2(\varepsilon - 1)}{1 + 2\varepsilon}\right]\frac{\boldsymbol{d}\cdot\boldsymbol{d}}{a^3}, \tag{3.8.2}$$

where $\boldsymbol{d}$ is the dipole vector. For a polarizable dipole, the energy is the sum of two contributions, the cost of polarization and the energy of the total dipole in the total reaction field[85, 86]

$$\mathrm{U}_{\alpha,d} = -\frac{1}{2}\frac{f\boldsymbol{d}\cdot\boldsymbol{d}}{1 - f\alpha}, \tag{3.8.3}$$

where α is the polarizability, $\boldsymbol{d}$ is the permanent dipole and f is the reaction field factor

$$f = \frac{1}{a^3}\left[\frac{2(\varepsilon - 1)}{1 + 2\varepsilon}\right]. \tag{3.8.4}$$

The analytic solution for the self-energy of a traceless Cartesian quadrupole can be derived beginning from the energy of a quadrupole in an electric field gradient

$$\mathrm{U}_\Theta = \frac{\Theta_{\gamma\delta}}{3}\nabla_\delta\nabla_\gamma\Phi, \tag{3.8.5}$$

where we are summing over the subscripts γ and δ, and the factor of 1/3 is due to use of traceless quadrupoles.[78] To determine the needed reaction field gradient, which for the moment will be assumed to come from any quadrupole component and not necessarily be a self-interaction, we begin from the reaction potential inside the cavity

$$\Phi_{\Theta} = -\frac{\Theta_{\alpha\beta}}{3}\left[\frac{3(\varepsilon-1)}{2+3\varepsilon}\right]\frac{3r_{\alpha}r_{\beta}}{a^{5}} \tag{3.8.6}$$

and take the first derivative

$$\nabla_{\gamma}\Phi_{\Theta} = -\frac{\Theta_{\alpha\beta}}{3}\left[\frac{3(\varepsilon-1)}{2+3\varepsilon}\right]\frac{3\left(r_{\beta}\delta_{\alpha\gamma}+r_{\alpha}\delta_{\beta\gamma}\right)}{a^{5}}, \tag{3.8.7}$$

followed by a second differentiation to achieve the reaction field gradient

$$\nabla_{\delta}\nabla_{\gamma}\Phi_{\Theta} = -\frac{\Theta_{\alpha\beta}}{3}\left[\frac{3(\varepsilon-1)}{2+3\varepsilon}\right]\frac{3\left(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\beta\gamma}\delta_{\alpha\delta}\right)}{a^{5}}. \tag{3.8.8}$$

Substituting Eq. (3.8.8) into Eq. (3.8.5) and taking into account that half the energy is lost due to polarizing the continuum gives the self-energy of a traceless quadrupole in its own reaction field gradient

$$\mathrm{U}_{\Theta} = -\frac{1}{2}\left[\frac{3(\varepsilon-1)}{2+3\varepsilon}\right]\frac{\Theta_{\alpha\beta}\left(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\beta\gamma}\delta_{\alpha\delta}\right)\Theta_{\gamma\delta}}{3a^{5}}. \tag{3.8.9}$$

From Eq. (3.8.9), it is seen that all non-self interactions, for example Θ_{xx} with Θ_{yy}, are zero, and the quadrupole self-energy is simply the sum of nine terms,

$$\mathrm{U}_{\Theta} = -\frac{1}{2}\left[\frac{3(\varepsilon-1)}{2+3\varepsilon}\right]\frac{2\Theta_{\alpha\beta}^{2}}{3a^{5}}. \tag{3.8.10}$$

The first series of numerical tests used a step function at the dielectric boundary rather than the smooth transition that is required for continuous energy gradients described previously. This simplification is necessary in order to compare the known analytic results directly with the numerical solver. In each case, the solution domain was

a 10.0 Å cube with the low-dielectric sphere located at the center. In Table 3.4 it is shown that each test case converges toward the analytic result as grid spacing is decreased.

Table 3.4. As grid spacing decreases, the numerical solution to the PE approaches the analytic solution for four canonical test cases including a charge, dipole, polarizable dipole and quadrupole. Each test case involved a 3 Å sphere of dielectric 1 and solvent dielectric of 78.3 with a step-function transition between solute and solvent (kcal/mole).

Grid Points	**Grid Spacing**	**Charge**	**Dipole**	**Polarizable Dipole**	**Quadrupole**
33 x 33 x 33	0.313	-55.6514	-5.3556	-5.8011	-1.8487
65 x 65 x 65	0.156	-54.9150	-5.1450	-5.5548	-1.7211
129 x 129 x 129	0.078	-54.8024	-5.1134	-5.5180	-1.7038
225 x 225 x 225	0.045	-54.7236	-5.0915	-5.4925	-1.6912
Analytic	-	-54.6355	-5.0675	-5.4645	-1.6783

Our next goal was to determine the energy change due to introduction of a smooth dielectric boundary with a window width of 0.6 Å. Using a grid spacing slightly less than 0.1 Å, it can be seen in Table 3.5 that the smooth dielectric boundary increases the solvation energy over the analogous step function boundary. By increasing the radius of the low dielectric cavity by approximately 0.2 Å, the energy of the charge, dipole, polarizable dipole and quadrupole can be adjusted to simultaneously mimic the known analytic results.

Table 3.5. The tests from Table 3.4 are repeated using 129 grid points (0.078 Å spacing), however, the transition between solute and solvent is defined by a 7th order polynomial, which acts over a total window width of 0.6 Å. Increasing the radius of the low dielectric sphere by approximately 0.2 Å raises the energies to mimic the step function transition results (kcal/mole).

Radius Increase	**Charge**	**Dipole**	**Polarizable Dipole**	**Quadrupole**
0.0	-58.4926	-6.2126	-6.8202	-2.3555
0.1	-56.4762	-5.5922	-6.0798	-1.9767
0.2	-54.5941	-5.0518	-5.4463	-1.6687
Step Function	-54.8024	-5.1134	-5.5180	-1.7038

3.8.2 Energy Gradient

Our first goal is to show that the energy gradient is continuous for higher order moments as a result of using 5th order B-splines. This is seen in Figure 3.5 through Figure 3.7 for a charge, dipole and quadrupole interacting with a neutral cavity, respectively. It is also clear that the sum of the forces between the neutral and charged site (and a third reference site that defines the local multipole frame in the cases of the dipole and quadrupole) is zero, indicating conservation of energy.

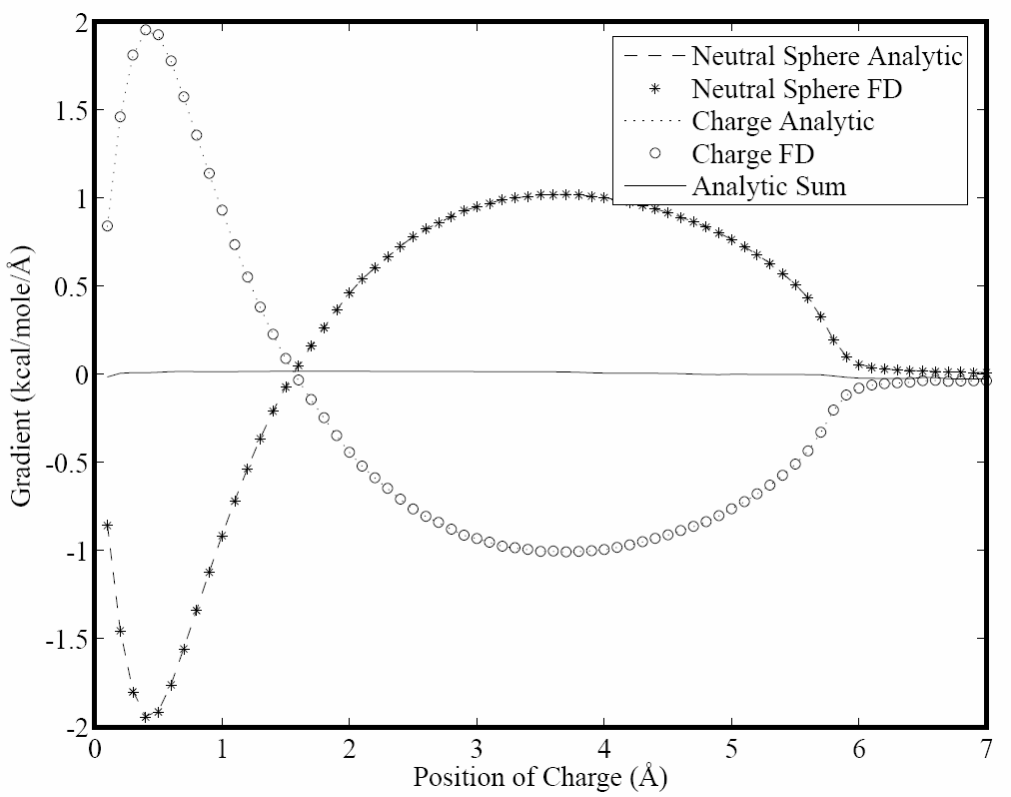

Figure 3.5. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with unit positive charge vs. separation. Both spheres have a radius of 3.0 Å and the solvent dielectric is 78.3. The gradient of the neutral cavity is due entirely to the dielectric boundary force and cancels exactly the force on the charged sphere.

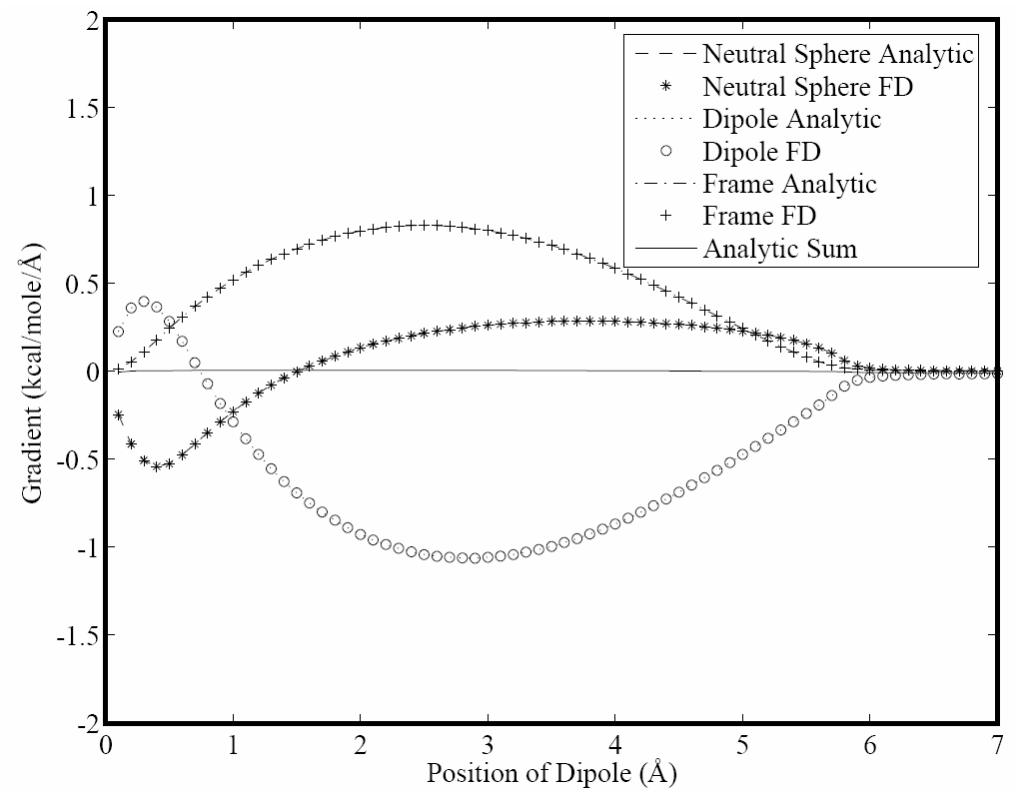

Figure 3.6. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with dipole moment components of (2.54, 2.54, 2.54) debye vs. separation. Both spheres have a radius of 3.0 Å and movement of the dipole is along the x-axis. The gradient of the neutral cavity is due entirely to the dielectric boundary force and cancels exactly the sum of the forces on the dipole and a third site (that has no charge density or dielectric properties) that defines the local coordinate system of the dipole.

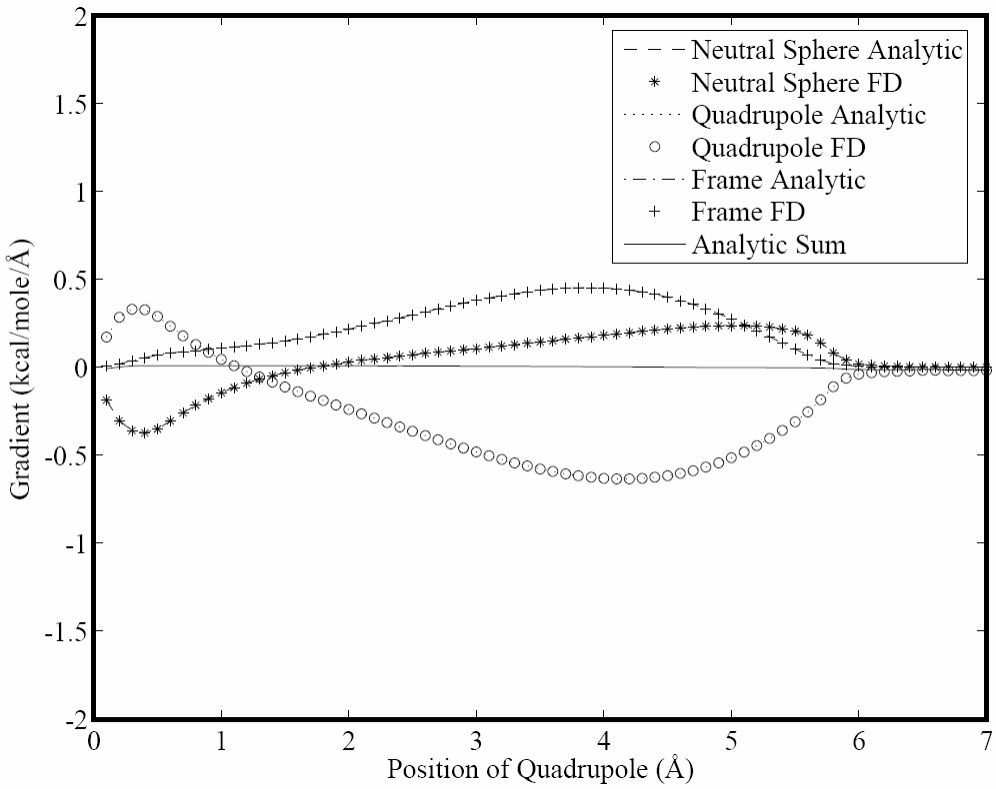

Figure 3.7. Analytic and finite-difference gradients for a neutral cavity fixed at the origin and a sphere with quadrupole moment components of (5.38, 2.69, 2.69, 2.69, -2.69, 2.69, 2.69, 2.69, -2.69) Buckinghams vs. separation. Both spheres have a radius of 3.0 Å and movement of the quadrupole is along the x-axis. The gradient of the neutral cavity cancels exactly the sum of the forces on the quadrupole and a third site (that has no charge density or dielectric properties) that defines the local coordinate system of the quadrupole.

Similarly, the reaction field and dielectric boundary gradients of the polarization energy for both the direct and mutual models are smooth and demonstrate conservation of energy, as shown in Figure 3.8 and Figure 3.9, respectively. Finally, it is clear that polarization catastrophes are avoided even when a charged site is moved toward superimposition with a polarizable site, due to use of a modified Thole model that damps mutual polarization at short range.[20]

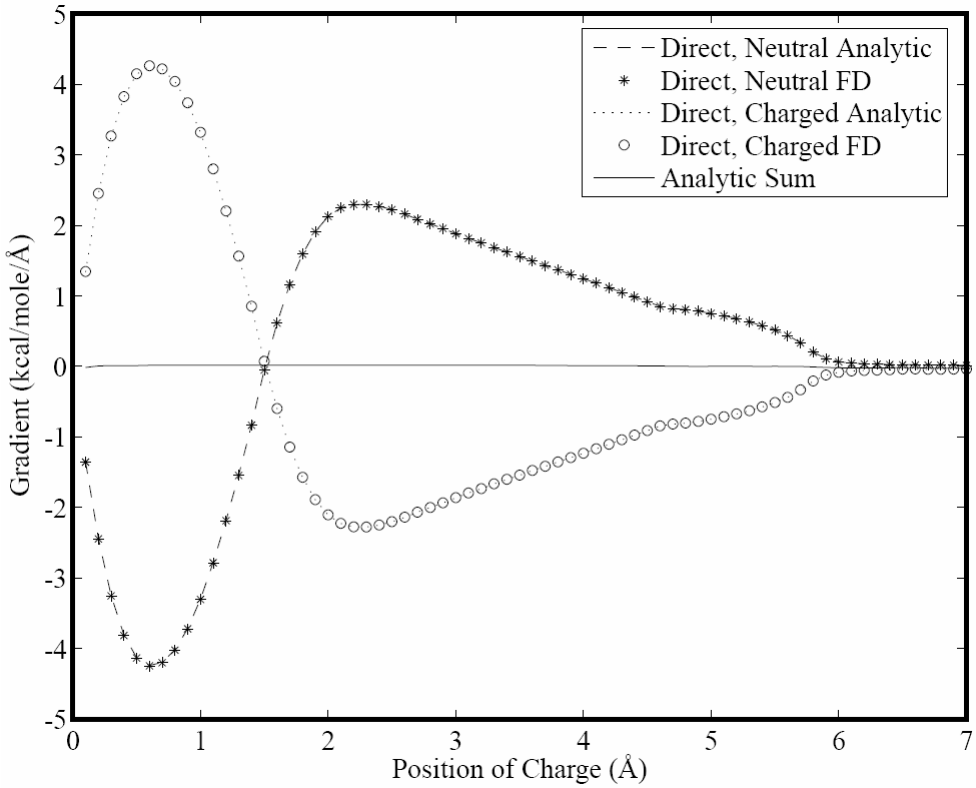

Figure 3.8. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin and a sphere with unit positive charge vs. separation using the *direct* polarization model. Both spheres have a radius of 3.0 Å. The gradient can be seen to approach zero at a number of points, notably when the spheres are separated by approximately 1.5 Å leading to a maximum in the reaction field produced by the charge at the polarizable site, and again when the spheres are superimposed and the reaction field is zero at the polarizable site.

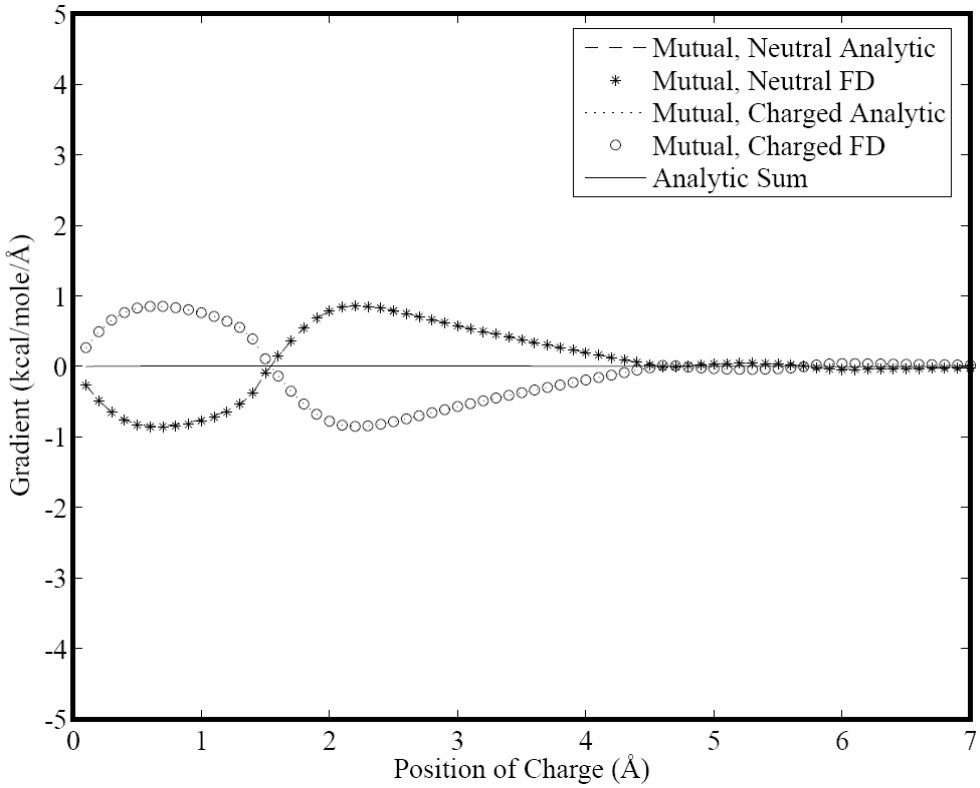

Figure 3.9. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin and a polarizable sphere with unit positive charge vs. separation using the *mutual* polarization model. Both spheres have a radius of 3.0 Å and a polarizability of 1.0 $Å^{-3}$. Note that the mutual polarization gradients are smaller than those in Fig 8. for the otherwise equivalent *direct* polarization model.

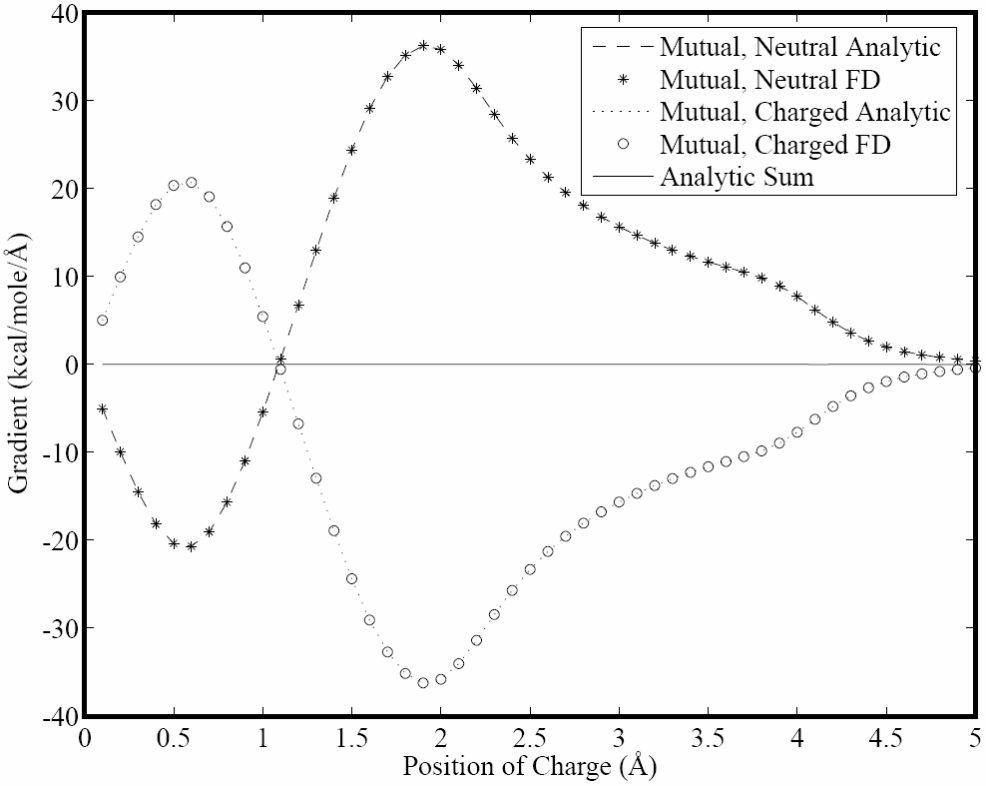

Figure 3.10. The dielectric of the solvent and test spheres are both set to 1 in this case, while a salt concentration of 150 mM is used to isolate the ionic boundary gradients. Analytic and finite-difference gradients for a neutral, polarizable cavity fixed at the origin (3.0 Å radius) and a polarizable sphere with a unit positive charge (1.0 Å radius) vs. separation using the mutual polarization model. Both spheres have a polarizability of 1.0 $Å^{-3}$, and the ionic radius is set to 0.0 Å.

3.8.3 The Electrostatic Response of Solvated Proteins

As described in the introduction, a motivation for the current work is study of polar macromolecules by an improved electrostatic model within an empirical molecular mechanics framework. From explicit water simulations it is possible to measure the total dipole moment of a solvated protein in a fixed folded conformation by sampling over the

water degrees of freedom. The resulting ensemble average electrostatic response can then be directly compared to the PMPB model.

Simulations of five proteins taken from the Protein Databank[87] (1CRN[88], 1ENH[89], 1FSV[90], 1PGB[91] and 1VII[92]) were equilibrated under NPT conditions (1 atm, 298 K) using a standard protocol. Formal charge and system size are given in Table 3.6. A single snapshot for each protein system was taken from equilibrated molecular dynamics simulations using the AMOEBA force field. The protein coordinates were frozen, and sampling of the solvent degrees of freedom continued for 150 psec under the same NPT conditions, with the first 50 psec discarded prior to analysis. For all simulations the Berendsen weak coupling thermostat and barostat were employed with time constants of 0.1 and 2.0 psec, respectively.[93] Long range electrostatics were treated using particle mesh Ewald (PME) summation with a cutoff for real space interactions of 7.0 Å and an Ewald coefficient of 0.54 $Å^{-1}$.[94] The PME methodology used tinfoil boundary conditions, a 54 x 54 x 54 charge grid and 6th order B-spline interpolation. van der Waals interactions were smoothly truncated to zero at 12.0 Å using a switching window of width 1.2 Å. Simulations were run using TINKER version 4.2.[95]

Table 3.6. Synopsis of the protein systems studied in explicit and continuum solvent.

		Number of Atoms	
Protein	**Formal Charge**	**Protein**	**Protein +Water**
CRN	0	642	4980
ENH	+7	947	5039
FSV	+5	504	6435
PGB	-4	855	6143
VII	+2	596	4271

The same conformation of each protein studied in explicit water was examined using the LPBE methodology developed in this work at a range of grid spacings using the direct and mutual polarization models. In addition, 150 mM electrolyte was used in conjunction with the mutual polarization model to determine the relative effect of salt on the electrostatic response. The results are summarized in Table 3.7. Similar to the analytic test cases, as grid spacing is reduced the total electrostatic energy rises monotonically toward the converged solution.

Table 3.7. The energy (kcal/mole) and dipole moment (debye) of each protein system was studied using a range of grid spacings under the direct polarization model, mutual polarization model, and mutual polarization model with 150 mM salt. The cavity was defined using AMOEBA R_{min} values for each atom and smooth dielectric and ionic boundaries via a total window width of 0.6 Å.

		Direct Polarization		**Mutual Polarization**		**150 mM Salt**	
Protein	**Grid Spacing**	**Energy**	μ	**Energy**	μ	**Energy**	μ
CRN	0.61	-597.4	83.9	-679.1	81.0	-680.9	81.6
	0.31	-563.3	83.4	-641.3	80.6	-643.0	81.1
	0.18	-554.7	83.4	-632.1	80.6	-633.8	81.1
ENH	0.63	-1892.6	265.2	-2055.1	265.1	-2067.5	266.8
	0.32	-1851.1	265.8	-2008.8	265.8	-2021.2	267.4
	0.18	-1834.9	265.7	-1991.1	265.7	-2003.6	267.3
FSV	0.66	-1207.0	208.1	-1293.8	215.7	-1301.0	216.3
	0.33	-1184.3	208.4	-1269.3	216.0	-1276.5	216.6
	0.19	-1173.1	208.4	-1257.1	215.9	-1264.3	216.5
PGB	0.71	-1327.7	128.4	-1453.5	132.7	-1458.9	133.4
	0.36	-1275.7	127.8	-1400.5	132.0	-1405.9	132.6
	0.20	-1259.3	127.7	-1380.3	131.9	-1385.7	132.5
VII	0.62	-902.4	194.2	-1009.8	197.1	-1014.3	198.1
	0.31	-866.0	194.4	-970.6	197.3	-975.0	198.3
	0.18	-858.3	194.3	-962.0	197.2	-966.5	198.2

The total dipole moments are less sensitive to grid spacing than are the energies, with little change observed in moving from 0.3 Å to 0.2 Å. Adding 150 mM salt lowers the electrostatic energy by 1.7-12.5 kcal/mole at the smallest grid spacing studied, with CRN (neutral) and ENH (+7) showing the smallest and largest response, respectively. The magnitude of the energetic change indicates that salt concentration plays an important role in protein energetics, especially for highly charged species. For these calculations we have chosen an ionic radius of 2.0 Å, however smaller or larger values increase or decrease the energetic response, respectively.

Finally, we compare the increase in dipole moment between the explicit water simulations and the continuum LPBE environment for each protein. As shown in Table 3.8, both the direct and mutual models lead to total moments that are in good agreement with those found by molecular dynamics sampling of explicit water degrees of freedom. On average, the dipole moment increased by a factor of 1.27 in explicit water and 1.26 using the mutual polarization model. This result, which was achieved without detailed parameterization of atomic radii (AMOEBA Buffered-14-7 R_{min} values were used), indicates that at the length scale of whole proteins the continuum assumption is justified. Timings and memory requirements for the LPBE calculations as a function of grid size are shown in Table 3.9.

Table 3.8. The dipole moment (debye) of each protein in vacuum μ_v, under the direct and mutual polarization models interacting with a continuum of permittivity 78.3, and in explicit water. Ensemble averages were taken over 100 psec trajectories and each has a std. err. of less than ± 0.3. The ratio of the solvated to vacuum dipole moment is given in each case. The cavity was defined using AMOEBA R_{min} values for each atom and smooth dielectric and ionic boundaries via a total window width of 0.6 Å.

	Vacuum	**Direct Polarization**		**Mutual Polarization**		**Explicit Water**	
Protein	μ_v	μ	μ/μ_v	μ	μ/μ_v	$<\mu>$	$<\mu>/\mu_v$
CRN	62.1	83.4	1.34	80.6	1.30	81.8	1.32
ENH	208.3	265.7	1.28	265.7	1.28	267.0	1.28
FSV	184.7	208.4	1.13	215.9	1.17	213.5	1.16
PGB	101.4	127.7	1.26	131.9	1.30	134.3	1.32
VII	158.3	194.4	1.23	197.3	1.25	197.7	1.25
Average	143.0	175.9	1.25	178.3	1.26	178.9	1.27

Table 3.9. Memory requirements and wall clock timings for each protein system are shown. All calculations were run on a 2.4 Ghz Opteron.

Protein	Cubic Box Size	Grid Points	Memory (MB)	Direct (s)	Mutual (s)
CRN	39.31	65	84	6.9	50.2
		129	487	34.5	276.8
		225	2027	189.3	1414.7
ENH	40.50	65	81	9.2	80.3
		129	491	45.5	414.7
		225	1983	253.0	2457.3
FSV	42.35	65	73	5.7	42.7
		129	487	31.9	234.6
		225	2040	188.1	1463.8
PGB	45.48	65	80	8.2	60.7
		129	375	30.6	230.9
		225	1880	156.9	1176.5
VII	39.47	65	73	6.4	65.2
		129	487	37.2	360.6
		225	2040	194.8	2062.8

4 Generalized Kirkwood

The description of GK will be subdivided into five sections. First, determination of the self-energy for a permanent multipole will be considered. Second, we will propose a functional form for the cross-term energy between arbitrary order multipole moments. Third, we suggest a factoring of the resulting tensors that facilitates their generation up to arbitrary order. Fourth, given the underlying GK theory, we continue on to the derivation of the electrostatic solvation energy and gradient in the specific case of solutes described by the AMOEBA force field. Finally, we apply the GK continuum model to 55 proteins and compare their electrostatic solvation free energy and total dipole moment to analogous calculations with the PMPB continuum.

4.1 Effective Radii and the Multipole Self-Energy

We begin by reiterating that the self-energy of a multipole depends on not only the reaction potential, but on the reaction field, the reaction field gradient, and so on. Unlike GB, the perfect effective radius is simply not enough information to guarantee the higher order features of the reaction potential are correct, unless the multipole site happens to be at the center of a spherical cavity. Two methods have been investigated to

describe the self-energy of a permanent atomic multipole. The first method reduces to the Coulomb-field approximation (CFA) for a monopole and requires knowledge of the analytic solution for the field in solvent based on a multipole at the center of a spherical dielectric cavity.[27] We term this the solvent field approximation (SFA), as it is consistent with the CFA, but requires more information. A second approach makes use of Grycuk's method for determining effective radii based on the reaction potential of an off-center charge within a spherical solute.[96] We refer to this approach as the reaction potential approximation (RPA).

Before detailing the SFA and RPA methods, a brief introduction to the electrostatic energy of a dielectric media will be given. The work required to assemble a fixed charge distribution in a linearly polarizable medium[54, 85, 86, 97] can be formulated by a volume integral of the product of the charge density $\rho(r)$ with the potential $\phi(r)$ or by the scalar product of the electric field **E** with the electric displacement **D**

$$\begin{aligned} W &= \frac{1}{2}\int_V \rho(r)\phi(r)\,dV \\ &= \frac{1}{8\pi}\int_V \mathbf{E}\cdot\mathbf{D}\,dV \end{aligned} \tag{4.1.1}$$

where we have assumed that $\rho(r)$ is localized and the displacement is proportional to the electric field in regions of constant permittivity ε

$$\mathbf{D} = \varepsilon\mathbf{E}\,. \tag{4.1.2}$$

For our purposes, the system of interest is composed of a solute with a different permittivity than the solvent. The electrostatic free energy of this system relative to a homogeneous reference state is

$$\Delta G = \frac{1}{8\pi}\int_V \left(\mathbf{E}\cdot\mathbf{D} - \mathbf{E}_h\cdot\mathbf{D}_h\right)dV \tag{4.1.3}$$

where in the homogeneous case the field is Coulombic and can be defined relative to the vacuum field as $\mathbf{E}_h = \mathbf{E}_{vac}/\varepsilon_h$ using the homogeneous permittivity ε_h and therefore the homogeneous displacement is simply $\mathbf{D}_h = \mathbf{E}_{vac}$. A less intuitive, but equivalent definition of the electrostatic free energy given in Eq. (4.1.3) is[54, 97]

$$\Delta G = \frac{1}{8\pi}\int_V \left(\mathbf{E}\cdot\mathbf{D}_h - \mathbf{D}\cdot\mathbf{E}_h\right)dV \ . \tag{4.1.4}$$

This expression can be subdivided into integrals over the solute and solvent as

$$\Delta G = \frac{1}{8\pi}\int_{solute}\left(\mathbf{E}\cdot\mathbf{D}_h - \mathbf{D}\cdot\mathbf{E}_h\right)dV + \frac{1}{8\pi}\int_{solvent}\left(\mathbf{E}\cdot\mathbf{D}_h - \mathbf{D}\cdot\mathbf{E}_h\right)dV \ . \tag{4.1.5}$$

In both the homogeneous and mixed permittivity states the solute retains the homogeneous permittivity. By using the relationships for the homogeneous field and displacement described above it can be seen that the integral over the solute vanishes

$$\begin{aligned}\Delta G = &\frac{1}{8\pi}\int_{solute}\left(\mathbf{E}\cdot\mathbf{E}_{vac} - \varepsilon_h\mathbf{E}\cdot\mathbf{E}_{vac}/\varepsilon_h\right)dV \\ &+\frac{1}{8\pi}\int_{solvent}\left(\mathbf{E}_s\cdot\mathbf{E}_{vac} - \varepsilon_s\mathbf{E}_s\cdot\mathbf{E}_{vac}/\varepsilon_h\right)dV\end{aligned} \tag{4.1.6}$$

to leave only the integral over the solvent

$$\Delta G = \frac{1}{8\pi}\left(1 - \frac{\varepsilon_s}{\varepsilon_h}\right) \int_{solvent} \left(\mathbf{E}_s \cdot \mathbf{E}_{vac}\right) dV \,. \tag{4.1.7}$$

Having made no assumptions to this point, the remaining challenge can be simplified to defining the field within the solvent $\mathbf{E}_s$ for the mixed permittivity case. This is the starting point for the SFA. In general, the solvent field does not have an exact analytic form for a union of spheres. However, many molecular systems of interest are globular, and therefore an approximation based on the assumption of a spherical solute is not only qualitatively reasonable, but in many cases quantitative.

4.1.1 The Solvent Field Approximation

The SFA is similar to the CFA, but is based on evaluating Eq. (4.1.7) using Kirkwood's solution for the field outside a spherical solute with a central multipole moment[27]

$$\mathbf{E}_s = \sum_{l=0}^{\infty} \frac{\left(2l+1\right)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h} \mathbf{E}_{vac}^{(l)} \tag{4.1.8}$$

where $\mathbf{E}_{vac}^{(l)}$ is the vacuum field due to all multipole moments of degree l, defined using either irregular spherical harmonics or Cartesian tensors. Throughout the current work we neglect salt effects, although their addition to a future GK formulation is straightforward. This definition of the self-energy is equivalent to the CFA for a monopole and becomes approximate for off-center multipole sites or for non-spherical solute geometries.

Under the SFA, the self-energy of a permanent multipole site i is given by

$$\Delta G_i^{SFA} = \frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{solvent} \mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV\,. \tag{4.1.9}$$

It is possible to invert the integration domain by adding and subtracting an integral over the solute region outside the radius R_i of atom i to Eq. (4.1.9) giving

$$\begin{aligned}\Delta G_i^{SFA} = &\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{r>R_i} \mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV \\ &-\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{\substack{solute,\\ r>R_i}} \mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV\end{aligned}\,. \tag{4.1.10}$$

The first integral is the solvation energy of a lone multipole ΔG_i^M and the second represents the effect of descreening sites. Substituting ΔG_i^M into Eq. (4.1.10) gives

$$\Delta G_i^{SFA} = \Delta G_i^M - \frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\underbrace{\int_{\substack{solute,\\ r>R_i}} \mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV}_{I_i} \tag{4.1.11}$$

where

$$\Delta G_i^M = \frac{1}{2}\left[c_0\frac{q_i^2}{a_i} + c_1\frac{\mu_{i,\alpha}^2}{a_i^3} + c_2\frac{2}{3}\frac{\Theta_{i,\alpha\beta}^2}{a_i^5}\right] \tag{4.1.12}$$

and

$$c_l = \frac{1}{\varepsilon_h}\frac{(l+1)(\varepsilon_h-\varepsilon_s)}{(l+1)\varepsilon_s + l\varepsilon_h}\,. \tag{4.1.13}$$

In Eq. (4.1.12) we have assumed the Einstein convention for summation over Greek subscripts α and β, which can take the value x, y, or z. The descreening integral I_i can be decomposed into a sum of pairwise integrals I_{ij}[32, 33]

$$
\begin{aligned}
I_i\left(r_{ij}, R_i, R_j\right) &= \sum_{j\neq i} \int \int_0^{\xi_{ij}} \int_0^{2\pi} \mathbf{E}_{vac,i} \cdot \sum_{l=0}^{\infty} \left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h} \mathbf{E}_{vac,i}^{(l)} \right) r^2 \sin\theta d\phi d\theta dr \\
&= \sum_{j\neq i} I_{ij}\left(r_{ij}, R_i, R_j\right)
\end{aligned}
\tag{4.1.14}
$$

where ξ_{ij} is the angle formed between the pairwise axis and any ray that begins at the center of atom i and passes through the circle of intersection between the integration shell and atom j

$$
\xi_{ij} = \cos^{-1}\left(\frac{r_{ij}^2 - R_j^2 + r^2}{2 r_{ij} r} \right) \tag{4.1.15}
$$

where r_{ij} is the distance between atoms i and j, R_j is the radius of atom j and r is the radial integration variable. The integration limits for the radial coordinate depend on what extent atoms i and j intersect, and therefore the solution to Eq. (4.1.14) is presented as an indefinite integral that is to be evaluated at limits described below. Typically the radius of the descreening atom is scaled down to prevent over counting due to atomic overlap, although parameter free approaches are being explored.[48, 98]

Unlike the field due to a partial charge, the field due to a multipole of arbitrary order has an angular dependence. Our approach has been to represent the field using a spherical harmonic basis, rather than Cartesian tensors, to determine the analytic solution to Eq. (4.1.14) through quadrupole order. Additionally, it is assumed that the positive z-axis of the multipole frame is directed towards the center of the descreening atom. This imposes symmetry that greatly reduces the number of non-vanishing terms in the solution, but requires rotation of multipole moments for each pairwise descreening interaction.

A complex definition of spherical harmonics is commonly used in the formulation of quantum mechanics, however this work uses the following real form

$$Y_l^{(m)}(\theta,\phi)=\begin{cases}(-1)^m\sqrt{2}\sqrt{\dfrac{(l-m)!}{(l+m)!}}P_l^{(m)}(\cos\theta)\cos m\phi & m>0\\ \sqrt{\dfrac{(l-m)!}{(l+m)!}}P_l^{(m)}(\cos\theta) & m=0\\ (-1)^{|m|}\sqrt{2}\sqrt{\dfrac{(l-|m|)!}{(l+|m|)!}}P_l^{(|m|)}(\cos\theta)\sin|m|\phi & m<0\end{cases} \tag{4.1.16}$$

where $Y_l^{(m)}(\theta,\phi)$ is of degree $l\geq 0$ and order $|m|\leq l$, $P_l^{(m)}$ are the associated Legendre polynomials, the polar angle ranges from $0\leq\theta\leq\pi$ and the azimuth ranges from $0\leq\varphi\leq 2\pi$. We chose to use the Racah normalization, which has the property that $Y_l^{(0)}(0,0)=1$. In combination with our choice of phase factors, this ensures formulas for the conversion between Cartesian multipole moments and those consistent with this definition of real spherical harmonics are identical to the conversions commonly used for complex spherical harmonics. The conversion formulas through quadrupole degree are given in Table 4.1.[78]

Table 4.1. Multipole moment conversions.

$Q_0^{(0)} = q$
$Q_1^{(0)} = \mu_z$
$Q_1^{(1)} = \mu_x$
$Q_1^{(-1)} = \mu_y$
$Q_2^{(0)} = \Theta_{zz}$
$Q_2^{(1)} = \frac{2}{\sqrt{3}}\Theta_{xz}$
$Q_2^{(-1)} = \frac{2}{\sqrt{3}}\Theta_{yz}$
$Q_2^{(2)} = \frac{1}{\sqrt{3}}\left(\Theta_{xx} - \Theta_{yy}\right)$
$Q_2^{(-2)} = \frac{2}{\sqrt{3}}\Theta_{xy}$

The potential due to a unit magnitude multipole moment $\Phi_l^{(m)}\left(r,\theta,\phi\right)$ is obtained by multiplication of the real spherical harmonics by a radial factor of $1/r^{l+1}$ to give

$$\Phi_l^{(m)}\left(r,\theta,\phi\right) = \frac{Y_l^{(m)}\left(\theta,\phi\right)}{r^{l+1}} \tag{4.1.17}$$

and are listed in Table 4.2 through quadrupole order.

Table 4.2. Unit vacuum potentials.

l	m	$\Phi_l^{(m)}(r,\theta,\phi)$
0	0	$\frac{1}{r}$
1	0	$\frac{\cos\theta}{r^2}$
	1	$\frac{\sin\theta\cos\phi}{r^2}$
	-1	$\frac{\sin\theta\sin\phi}{r^2}$
2	0	$\frac{1}{2}\frac{3\cos^2\theta-1}{r^3}$
	1	$\frac{\sqrt{3}\cos\theta\sin\theta\cos\phi}{r^3}$
	-1	$\frac{\sqrt{3}\cos\theta\sin\theta\sin\phi}{r^3}$
	2	$-\frac{\sqrt{3}}{2}\frac{\sin^2\theta+2\cos^2\theta\cos^2\phi-2\cos^2\phi}{r^3}$
	-2	$\frac{\sqrt{3}\sin^2\theta\sin\phi\cos\phi}{r^3}$

The unit field can then be calculated as the negative gradient of the unit potential

$$\begin{aligned}\mathbf{E}_l^{(m)} &= -\nabla\Phi_l^{(m)}(r,\theta,\phi) \\ &= -\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial r}\hat{\mathbf{r}} - \frac{1}{r}\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial\theta}\hat{\boldsymbol{\theta}} - \frac{1}{r\sin\theta}\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial\phi}\hat{\boldsymbol{\phi}}\,.\end{aligned} \tag{4.1.18}$$

The field for 9 multipole components through degree 2, which are listed in Table 4.3, lead to 36 scalar products that must be integrated via Eq. (4.1.14) to determine the descreening energy due to atom j.

Table 4.3. Unit vacuum fields.

l	m	$\mathbf{E}_l^{(m)}$
0	0	$\frac{1}{r^2}\hat{\mathbf{r}}$
1	0	$\frac{2\cos\theta}{r^3}\hat{\mathbf{r}}+\frac{\sin\theta}{r^3}\hat{\boldsymbol{\theta}}$
	1	$\frac{2\sin\theta\cos\phi}{r^3}\hat{\mathbf{r}}-\frac{\cos\theta\cos\phi}{r^3}\hat{\boldsymbol{\theta}}+\frac{\sin\phi}{r^3}\hat{\boldsymbol{\phi}}$
	-1	$\frac{2\sin\theta\sin\phi}{r^3}\hat{\mathbf{r}}-\frac{\cos\theta\sin\phi}{r^3}\hat{\boldsymbol{\theta}}-\frac{\cos\phi}{r^3}\hat{\boldsymbol{\phi}}$
2	0	$\frac{3}{2}\frac{\left(3\cos^2\theta-1\right)}{r^4}\hat{\mathbf{r}}+\frac{3\cos\theta\sin\theta}{r^4}\hat{\boldsymbol{\theta}}$
	1	$\frac{3\sqrt{3}\cos\theta\sin\theta\cos\phi}{r^4}\hat{\mathbf{r}}-\frac{\sqrt{3}\left(2\cos^2\theta-1\right)\cos\phi}{r^4}\hat{\boldsymbol{\theta}}+\frac{\sqrt{3}\cos\theta\sin\phi}{r^4}\hat{\boldsymbol{\phi}}$
	-1	$\frac{3\sqrt{3}\cos\theta\sin\theta\sin\phi}{r^4}\hat{\mathbf{r}}-\frac{\sqrt{3}\left(2\cos^2\theta-1\right)\sin\phi}{r^4}\hat{\boldsymbol{\theta}}-\frac{\sqrt{3}\cos\theta\cos\phi}{r^4}\hat{\boldsymbol{\phi}}$
	2	$\frac{3\sqrt{3}}{2}\frac{\sin^2\theta\cos2\phi}{r^4}\hat{\mathbf{r}}-\frac{\sqrt{3}\sin\theta\cos\theta\cos2\phi}{r^4}\hat{\boldsymbol{\theta}}+\frac{\sqrt{3}\sin\theta\sin2\phi}{r^4}\hat{\boldsymbol{\phi}}$
	-2	$\frac{3\sqrt{3}}{2}\frac{\sin^2\theta\sin2\phi}{r^4}\hat{\mathbf{r}}-\frac{\sqrt{3}\sin\theta\cos\theta\sin2\phi}{r^4}\hat{\boldsymbol{\theta}}-\frac{\sqrt{3}\sin\theta\cos2\phi}{r^4}\hat{\boldsymbol{\phi}}$

However, due to the symmetry of the integration domain only 14 scalar products lead to non-zero integrals, and these are listed in Table 4.4. The integration results are given in Table 4.5, showing 11 unique terms and 3 duplicates. Schaeffer et al. originally presented the same result for a monopole[31], and the higher order formulas are presented here for the first time. If the descreening angle ξ_{ij} is π as a result of atom j completely engulfing atom i, then the indefinite integrals simplify to those given in Table 4.6. This situation can occur for hydrogen atoms bonded to a heavy atom, for example, or in more artificial structures where one still wishes to have a continuous potential.

Table 4.4. Selected scalar products of unit magnitude vacuum spherical harmonic fields.

$(l,m)_1$	$(l,m)_2$	$\mathbf{E}_{l_1}^{(m_1)} \cdot \mathbf{E}_{l_2}^{(m_2)}$
(0, 0)	(0, 0)	$\frac{1}{r^4}$
	(1, 0)	$\frac{2\cos\theta}{r^5}$
	(2, 0)	$\frac{3}{2}\frac{3\cos^2\theta-1}{r^6}$
(1, 0)	(1, 0)	$\frac{3\cos^2\theta+1}{r^6}$
	(2, 0)	$\frac{6\cos^3\theta}{r^7}$
(1, 1)	(1, 1)	$\frac{\left(4\sin^2\theta+\cos^2\theta\right)\cos^2\phi+\sin^2\phi}{r^6}$
	(2, 1)	$\frac{\sqrt{3}\cos\theta\left(6\cos^2\phi\sin^2\theta-\cos^2\phi+2\cos^2\theta\cos^2\phi+\sin^2\phi\right)}{r^7}$
(1,-1)	(1,-1)	$\frac{\left(4\sin^2\theta+\cos^2\theta\right)\sin^2\phi+\cos^2\phi}{r^6}$
	(2,-1)	$\frac{\sqrt{3}\cos\theta\left(6\sin^2\phi\sin^2\theta-\sin^2\phi+2\sin^2\phi\cos^2\theta+\cos^2\phi\right)}{r^7}$
(2, 0)	(2, 0)	$\frac{1}{4}\frac{9+81\cos^4\theta+\left(36\sin^2\theta-54\right)\cos^2\theta}{r^8}$
(2, 1)	(2, 1)	$\frac{12\cos^2\phi\cos^4\theta+\left(\left(27\sin^2\theta-12\right)\cos^2\phi+3\sin^2\phi\right)\cos^2\theta+3\cos^2\phi}{r^8}$
(2,-1)	(2,-1)	$\frac{12\sin^2\phi\cos^4\theta+\left(\left(27\sin^2\theta-12\right)\sin^2\phi+3\cos^2\phi\right)\cos^2\theta+3\sin^2\phi}{r^8}$
(2, 2)	(2, 2)	$\frac{1}{4}\frac{\left(27\cos^2\theta+27+\left(12\sin^2\theta-54\right)\cos^2\theta\right)\cos^2 2\phi+12\sin^2 2\phi\sin^2\theta}{r^8}$
(2,-2)	(2,-2)	$\frac{3}{4}\frac{\sin^2\theta\left(9\sin^2 2\phi\sin^2\theta+4\sin^2 2\phi\cos^2\theta+4\cos^2 2\phi\right)}{r^8}$

Table 4.5. Indefinite integrals for the pairwise descreening of multipoles.

$(l,m)_1$	$(l,m)_2$	$D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)$
(0,0)	(0,0)	$-\left(2\ln\left(r\right)r^2+4r_{ij}r-r_{ij}^2+R_j^2\right)\Big/16\,r_{ij}r^2$
	(1,0)	$-\left(4r^4\ln\left(r\right)+4r_{ij}^2r^2+4r^2R_j^2-r_{ij}^4+2r_{ij}^2R_j^2-R_j^4\right)\Big/64r^4r_{ij}^2$
	(2,0)	$-\Big(12r^6\ln\left(r\right)+6r_{ij}^2r^4+18r^4R_j^2+3r^2r_{ij}^4$
		$+6r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6\Big)\Big/256r_{ij}^3r^6$
(1,0)	(1,0)	$-\Big(12r^6\ln\left(r\right)-42r_{ij}^2r^4+18r^4R_j^2+64r^3r_{ij}^3-21r^2r_{ij}^4$
		$+30r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6\Big)\Big/384r^6r_{ij}^3$
	(2,0)	$-\Big(24r^8\ln\left(r\right)-48r^6r_{ij}^2+48r^6R_j^2+60r_{ij}^4r^4$
		$+72r_{ij}^2r^4R_j^2-36r^4R_j^4-16r^2r_{ij}^6+48r^2r_{ij}^4R_j^2-48r^2r_{ij}^2R_j^4$
		$+16r^2R_j^6-3r_{ij}^8+12r_{ij}^6R_j^2-18r_{ij}^4R_j^4+12r_{ij}^2R_j^6-3R_j^8\Big)\Big/1024r^8r_{ij}^4$
(1,1) (1,-1)	(1,1) (1,-1)	$\Big(12r^6\ln\left(r\right)+102r_{ij}^2r^4+18r^4R_j^2-128r^3r_{ij}^3+51r^2r_{ij}^4$
		$-42r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6\Big)\Big/768r^6r_{ij}^3$
	(2,1) (2,-1)	$\sqrt{3}\Big(24r^8\ln\left(r\right)+96r^6r_{ij}^2+48r^6R_j^2-84r_{ij}^4r^4$
		$-72r_{ij}^2r^4R_j^2-36r^4R_j^4+32r^2r_{ij}^6-48r^2r_{ij}^4R_j^2$
		$+16r^2R_j^6-3r_{ij}^8+12r_{ij}^6R_j^2-18r_{ij}^4R_j^4+12r_{ij}^2R_j^6-3R_j^8\Big)\Big/3072r^8r_{ij}^4$
(2,0)	(2,0)	$-3\Big(120r^{10}\ln\left(r\right)-140r^8r_{ij}^2+300r^8R_j^2-540r^6r_{ij}^4+360r^6r_{ij}^2R_j^2$
		$-300r^6R_j^4+1024r^5r_{ij}^5-360r^4r_{ij}^6+600r^4r_{ij}^4R_j^2-440r^4r_{ij}^2R_j^4$
		$+200r^4R_j^6-35r^2r_{ij}^8+180r^2r_{ij}^6R_j^2-330r^2r_{ij}^4R_j^4+260r^2r_{ij}^2R_j^6-75r^2R_j^8$
		$-12r_{ij}^{10}+60r_{ij}^8R_j^2-120r_{ij}^6R_j^4+120r_{ij}^4R_j^6-60r_{ij}^2R_j^8+12R_j^{10}\Big)\Big/20480r^{10}r_{ij}^5$
(2,1) (2,-1)	(2,1) (2,-1)	$\Big(120r^{10}\ln\left(r\right)+180r^8r_{ij}^2+300r^8R_j^2+900r^6r_{ij}^4-120r^6r_{ij}^2R_j^2$
		$-300r^6R_j^4-1536r^5r_{ij}^5+600r^4r_{ij}^6-680r^4r_{ij}^4R_j^2-120r^4r_{ij}^2R_j^4$
		$+200r^4R_j^6+45r^2r_{ij}^8-60r^2r_{ij}^6R_j^2-90r^2r_{ij}^4R_j^4+180r^2r_{ij}^2R_j^6-75r^2R_j^8$
		$-12r_{ij}^{10}+60r_{ij}^8R_j^2-120r_{ij}^6R_j^4+120r_{ij}^4R_j^6-60r_{ij}^2R_j^8+12R_j^{10}\Big)\Big/10240r^{10}r_{ij}^5$

(2,2) (2,-2)	(2,2) (2,-1)	$-\left(120r^{10}\ln\left(r\right)+1140r^{8}r_{ij}^{2}+300r^{8}R_{j}^{2}-4380r^{6}r_{ij}^{4}-1560r^{6}r_{ij}^{2}R_{j}^{2}\right.$ $-300r^{6}R_{j}^{4}+6144r^{5}r_{ij}^{5}-2920r^{4}r_{ij}^{6}+1880r^{4}r_{ij}^{4}R_{j}^{2}+840r^{4}r_{ij}^{2}R_{j}^{4}$ $+200r^{4}R_{j}^{6}+285r^{2}r_{ij}^{8}-780r^{2}r_{ij}^{6}R_{j}^{2}+630r^{2}r_{ij}^{4}R_{j}^{4}-60r^{2}r_{ij}^{2}R_{j}^{6}-75r^{2}R_{j}^{8}$ $\left.-12r_{ij}^{10}+60r_{ij}^{8}R_{j}^{2}-120r_{ij}^{6}R_{j}^{4}+120r_{ij}^{4}R_{j}^{6}-60r_{ij}^{2}R_{j}^{8}+12R_{j}^{10}\right)\Big/40960r^{10}r_{ij}^{5}$

Table 4.6. Indefinite integrals for the pairwise descreening of multipoles when $\xi_{ij}=\pi$.

l_i	D_{l_i}
0	$-\frac{1}{2r}$
1	$-\frac{1}{3r^{3}}$
2	$-\frac{3}{10r^{5}}$

We note that after performing the integral no angular dependence remains. Therefore, although the derivation is based on spherical harmonics, our solution is equally useful for Cartesian tensors by using the conversion formulas in Table 4.1. We can now define the pairwise descreening integral for a permanent atomic multipole at site *i* being descreened by site *j* under the SFA as

$$I_{ij}\left(r_{ij},R_i,R_j\right)=\sum_{l_i=0}^{n}\frac{\left(2l_i+1\right)\varepsilon_h}{(l_i+1)\varepsilon_s+l_i\varepsilon_h}\sum_{m_i=-l_i}^{l_i}Q_{l_i}^{m_i}\sum_{l_j=0}^{n}\sum_{m_j=-l_j}^{l_j}Q_{l_j}^{m_j}D_{(l,m)_j,(l,m)_j}\left(r_{ij},R_i,R_j\right) \quad (4.1.19)$$

where $Q_{l_i}^{m_i}$ is the magnitude of a spherical harmonic of site *i*, $Q_{l_j}^{m_j}$ is the magnitude of a spherical harmonic of site *j* and $D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_i,R_j\right)$ is given by

$$D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_i,R_j\right)=\begin{cases}\underbrace{\begin{array}{ll}\delta_{(l_1,l_2)}\delta_{(m_1,m_2)}\,D_{l_i}\Big|_{r=R_i}^{r=R_j-r_{ij}} & \\ +D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=R_j-r_{ij}}^{r=r_{ij}+R_j} & R_j-r_{ij}>R_i\end{array}}_{\text{Case 1: Engulfment by the descreener.}} \\ \underbrace{D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=R_i}^{r=r_{ij}+R_j} \quad \begin{array}{l}R_j-r_{ij}<=R_i\\ r_{ij}<R_i+R_j\end{array}}_{\text{Case 2: Partial overlap.}} \\ \underbrace{D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=r_{ij}-R_j}^{r=r_{ij}+R_j} \quad r_{ij}>R_i+R_j}_{\text{Case 3: No overlap.}}\end{cases}. \tag{4.1.20}$$

Radial limits are given for three cases including engulfment by the descreener, partial overlap and no overlap. These limits are applied in conjunction with the indefinite integrals $D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)$ and D_{l_i} listed in Table 4.5 and Table 4.6, respectively. We note that the Kronecker delta functions δ specify that the engulfment integrals between orthogonal spherical harmonics vanish. In our implementation of Eq. (4.1.19), the magnitude of the spherical harmonics moments are found via conversion from AMOEBA traceless Cartesian multipoles.

4.1.2 The Reaction Potential Approximation

An alternative to the CFA for determining effective radii based on the analytic solution for the reaction potential of an off-center charge within a spherical dielectric cavity[27, 99] has been proposed by Grycuk.[96] We briefly outline this RPA method and its application to the self-energy of a permanent multipole.

The reaction potential at $\boldsymbol{r}$ due to an off-center charge at $\boldsymbol{r_0}$ inside a spherical dielectric cavity of permittivity ε_h surrounded by solvent with permittivity ε_s is given by

$$\Phi(\mathbf{r}) = \frac{q}{a\varepsilon_h}\sum_{l=0}^{\infty}\frac{(l+1)(\varepsilon_h-\varepsilon_s)}{(l+1)\varepsilon_s + l\varepsilon_h}\left(\frac{rr_0}{a^2}\right)^l P_l(\cos\theta) \tag{4.1.21}$$

where a is the cavity radius, q is the magnitude of the charge and P_l is the Legendre polynomial of degree l whose argument is the cosine of the angle θ between r and r_0.[27, 99] The self-energy of a charge based on Eq. (4.1.21) is

$$W(r) = \frac{1}{2}\frac{q^2}{a\varepsilon_h}\sum_{l=0}^{\infty}\frac{(l+1)(\varepsilon_h-\varepsilon_s)}{(l+1)\varepsilon_s + l\varepsilon_h}\left(\frac{d^2}{a^2}\right)^l P_l(1) \tag{4.1.22}$$

where d is used to specify the distance between the multipole site and the center of the sphere. For d=0, all asymmetric self-interactions vanish, for example the charge with a dipole component, but for off-center multipole sites these interactions are generally nonzero.[28] Noting that $P_l(1) = 1$ for all l, the summation in Eq. (4.1.22) can be reduced to a closed form if the factor (l+1) can be canceled by setting $l\varepsilon_h$ in the denominator to (l+1)ε_h or to 0, such that the self-energy is more positive or more negative than the true self-energy, respectively

$$
\begin{aligned}
W(d) &< \frac{1}{2}\frac{(\varepsilon_h-\varepsilon_s)}{\varepsilon_s\varepsilon_h+\varepsilon_h{}^2}\frac{q^2}{a}\sum_{l=0}^{\infty}\left(\frac{d^2}{a^2}\right)^l \\
&< \frac{1}{2}\frac{(\varepsilon_h-\varepsilon_s)}{\left(\varepsilon_s\varepsilon_h+\varepsilon_h{}^2\right)}q^2\frac{a}{\left(a^2-d^2\right)} \\
W(d) &> \frac{1}{2}\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{q^2}{a}\sum_{l=0}^{\infty}\left(\frac{d^2}{a^2}\right)^l \\
&> \frac{1}{2}\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q^2\frac{a}{\left(a^2-d^2\right)}
\end{aligned}
\quad . \tag{4.1.23}
$$

Both the upper and lower bound approach the true self-energy if $\varepsilon_s \gg \varepsilon_h$ allowing the simpler form to be used as an approximation

$$
w(d) = \frac{1}{2}\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q^2\frac{a}{\left(a^2-d^2\right)} \tag{4.1.24}
$$

As shown by Grycuk, it is possible to calculate the factor $a_r = a\big/\left(a^2-d^2\right)$, which is equivalent to the inverse of an effective radius, as

$$
a_r = \left(\frac{3}{4\pi}\int_{ex}\frac{1}{r'^6}dV\right)^{1/3} . \tag{4.1.25}
$$

This expression is motivated by the analytic solution for a spherical geometry

$$
\begin{aligned}
\int_{ex}\frac{1}{r'^6}dV &= 2\pi\int_a^{\infty}\int_0^{\pi}\frac{r^2\sin\theta}{\left(r^2+d^2-2dr\cos\theta\right)^3}d\theta dr \\
&= \frac{4\pi}{3}\frac{a^3}{\left(a^2-d^2\right)^3}
\end{aligned}
\tag{4.1.26}
$$

As d approaches zero, the multipole approaches the center of the dielectric sphere such that a_r simply equals the radius of the sphere a. In practice this integral is evaluated using

the pairwise descreening approach described in the previous section for the SFA and elsewhere.[32, 33] After determining effective radii, the self-energy for each permanent atomic multipole under the RPA is evaluated via Eq. (4.1.12).

4.1.3 Self-energy accuracy

We now demonstrate that for a series of proteins the RPA is superior to the SFA, which is consistent with findings for fixed partial charge models.[3] The perfect self-energy and perfect effective radii for all permanent atomic multipole sites for five proteins structures retrieved from the Protein Databank[87] (1CRN[88], 1ENH[89], 1FSV[90], 1PGB[91] and 1VII[92]) were determined using the PMPB model.[13] The grid size for all calculations was 257×257×257 using a grid spacing of 0.31 Å to give approximately 10 Å of continuum solvent between the low dielectric boundary and the grid boundary. The Bondi radii set (H 1.2, C 1.7, N 1.55, O 1.52, S 1.8) was used to define a step-function solute-solvent boundary with the solute dielectric set to unity and that of the solvent to 78.3.[100] Multiple Debye-Hückel boundary conditions were used to complete the definition of the Dirichlet problem. We also tried larger grids, up to 353×353×353, and therefore smaller grid spacing, which leads to the PMPB electrostatic solvation energy increasing by less than 2%. We opted for efficiency, since the important conclusion of this section, that the RPA is superior to the SFA, is not altered.

The SFA was fit using nonlinear optimization to determine one HCT scale factor per atomic number that minimized the RMS percent error in the permanent atomic multipole self-energies against numerical PMPB results for 3,032 data points. As

discussed previously, these HCT parameters scale down the radius of the descreening atom to prevent over counting due to atomic overlap. This leads o a mean unsigned relative difference (MURD) between the perfect self-energy for each multipole site and the SFA self-energy of 5.5%. However, using only a single scale factor (0.568), rather than one per atomic number, increased the MURD by just 0.4 to 5.9%.

Similarly, the RPA was fit using nonlinear optimization to determine a second set of scale factors to minimize the RMS percent difference between analytic effective radii and perfect effective radii. The achieved MURD in the effective radii was 1.1%. Alternatively, using a single scale factor (0.690) increased the MUPD by only 0.2% to 1.3%. Therefore, given the negligible improvements of using one HCT parameter per atomic number, we prefer implementations of the SFA and RPA that are each based on a single parameter.

The total analytic self-energy for each protein is compared to the total computed by summing the numerical permanent multipole self-energies as shown in Table 4.7. Fitting of a single HCT parameter for each method as described above eliminated the systematic error for both the SFA and RPA. However, the mean unsigned percent difference of the RPA (0.5) is smaller than that of the SFA (0.8). Considering that the RPA is more efficient and more accurate than the SFA, it is our preferred method to compute effective radii and permanent multipole self-energies.

Table 4.7. Shown is a comparison of the performance of the SFA and RPA in determining the perfect self-energy (kcal/mole) for a series of five folded proteins. Optimization of a single HCT scale factor for each method removes systematic error as shown by the mean signed percent differences. However, the mean RPA unsigned percent difference of 0.5 is smaller than that of the SFA.

	Self-Energy			Signed % Difference		Unsigned % Difference	
	PMPB	**SFA**	**RPA**	**SFA**	**RPA**	**SFA**	**RPA**
1CRN	-8141	-8191	-8196	-0.6	-0.7	0.6	0.7
1ENH	-11919	-11852	-11878	0.6	0.3	0.6	0.3
1FSV	-6254	-6341	-6287	-1.4	-0.5	1.4	0.5
1PGB	-11794	-11743	-11803	0.4	-0.1	0.4	0.1
1VII	-7206	-7132	-7133	1.0	1.0	1.0	1.0
Mean				0.0	0.0	0.8	0.5

4.2 Multipole Cross-Term Energy

There are two concepts needed to extend the GB cross-term to the interaction between two arbitrary multipole components. First, we describe the simplest possible definition for the reaction potential of any multipole component in the presence of a second multipole site, where an effective radius characterizes each site. Second, using this auxiliary definition of the reaction potential for each site, we formulate the cross-term energy in a consistent fashion. The electrostatic solvation free energy for the interaction between multipole components will be reproduced in the limiting cases of superimposition and wide separation.

4.2.1 Generalized Kirkwood Auxiliary Reaction Potential

The generalized Kirkwood auxiliary reaction potential is a building block for defining the interaction energy and its gradients for any pair of multipole components. It is motivated by noting that the only difference between the analytic solution for the

reaction potential inside and outside of a spherical solute with central multipole is exchange of the solute radius a in the former case with separation distance r_{ij} in the latter, where $\mathbf{r}_{ij} = \left(x_j - x_i, y_j - y_i, z_j - z_i\right)$.[86] For example, substitution for f in Eq. (4.2.1) below by a or r_{ij} gives the analytic formulas for the reaction potential inside and outside of the dielectric boundary, respectively.

Rather than using radial factors of $1/r_{ij}^{l+1}$ as was done earlier in defining the unit vacuum potential in terms of real spherical harmonics, the factor r_{ij}^{l}/f^{2l+1} is used to define the unit GK auxiliary reaction potential $A_l^{(m)}$ for a multipole component of degree l and order m,

$$A_l^{(m)}\left(\mathbf{r}_{ij}, a_i, a_j, \theta, \phi\right) = c_l \frac{r_{ij}^{l}}{f^{2l+1}} Y_l^{(m)}\left(\theta, \phi\right) \tag{4.2.1}$$

where f is the generalizing function defined in Eq. (2.3.6) and c_l is a function of the permittivity inside and outside the solute defined in Eq. (4.1.13). We note that for $r_{ij}^2 >> a_i a_j$, r_{ij}^{l}/f^{2l+1} approaches $1/r_{ij}^{l+1}$ to give the reaction potential in solvent. When $r_{ij} = 0$ and therefore $a_i = a_j = a$, then r_{ij}^{l}/f^{2l+1} simplifies to r_{ij}^{l}/a^{2l+1} to give the reaction potential at the center of the two concentric atoms. In this case the reaction potential is nonzero only for the monopole.

A definition in terms of Cartesian tensors is possible by first taking successive gradients of $1/r_{ij}$ and then substituting for factors of r_{ij} in the denominator with factors of f. For example, neglecting the ij subscript, the vacuum tensors are[78]

$$
\begin{aligned}
T &= \frac{1}{r} \\
T_\alpha &= \nabla_\alpha \frac{1}{r} = -\frac{r_\alpha}{r^3} \\
T_{\alpha\beta} &= \nabla_\alpha \nabla_\beta \frac{1}{r} = \frac{3 r_\alpha r_\beta}{r^5} - \frac{\delta_{\alpha\beta}}{r^3} \\
T_{\alpha\beta\gamma} &= \nabla_\alpha \nabla_\beta \nabla_\gamma \frac{1}{r} = -\frac{15 r_\alpha r_\beta r_\gamma}{r^7} + \frac{3\left(r_\alpha \delta_{\beta\gamma} + r_\beta \delta_{\alpha\gamma} + r_\gamma \delta_{\alpha\beta}\right)}{r^5} \\
T_{\alpha\beta\gamma\delta} &= \nabla_\alpha \nabla_\beta \nabla_\gamma \nabla_\delta \frac{1}{r} = \frac{105 r_\alpha r_\beta r_\gamma r_\delta}{r^9} \\
&\quad - \frac{15\left(r_\alpha r_\beta \delta_{\gamma\delta} + r_\alpha r_\gamma \delta_{\beta\delta} + r_\alpha r_\delta \delta_{\beta\gamma} + r_\beta r_\gamma \delta_{\alpha\delta} + r_\beta r_\delta \delta_{\alpha\gamma} + r_\gamma r_\delta \delta_{\alpha\beta}\right)}{r^7} \\
&\quad + \frac{3\left(\delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}\right)}{r^5}
\end{aligned}
\tag{4.2.2}
$$

where α, β, γ, and δ can take the values x, y, or z and the Kronecker delta function is unity if its subscripts are equal, but zero otherwise. Applying the substitution gives

$$
\begin{aligned}
A &= c_0 \frac{1}{f} \\
A_\alpha &= -c_1 \frac{r_\alpha}{f^3} \\
A_{\alpha\beta} &= c_2 \frac{3 r_\alpha r_\beta}{f^5} \\
A_{\alpha\beta\gamma} &= -c_3 \frac{15 r_\alpha r_\beta r_\gamma}{f^7} \\
A_{\alpha\beta\gamma\delta} &= c_4 \frac{105 r_\alpha r_\beta r_\gamma r_\delta}{f^9}
\end{aligned}
\tag{4.2.3}
$$

and represent the GK auxiliary reaction potential tensors. We have removed terms that require summing over a trace by requiring use of traceless multipoles. Unlike the vacuum case, the GK auxiliary reaction potential tensor of degree *l* is not simply a gradient of the degree *l*-1 tensor.

The total auxiliary reaction potential due to multipole *i*, up to quadrupole order, at site *j* is

$$\phi^{(i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) = q_i A - \mu_{i,\alpha} A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta} A_{\alpha\beta} \tag{4.2.4}$$

where the Einstein convention for repeated summation over Greek subscripts is implied. The total auxiliary potential due to multipole *j*, up to quadrupole order, at site *i* is given by

$$\phi^{(j)}\left(\mathbf{r}_{ji}, a_i, a_j\right) = q_j A - \mu_{j,\alpha} A_\alpha + \frac{1}{3}\Theta_{j,\alpha\beta} A_{\alpha\beta} \tag{4.2.5}$$

where $\boldsymbol{r}_{ji}$ is defined from site *j* to site *i*.

4.2.2 Generalized Kirkwood Cross-Term

Given the auxiliary reaction potentials, we define the auxiliary cross-term energy using Eq. (4.2.4) to be

$$U^{(i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) = \frac{1}{2}\left(q_j \phi^{(i)} + \mu_{j,\gamma}\nabla_\gamma \phi^{(i)} + \frac{1}{3}\Theta_{j,\gamma\delta}\nabla_\gamma \nabla_\delta \phi^{(i)}\right) \tag{4.2.6}$$

such that substituting for $\phi^{(i)}$ gives

$$\begin{aligned} U^{(i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) = \frac{1}{2}\Bigg[& q_j\left(q_i A - \mu_{i,\alpha} A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta} A_{\alpha\beta}\right) \\ & + \mu_{j,\gamma}\nabla_\gamma\left(q_i A - \mu_{i,\alpha} A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta} A_{\alpha\beta}\right) \\ & + \frac{1}{3}\Theta_{j,\gamma\delta}\nabla_\gamma\nabla_\delta\left(q_i A - \mu_{i,\alpha} A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta} A_{\alpha\beta}\right)\Bigg] \end{aligned} \tag{4.2.7}$$

while the auxiliary cross-term energy using Eq. (4.2.5) is

$$U^{(j)}\left(\mathbf{r}_{ji},a_i,a_j\right)=\frac{1}{2}\left(q_i\phi^{(j)}+\mu_{i,\gamma}\nabla_\gamma\phi^{(j)}+\frac{1}{3}\Theta_{i,\gamma\delta}\nabla_\gamma\nabla_\delta\phi^{(j)}\right) \tag{4.2.8}$$

such that substituting for $\phi^{(j)}$ gives

$$\begin{aligned}U^{(j)}\left(\mathbf{r}_{ji},a_i,a_j\right)=\frac{1}{2}\Bigg[&q_i\left(q_jA-\mu_{j,\alpha}A_\alpha+\frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right)\\&+\mu_{i,\gamma}\nabla_\gamma\left(q_jA-\mu_{j,\alpha}A_\alpha+\frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right)\\&+\frac{1}{3}\Theta_{i,\gamma\delta}\nabla_\gamma\nabla_\delta\left(q_jA-\mu_{j,\alpha}A_\alpha+\frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right)\Bigg]\end{aligned} \tag{4.2.9}$$

In the case of superimposition, either $U^{(i)}$ or $U^{(j)}$ exactly reproduces the correct self-energies. In the case of wide separation, both $\phi^{(i)}$ and $\phi^{(j)}$ neglect the bending of field lines near the spherical dielectric cavity surrounding site *j* and site *i*, respectively. The density of field lines in the case of wide separation is not an issue for a fixed partial charge interaction, although neglect of this effect introduces an error of less than 1% for dipole interactions in the case of a solute with unit permittivity in water.

Gradients of the auxiliary reaction potential can easily be obtained, although it is important to note that

$$\nabla_\alpha A\neq A_\alpha\,. \tag{4.2.10}$$

Namely, $\nabla_\alpha A$ includes a factor of $\left(1-e^{-r_{ij}^2/c_f a_i a_j}/c_f\right)$ relative to A_α such that equality is only achieved for r_{ij} equal to infinite. This subtle point implies, not surprisingly, the auxiliary reaction potential is too simple for intermediate r_{ij}. An important consequence is that $U^{(i)}\neq U^{(j)}$. A consistent model requires that the α-component of the potential

gradient at site *j* of a unit charge at site *i* should equal the potential at site *i* of the dipole's unit magnitude α-component at site *j*. This reciprocity condition is a well-known property of linear dielectric continuums.[85, 86] We note that in practice $\nabla_\alpha A \approx A_\alpha$ and therefore we simply take the average of the energies to obtain a consistent interaction model.

$$\Delta G_{ij} = \frac{1}{2}\left(U^{(i)} + U^{(j)}\right) \tag{4.2.11}$$

The qualitative behavior of the GK cross-term formulation for multipole permutations through quadrupole degree is seen in Figure 4.1. The system is composed of two spheres, each with a radius of 3.0 Å and unit permittivity, in a solvent with permittivity 78.3. The solute-solvent boundary is defined using a step-function boundary, although use of a smooth boundary does not qualitatively change the resulting plot. The total electrostatic solvation energy was evaluated using the PMPB and GK models. In the case of superimposition, the GK value is exact. When the two spheres are widely separated, GK asymptotes to the PMPB results for all permutations. For intermediate separations, the behavior is promising, but not exact.

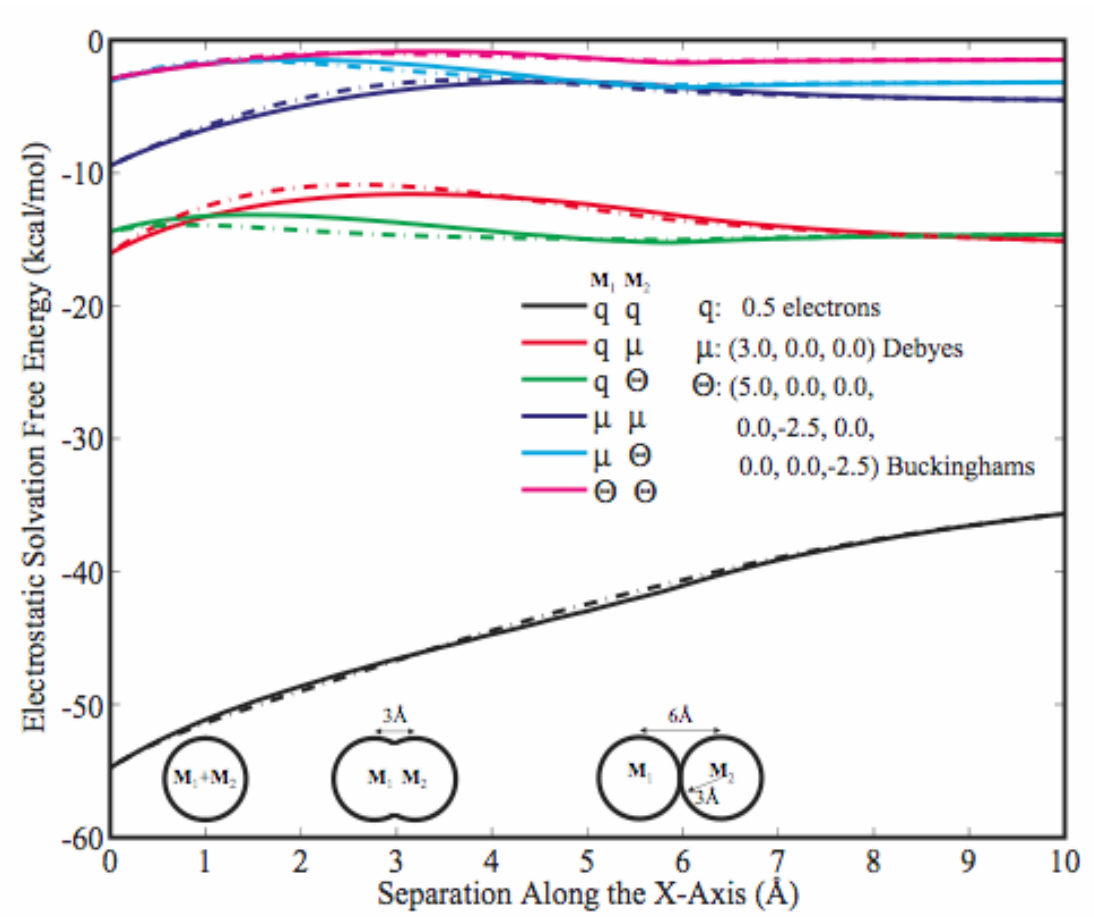

Figure 4.1. The solvation energy for a system composed two spheres, each with a radius of 3 Å and permittivity of 1, and a variety of multipole combinations are computed as a function of separation along the x-axis using numerical Poisson solutions (solid lines) and generalized Kirkwood (dashed lines). The solvent permittivity was 78.3. The limiting cases of wide separation and superimposition are exact in all cases, while intermediate separations are seen to be a reasonable approximation.

4.3 Factoring of Generalized Kirkwood Tensors

Cartesian multipole interaction tensors can be computed via recurrence relationships, which can greatly improve the efficiency of their use for high degree expansions.[101, 102] Unfortunately, we have not found an analogous approach for Generalized Kirkwood tensors. In this section we present a practical factoring of the associated algebra that mirrors our implementation of GK, but it is conceivable superior alternatives exist.

The GK auxiliary potential tensor $\mathbf{A}^{(n)}$ of rank n has 3^n elements, but because it is totally symmetric only $(n+1)(n+2)/2$ elements are distinct. For example, the dipole auxiliary potential tensor $\mathbf{A}^{(1)}$ has 3 elements $\left(-c_1\frac{x}{f^3}, -c_1\frac{y}{f^3}, -c_1\frac{z}{f^3}\right)$, where the generalizing function f was defined in Eq. (2.3.6) and c_l was defined in Eq. (4.1.13). In compressed tensor notation, elements are denoted as $\mathrm{A}^{(n)}_{\{n_1,n_2,n_3\}}$, where n_1, n_2, and n_3 are called degree indices that satisfy the constraint $n_1+n_2+n_3=n$.[82, 83] All components of the GK auxiliary potential tensor of any order can be decomposed as $\mathrm{A}^{(n)}_{\{n_1,n_2,n_3\}} = x^{n_1}y^{n_2}z^{n_3}t_{(n,0)}$ where $t_{(n,0)}$ is an entry in the first column of a matrix $\boldsymbol{t}$ that we construct purely for convenience, and detail below. Using this notation, one element of the auxiliary dipole potential is $\mathrm{A}^{(1)}_{\{1,0,0\}} = xt_{(n,0)}$, where $t_{(1,0)}$ is $-c_1\frac{1}{f^3}$.

Generation of Cartesian derivatives for any element $\mathrm{A}^{(n)}_{\{n_1,n_2,n_3\}}$ will now be described. Unlike tensors built from derivatives of $1/r$, GK auxiliary potential tensors do not, in general, obey the relationship

$$\frac{\partial \mathrm{A}^{(n)}_{\{n_1,n_2,n_3\}}}{\partial x} \neq \mathrm{A}^{(n+1)}_{\{n_1+1,n_2,n_3\}} . \tag{4.3.1}$$

However, we present a factoring scheme that facilitates generation of the mth order auxiliary potential gradient for any GK auxiliary potential tensor, which we denote as

$$\mathrm{A}^{(n)}_{\{n_1,n_2,n_3\},\{m_1,m_2,m_3\}} = \frac{\partial^m}{\partial x^{m_1}\partial y^{m_2}\partial z^{m_3}} \mathrm{A}^{(n)}_{\{n_1,n_2,n_3\}} \tag{4.3.2}$$

where $m_1 + m_2 + m_3 = m$. All potential gradients are composed of sums of terms that have the form $p_1 x^{p_2} y^{p_3} z^{p_4} t_{(i,j)}$, where p_1, p_2, p_3, and p_4 are constants. These are enumerated in Table 7.1 through Table 7.10 for moments through quadrupole degree. For example Table 7.2, contains the auxiliary reaction potential $\mathrm{A}^{(1)}_{\{1,0,0\}}$ and its gradients for the x-component of a dipole.

The effective radii chain rule terms are denoted $\frac{\partial}{\partial a_1}\mathrm{A}^{(n)}_{\{n_1,n_2,n_3\},\{m_1,m_2,m_3\}}$, where a_1 denotes the derivative is with respect to the effective radii of site 1. They are composed of sums of terms that have the form $a_2 p_1 x^{p_2} y^{p_3} z^{p_4} b_{(i,j)}$ where $b_{(i,j)} = \frac{\partial t_{(i,j)}}{\partial a_1}\frac{1}{a_2} = \frac{\partial t_{(i,j)}}{\partial a_2}\frac{1}{a_1}$ is

an element from a second matrix **b**, again defined for convenience. This matrix contains the derivatives of the **t** matrix elements with respect to an effective radius, normalized by the effective radius of the other site in the pairwise interaction. The chain rule term with respect to an effective radius for any Table 7.1 through Table 7.10 is given by substituting the **t** matrix elements with corresponding elements from the **b** matrix and multiplying by the opposite effective radius. For example, the chain rule term for the first entry in Table 7.2 with respect to effective radius 1 is $\frac{\partial}{\partial a_1}\mathrm{A}^{(1)}_{\{1,0,0\},\{0,0,0\}} = a_2 x b_{(1,0)}$ and that for the 2nd entry is $\frac{\partial}{\partial a_1}\mathrm{A}^{(1)}_{\{1,0,0\},\{1,0,0\}} = a_2\left(b_{(1,0)} + x^2 b_{(1,1)}\right)$.

All that remains is to describe an efficient mechanism to generate all elements of the **t** and **b** matrices. The matrix **t** is of size $n \times m$, whose rows and columns are indexed from $0..n-1$ and $0..m-1$, respectively. The first column contains the GK auxiliary reaction potential tensors given in Eq. (4.2.3) without factors of x, y or z in the numerator. All other columns contain derivatives with respect to r_α of the previous column, where α represents x, y or z. The results are normalized by r_α such that the terms are independent of which derivative was taken.

$$\mathbf{t} = \begin{bmatrix} c_0 \dfrac{1}{f} & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(0,0)} & \cdots & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(0,m-2)} \\ -c_1 \dfrac{1}{f^3} & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(1,0)} & \cdots & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(1,m-2)} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{(n-1)} c_{n-1} \dfrac{\left(2(n-1)-1\right)!!}{f^{2(n-1)+1}} & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(n-1,0)} & \cdots & \dfrac{1}{r_\alpha}\dfrac{\partial}{\partial r_\alpha} t_{(n-1,m-2)} \end{bmatrix} \tag{4.3.3}$$

We note that all the elements of the 2nd column $t_{(i,1)}$ are related to elements of the first column by a constant factor f_1,

$$f_1 = \left(\frac{\partial t_{(i,0)}}{\partial r_\alpha}\right)\frac{1}{r_\alpha t_{(i+1,0)}} = 1 - \frac{e^{-r_{ij}^2/c_f a_1 a_2}}{c_f} \tag{4.3.4}$$

such that all $t_{(i,1)}$ can be found as

$$t_{(i,1)} = f_1 t_{(i+1,0)} \tag{4.3.5}$$

By the chain rule, all components in the 3rd column of **t** are related to those in the first two columns as

$$t_{(i,2)} = f_1 t_{(i+1,1)} + f_2 t_{(i+1,0)} \tag{4.3.6}$$

where f_2 is the derivative of f_1 normalized by r_α

$$f_2 = \left(\frac{\partial f_1}{\partial r_\alpha}\right)\frac{1}{r_\alpha} = \frac{2}{c_f^2 a_1 a_2} e^{-r_{ij}^2/c_f a_1 a_2} \tag{4.3.7}$$

All higher order (normalized) derivatives f_i can be determined as

$$f_i = f_r^{i-2} f_2, \quad i \geq 2 \tag{4.3.8}$$

where,

$$f_r = -\frac{2}{c_f a_1 a_2}. \tag{4.3.9}$$

For example, f_3 is the last such term needed for GK quadrupole-quadrupole energy gradient

$$f_3 = \left(\frac{\partial f_2}{\partial r_\alpha}\right)\frac{1}{r_\alpha} = f_r f_2 \tag{4.3.10}$$

and the final column needed for **t** is

$$t_{(i,3)} = f_1 t_{(i+1,2)} + 2 f_2 t_{(i+1,1)} + f_3 t_{(i+1,0)}. \tag{4.3.11}$$

However, for arbitrary order tensors, any entry can be determined from entries in previous columns as a sum

$$t_{(i,j)} = \left(\frac{\partial t_{(i,j-1)}}{\partial r_\alpha}\right)\frac{1}{r_\alpha} = \sum_{k=1}^{j} \omega_{j,k} f_k t_{(i+1,j-k)} \tag{4.3.12}$$

where each row of the coefficient matrix $\boldsymbol{\omega}$ can be determined from the previous row

$$\omega = \begin{bmatrix} 0 & & & & & & \\ 1 & 0 & & & & & \\ 1 & 1 & 0 & & & & \\ 1 & 2 & 1 & 0 & & & \\ 1 & 3 & 3 & 1 & 0 & & \\ 1 & 4 & 6 & 4 & 1 & 0 & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{4.3.13}$$

Note that row j of the coefficient matrix is used for all elements in column j of the matrix **t**.

The matrix of effective radius chain rule terms **b** is of size $n \times m - 1$. It has one fewer columns than **t** because the last column in **t** is itself only needed for energy gradients and not the energy. Therefore effective radius chain rule terms are not needed for this column. We note that any term in the first column $b_{(i,0)}$ is related to an element in the first column of **t** by a factor we label g

$$g = \left(\frac{\partial t_{(i,0)}}{\partial a_1} \right) \frac{1}{a_2 t_{(i+1,0)}} = \frac{1}{2} e^{-r_{ij}^2 / c_f a_1 a_2} \left(1 + \frac{r_{ij}^2}{c_f a_1 a_2} \right) \tag{4.3.14}$$

such that

$$b_{(i,0)} = g t_{(i+1,0)} . \tag{4.3.15}$$

Elements in the 2nd column of **b** can be found from elements in the first columns of **t** and **b** via the chain rule as

$$b_{(i,1)} = \left(\frac{\partial t_{(i,1)}}{\partial a_1} \right) \frac{1}{a_2} = g_1 t_{(i+1,0)} + f_1 b_{(i+1,0)} \tag{4.3.16}$$

where g_1 is defined as

$$g_1 = \left(\frac{\partial f_1}{\partial a_1} \right) \frac{1}{a_2} = -\frac{r_{ij}^2 e^{-r_{ij}^2 / c_f a_1 a_2}}{c_f^2 a_1^2 a_2^2} \,. \tag{4.3.17}$$

Similarly, the 3rd column is

$$b_{(i,2)} = \left(\frac{\partial t_{(i,2)}}{\partial a_1} \right) \frac{1}{a_2} = g_1 t_{(i+1,1)} + g_2 t_{(i+1,0)} + f_1 b_{(i+1,1)} + f_2 b_{(i+1,0)} \tag{4.3.18}$$

where

$$g_2 = \left(\frac{\partial f_2}{\partial a_1} \right) \frac{1}{a_2} \quad = \frac{2 e^{-r_{ij}^2 / c_f a_1 a_2}}{\left(c_f a_1 a_2 \right)^2} \left(\frac{r_{ij}^2}{c a_1 a_2} - 1 \right). \tag{4.3.19}$$

All further terms g_i are determined from f_i as

$$g_i = \left(\frac{\partial f_i}{\partial a_1} \right) \frac{1}{a_2} = \left(i - 2 \right) f_r^{i-3} g_r f_2 + f_r^{i-2} g_2, \quad i \geq 2 \tag{4.3.20}$$

where

$$g_r = \frac{2}{c_f \left(a_1 a_2 \right)^2} \,. \tag{4.3.21}$$

For example,

$$g_3 = g_r f_2 + f_r g_2 \,. \tag{4.3.22}$$

It is now possible to define all elements of the matrix **b**,

$$b_{(i,j)} = \left(\frac{\partial t_{(i,j)}}{\partial a_1} \right) \frac{1}{a_2} = \sum_{k=1}^{j} \omega_{j,k} \left(g_k t_{(i+1,j-k)} + f_k b_{(i+1,j-k)} \right) \tag{4.3.23}$$

to facilitate determination of energy gradients for any order multipole interaction.

4.4 AMOEBA Solutes in a Generalized Kirkwood Continuum

4.4.1 Electrostatic Solvation Free Energy

Derivation of the electrostatic solvation free energy for an AMOEBA solute within the GK continuum resembles the derivation of the PMPB electrostatic solvation free energy.[13] Each permanent atomic multipole site can be considered as a vector of coefficients including charge, dipole and quadrupole components

$$\mathbf{M}_i = \left[q_i, d_{i,x}, d_{i,y}, d_{i,z}, \Theta_{i,xx}, \Theta_{i,xy}, \Theta_{i,xz}, ..., \Theta_{i,zz} \right]^t \tag{4.4.1}$$

where the superscript t denotes the transpose. The interaction potential energy between two sites i and j separated by the distance r_{ij} in a homogeneous permittivity ε_h can then be represented in tensor notation as

$$
\begin{aligned}
\mathrm{U}\left(\mathbf{r}_{ij}\right) &= \mathbf{M}_i^t \mathbf{T}_{ij} \mathbf{M}_j \\
&= \begin{bmatrix} q_i \\ d_{i,x} \\ d_{i,y} \\ d_{i,z} \\ \Theta_{i,xx} \\ \vdots \end{bmatrix}^t
\begin{bmatrix}
1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\
\frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\
\frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\
\frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\frac{1}{\varepsilon_h r_{ij}}
\begin{bmatrix} q_j \\ d_{j,x} \\ d_{j,y} \\ d_{j,z} \\ \Theta_{j,xx} \\ \vdots \end{bmatrix}
\end{aligned}
\tag{4.4.2}
$$

Similarly, the GK energy for two multipoles (self or cross-term) is given by

$$
\Delta G_{ij}\left(\mathbf{r}_{ij}, a_i, a_j\right) = \frac{1}{2} \mathbf{M}_i^t \mathbf{K}_{ij} \mathbf{M}_j \tag{4.4.3}
$$

where the factor of ½ accounts for the cost of charging the continuum and the GK interaction matrix $\mathbf{K}_{ij}$ depends on the coordinates of all atoms via the effective radii a_i and a_j. As introduced above, GK requires averaging of the auxiliary reaction potentials and their respective gradients to obtain a consistent interaction matrix

$$\mathbf{K}_{ij} = \frac{1}{2}\left[\mathbf{K}^{(i)} + \mathbf{K}^{(j)}\right]$$

$$\mathbf{K}^{(i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) = \begin{bmatrix} A & \frac{\partial A}{\partial x} & \frac{\partial A}{\partial y} & \frac{\partial A}{\partial z} & \cdots \\ A_x & \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} & \cdots \\ A_y & \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} & \cdots \\ A_z & \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4.4.4}$$

$$\mathbf{K}^{(j)}\left(\mathbf{r}_{ji}, a_i, a_j\right) = \left(\mathbf{K}^{(i)}\left(\mathbf{r}_{ji}, a_i, a_j\right)\right)^t$$

Each site may also be polarizable, such that an induced dipole is formed in vacuum $\boldsymbol{\mu}_i^v$ proportional to the strength of the local field

$$\begin{aligned} \boldsymbol{\mu}_i^v &= \alpha_i \mathbf{E}_i^v \\ &= \alpha_i \left(\sum_{j \neq i} \mathbf{T}_{d,ij}^{(1)} \mathbf{M}_j + \sum_{k \neq i} \mathbf{T}_{ik}^{(11)} \boldsymbol{\mu}_k \right) \end{aligned} \tag{4.4.5}$$

Here α_i is an isotropic atomic polarizability and $\mathbf{E}_i^v$ is the total vacuum field, which can be decomposed into contributions from permanent multipole sites and induced dipoles, and the summations run over all multipole sites. The interaction tensors $\mathbf{T}_{d,ij}^{(1)}$ and $\mathbf{T}_{ik}^{(11)}$ are, respectively,

$$\mathbf{T}_{d,ij}^{(1)} = \begin{bmatrix} \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \end{bmatrix} \frac{1}{\varepsilon_h r_{ij}} \tag{4.4.6}$$

and

$$\mathbf{T}_{ik}^{(11)} = \begin{bmatrix} \frac{\partial^2}{\partial x_i \partial x_k} & \frac{\partial^2}{\partial x_i \partial y_k} & \frac{\partial^2}{\partial x_i \partial z_k} \\ \frac{\partial^2}{\partial y_i \partial x_k} & \frac{\partial^2}{\partial y_i \partial y_k} & \frac{\partial^2}{\partial y_i \partial z_k} \\ \frac{\partial^2}{\partial z_i \partial x_k} & \frac{\partial^2}{\partial z_i \partial y_k} & \frac{\partial^2}{\partial z_i \partial z_k} \end{bmatrix} \frac{1}{\varepsilon_h r_{ik}} \tag{4.4.7}$$

where the d in $\mathbf{T}_{d,ij}^{(1)}$ denotes that masking rules for the AMOEBA group-based polarization model are applied. Upon adding the GK reaction field due to the permanent multipoles and induced dipoles, the self-consistent induced dipoles are proportional to the self-consistent reaction field

$$\begin{aligned} \boldsymbol{\mu}_i &= \alpha_i \mathbf{E}_i \\ &= \alpha_i \left[\sum_j \left[\left(1-\delta_{ij}\right) \mathbf{T}_{d,ij}^{(1)} + \mathbf{K}_{ij}^{(1)} \right] \mathbf{M}_j + \sum_k \left[\left(1-\delta_{ik}\right) \mathbf{T}_{ik}^{(11)} + \mathbf{K}_{ik}^{(11)} \right] \boldsymbol{\mu}_k \right] \end{aligned} \tag{4.4.8}$$

where the sums now include self-contributions to the reaction field, but exclude Coulomb self-interactions via Kronecker delta functions. The GK interaction matrices $\mathbf{K}_{ij}^{(1)}$ and $\mathbf{K}_{ik}^{(11)}$ are, respectively,

$$\mathbf{K}_{ij}^{(1)} = \frac{1}{2}\left(\mathbf{K}_{ij}^{(1,i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) + \mathbf{K}_{ij}^{(1,j)}\left(\mathbf{r}_{ji}, a_i, a_j\right) \right) \tag{4.4.9}$$

where

$$\mathbf{K}_{ij}^{(1,i)}\left(\mathbf{r}_{ij},a_i,a_j\right)=\begin{bmatrix} A_x & \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} & \cdots \\ A_y & \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} & \cdots \\ A_z & \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} & \cdots \end{bmatrix} \tag{4.4.10}$$

$$\mathbf{K}_{ij}^{(1,j)}\left(\mathbf{r}_{ji},a_i,a_j\right)=\begin{bmatrix} \frac{\partial A}{\partial x} & \frac{\partial A_x}{\partial x} & \frac{\partial A_y}{\partial x} & \frac{\partial A_z}{\partial x} & \cdots \\ \frac{\partial A}{\partial y} & \frac{\partial A_x}{\partial y} & \frac{\partial A_y}{\partial y} & \frac{\partial A_z}{\partial y} & \cdots \\ \frac{\partial A}{\partial z} & \frac{\partial A_x}{\partial z} & \frac{\partial A_y}{\partial z} & \frac{\partial A_z}{\partial z} & \cdots \end{bmatrix} \tag{4.4.11}$$

and

$$\mathbf{K}_{ik}^{(11)}=\begin{bmatrix} \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} \\ \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} \\ \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} \end{bmatrix} \tag{4.4.12}$$

where averaging cancels for the matrix $\mathbf{K}_{ik}^{(11)}$ that produces the field at site *i* due to the induced dipole at site *k* as a result of symmetry.

The linear system of equations, both for the vacuum and solvated systems, can be solved via a number of approaches, including direct matrix inversion or iterative schemes such as successive over-relaxation (SOR). The total vacuum electrostatic energy U_{elec}^{v} includes pairwise permanent multipole interactions and many-body polarization

$$\mathrm{U}^{\mathrm{v}}_{\mathrm{elec}} = \frac{1}{2}\left[\mathbf{M}^t\mathbf{T} - \left(\boldsymbol{\mu}^{\mathrm{v}}\right)^t \mathbf{T}^{(1)}_{\mathrm{p}}\right]\mathbf{M} \tag{4.4.13}$$

where the factor of ½ avoids double-counting of permanent multipole interactions in the first term and accounts for the cost of polarizing the system in the second term. Furthermore, $\mathbf{M}$ is a column vector of 13N multipole components

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_{\mathrm{N}} \end{bmatrix} \tag{4.4.14}$$

$\mathbf{T}$ is a N x N supermatrix with $\mathbf{T}_{ij}$ off-diagonal elements

$$\mathbf{T} = \begin{bmatrix} 0 & \mathbf{T}_{12} & \mathbf{T}_{13} & \cdots \\ \mathbf{T}_{21} & 0 & \mathbf{T}_{23} & \cdots \\ \mathbf{T}_{31} & \mathbf{T}_{32} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4.4.15}$$

$\boldsymbol{\mu}^{\mathrm{v}}$ is a 3N column vector of converged induced dipole components in vacuum

$$\boldsymbol{\mu}^{\mathrm{v}} = \begin{bmatrix} \mu^v_{1,x} \\ \mu^v_{1,y} \\ \mu^v_{1,z} \\ \vdots \\ \mu^v_{N,z} \end{bmatrix} \tag{4.4.16}$$

and $\mathbf{T}^{(1)}_{\mathrm{p}}$ is a 3N x 13N supermatrix with $\mathbf{T}^{(1)}_{\mathrm{p},ij}$ as off-diagonal elements

$$\mathbf{T}^{(1)}_{\mathrm{p}} = \begin{bmatrix} 0 & \mathbf{T}^{(1)}_{\mathrm{p},12} & \mathbf{T}^{(1)}_{\mathrm{p},13} & \cdots \\ \mathbf{T}^{(1)}_{\mathrm{p},21} & 0 & \mathbf{T}^{(1)}_{\mathrm{p},23} & \cdots \\ \mathbf{T}^{(1)}_{\mathrm{p},31} & \mathbf{T}^{(1)}_{\mathrm{p},32} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.4.17}$$

The subscript p denotes a tensor matrix that operates on the permanent multipoles to produce the electric field in which the polarization energy is evaluated, while the subscript d is used to specify an analogous tensor matrix that produces the field that induces dipoles. The differences between the two are masking rules that leave out 1-2, 1-3, and 1-4 interactions in the former case and use the AMOEBA group based polarization scheme for the later.[3, 7, 9, 10]

For the solvated system, the total electrostatic energy is similar to the vacuum case

$$\mathrm{U}_{\mathrm{elec}} = \frac{1}{2}\left[\mathbf{M}^{t}\left(\mathbf{T}+\mathbf{K}\right) - \boldsymbol{\mu}^{t}\left(\mathbf{T}_{\mathrm{p}}^{(1)} + \mathbf{K}^{(1)}\right)\right]\mathbf{M} \tag{4.4.18}$$

where the GK matrices are

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} & \cdots \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \cdots \\ \mathbf{K}_{31} & \mathbf{K}_{32} & \mathbf{K}_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4.4.19}$$

and

$$\mathbf{K}^{(1)} = \begin{bmatrix} \mathbf{K}_{11}^{(1)} & \mathbf{K}_{12}^{(1)} & \mathbf{K}_{13}^{(1)} & \cdots \\ \mathbf{K}_{21}^{(1)} & \mathbf{K}_{22}^{(1)} & \mathbf{K}_{23}^{(1)} & \cdots \\ \mathbf{K}_{31}^{(1)} & \mathbf{K}_{32}^{(1)} & \mathbf{K}_{33}^{(1)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.4.20}$$

The total electrostatic solvation free energy is determined as the difference between the vacuum electrostatic energy and total electrostatic energy in solvent as

$$\mathrm{U}_{\mathrm{solv}} = \frac{1}{2}\left(\mathbf{M}^{t}\mathbf{K} - \boldsymbol{\mu}^{\Delta}\mathbf{T}_{\mathrm{p}}^{(1)} - \boldsymbol{\mu}\mathbf{K}^{(1)}\right)\mathbf{M} \tag{4.4.21}$$

where $\boldsymbol{\mu}^{\Delta}$ represents the change in the induced dipoles upon solvation

$$\boldsymbol{\mu}^{\Delta} = \boldsymbol{\mu} - \boldsymbol{\mu}^{\mathrm{v}} . \tag{4.4.22}$$

4.4.2 Permanent Multipole Energy Gradient

The permanent multipole electrostatic solvation energy gradient between sites *i* and *j* only depends on the gradient of the GK interaction tensor

$$\frac{\partial \mathbf{K}_{ij}}{\partial r_{i,\sigma}} = \left(\frac{\partial \mathbf{K}_{ij}}{\partial r_{i,\sigma}}\right)_{a_i,a_j} + \left(\frac{\partial \mathbf{K}_{ij}}{\partial a_i}\right)\frac{\partial a_i}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{K}_{ij}}{\partial a_j}\right)\frac{\partial a_j}{\partial r_{i,\sigma}} \tag{4.4.23}$$

where

$$\begin{aligned}
\left(\frac{\partial \mathbf{K}_{ij}}{\partial r_{i,\sigma}}\right)_{a_i,a_j} &= \frac{1}{2}\left(\frac{\partial \mathbf{K}^{(i)}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}^{(j)}}{\partial r_{i,\sigma}}\right)_{a_i,a_j} \\
\left(\frac{\partial \mathbf{K}_{ij}}{\partial a_i}\right)\frac{\partial a_i}{\partial \mathbf{r}_i} &= \frac{1}{2}\left(\frac{\partial \mathbf{K}^{(i)}}{\partial a_i} + \frac{\partial \mathbf{K}^{(j)}}{\partial a_i}\right)\frac{\partial a_i}{\partial r_{i,\sigma}} \\
\left(\frac{\partial \mathbf{K}_{ij}}{\partial a_j}\right)\frac{\partial a_i}{\partial \mathbf{r}_i} &= \frac{1}{2}\left(\frac{\partial \mathbf{K}^{(i)}}{\partial a_j} + \frac{\partial \mathbf{K}^{(j)}}{\partial a_j}\right)\frac{\partial a_j}{\partial r_{i,\sigma}}
\end{aligned} \tag{4.4.24}$$

and subscript a_i and a_j denote keeping the effective radii fixed in this case. Generation of the GK interaction tensors that make up $\frac{\partial \mathbf{K}^{(i)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}^{(i)}}{\partial a_i}$, $\frac{\partial \mathbf{K}^{(i)}}{\partial a_j}$, $\frac{\partial \mathbf{K}^{(j)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}^{(j)}}{\partial a_i}$ and $\frac{\partial \mathbf{K}^{(j)}}{\partial a_j}$ were described in the previous section. The derivatives of the effective radii with respect to an atomic displacement follow from the pairwise descreening implementation of the RPA and will not be discussed here.[32, 33] We also point out that there is a torque on the permanent dipoles due to the permanent reaction field and also on the permanent

quadrupoles due to the permanent reaction field gradient. All torques, including contributions from the polarization energy gradient discussed below, are converted to forces on adjacent atoms that define the local coordinate frame of the multipole.

4.4.3 Polarization Energy Gradient

The polarization energy gradient when using either the "direct" or "mutual" polarization models within the GK continuum will now be derived. The definition of the starting point for the iterative convergence of the self-consistent reaction field (SCRF) is the total "direct" field $\mathbf{E}_{\text{direct}}$ at each polarizable site. This field is the sum of the permanent atomic multipoles (PAM) intramolecular field

$$\mathbf{E}_{\text{d}} = \mathbf{T}_{\text{d}}^{(1)}\mathbf{M} \tag{4.4.25}$$

where $\mathbf{T}_{\text{d}}^{(1)}$ is analogous to the tensor matrix defined in deriving the AMOEBA vacuum energy in Eq. (2.1.6), and the PAM GK reaction field

$$\mathbf{E}_{\text{RF}} = \mathbf{K}^{(1)}\mathbf{M} \tag{4.4.26}$$

The product of the direct field $\mathbf{E}_{\text{direct}}$ with a vector of atomic polarizabilities determines the initial induced dipoles $\boldsymbol{\mu}_{\text{direct}}$

$$\begin{aligned} \boldsymbol{\mu}_{\text{direct}} &= \boldsymbol{\alpha}\mathbf{E}_{\text{direct}} \\ &= \boldsymbol{\alpha}\left(\mathbf{T}_{\text{d}}^{(1)} + \mathbf{K}^{(1)}\right)\mathbf{M} \end{aligned} \tag{4.4.27}$$

At this point the induced dipoles do not act upon each other nor do they elicit a reaction field. This is defined as the direct model of polarization.

In contrast to the direct polarization model, the total SCRF $\mathbf{E}$ has two additional contributions due to the induced dipoles and their reaction field,

$$\mathbf{E} = \left(\mathbf{T}_{\mathrm{d}}^{(1)} + \mathbf{K}^{(1)}\right)\mathbf{M} + \left(\mathbf{T}^{(11)} + \mathbf{K}^{(11)}\right)\boldsymbol{\mu} \tag{4.4.28}$$

for a sum of 4 contributions. The induced dipoles

$$\boldsymbol{\mu} = \boldsymbol{\alpha}\left[\left(\mathbf{T}_{\mathrm{d}}^{(1)} + \mathbf{K}^{(1)}\right)\mathbf{M} + \left(\mathbf{T}^{(11)} + \mathbf{K}^{(11)}\right)\boldsymbol{\mu}\right] \tag{4.4.29}$$

can be solved for in an iterative fashion using successive over-relaxation (SOR) to accelerate convergence.[84] Alternatively, the induced dipoles can be solved for directly as a mechanism for deriving the polarization energy gradient with respect to an atomic displacement. Moving all terms containing the induced dipoles to the LHS allows their isolation

$$\left(\boldsymbol{\alpha}^{-1} - \mathbf{T}^{(11)} - \mathbf{K}^{(11)}\right)\boldsymbol{\mu} = \left(\mathbf{T}_{\mathrm{d}}^{(1)} + \mathbf{K}^{(1)}\right)\mathbf{M} \tag{4.4.30}$$

For convenience, a matrix C is defined as

$$\mathbf{C} = \boldsymbol{\alpha}^{-1} - \mathbf{T}^{(11)} - \mathbf{K}^{(11)} \tag{4.4.31}$$

which is substituted into Eq. (4.4.30) above to show the induced dipoles are a linear function of the PAM $\mathbf{M}$, directly via the intramolecular interaction tensor $\mathbf{T}_{\mathrm{d}}^{(1)}$ that implicitly contains the AMOEBA group based polarization scheme, and also through their reaction field

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{C}^{-1}\left(\mathbf{T}_{\mathrm{d}}^{(1)} + \mathbf{K}^{(1)}\right)\mathbf{M} \\ &= \mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}\right)\end{aligned} \tag{4.4.32}$$

The polarization energy can now be described in terms the permanent reaction field and solute field $\mathbf{E}_{\mathrm{p}}$

$$\mathrm{U}_{\mu} = -\frac{1}{2}\left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}\right)^{t}\boldsymbol{\mu} \tag{4.4.33}$$

To find the polarization energy gradient, we wish to avoid terms that rely on the change in induced dipoles with respect to an atomic displacement. Therefore, the induced dipoles in Eq. (4.4.33) are substituted for using Eq. (4.4.32) to yield

$$\mathrm{U}_{\mu} = -\frac{1}{2}\left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}\right)^{t}\mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}\right) \tag{4.4.34}$$

By the chain rule, the polarization energy gradient is

$$\begin{aligned}\frac{\partial \mathrm{U}_{\mu}}{\partial r_{i,\sigma}} = -\frac{1}{2}\Bigg[&\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}\right) + \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}\right)^{t}\frac{\partial \mathbf{C}^{-1}}{\partial r_{i,\sigma}}\left(\mathbf{E}_{\mathrm{d}} + \mathbf{E}_{\mathrm{RF}}\right) \\ &+ \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}\right)^{t}\mathbf{C}^{-1}\left(\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)\Bigg]\end{aligned} \tag{4.4.35}$$

For convenience a mathematical quantity $\boldsymbol{\nu}$ is defined, which is similar to $\boldsymbol{\mu}$, as

$$\boldsymbol{\nu} = \left(\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}}\right)\mathbf{C}^{-1} \tag{4.4.36}$$

We can now greatly simplify Eq. (4.4.35) above using Eqs. (4.4.32) and (4.4.36) along with the identity $\frac{\partial \mathbf{C}^{-1}}{\partial r_{i,\sigma}} = -\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial r_{i,\sigma}}\mathbf{C}^{-1}$ to give

$$\frac{\partial \mathrm{U}_{\mu}}{\partial r_{i,\sigma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}} - \boldsymbol{\nu}^{t}\frac{\partial \mathbf{C}}{\partial r_{i,\sigma}}\boldsymbol{\mu}\right] \quad (4.4.37)$$

Under the direct polarization model, $\mathbf{C}$ is an identity matrix whose derivative is zero, and therefore Eq. (4.4.37) simplifies to

$$\frac{\partial \mathrm{U}_{\mu_{\mathrm{direct}}}}{\partial r_{i,\sigma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right] \quad (4.4.38)$$

The first two terms on the RHS appear in the polarization energy gradient even in the absence of a continuum reaction field and are described elsewhere.[2, 3, 7, 9, 10] The third and fourth terms are specific to GK and can be combined. We require the derivative of the GK reaction field due to permanent multipoles with respect to movement of any atom

$$\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}} = \frac{\partial \mathbf{K}^{(1)}}{\partial r_{i,\sigma}}\mathbf{M} \quad (4.4.39)$$

It is therefore sufficient to describe the gradient of any $\mathbf{K}_{ij}^{(1)}$ sub-matrix of $\mathbf{K}^{(1)}$ as

$$\frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial r_{i,\sigma}} = \left(\frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial r_{i,\sigma}} \right)_{a_i, a_j} + \frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial a_i} \frac{\partial a_i}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial a_j} \frac{\partial a_j}{\partial r_{i,\sigma}} \tag{4.4.40}$$

where,

$$\begin{aligned} \left(\frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial r_{i,\sigma}} \right)_{a_i, a_j} &= \frac{1}{2} \left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial r_{i,\sigma}} \right)_{a_i, a_j} \\ \frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial a_i} \frac{\partial a_i}{\partial r_{i,\sigma}} &= \frac{1}{2} \left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_i} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_i} \right) \frac{\partial a_i}{\partial r_{i,\sigma}} \\ \frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial a_j} \frac{\partial a_j}{\partial r_{i,\sigma}} &= \frac{1}{2} \left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_j} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_j} \right) \frac{\partial a_j}{\partial r_{i,\sigma}} \end{aligned} \tag{4.4.41}$$

The tensors that make up $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_i}$, $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_j}$, $\frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}_{ij}^{(1\ j)}}{\partial a_i}$ and $\frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_j}$ were described in the previous section. In this case there is a torque on the permanent dipoles and quadrupoles due to the reaction field and reaction field gradient of $(\mu + \nu)/2$, respectively.

The full mutual polarization gradient has an additional term compared to the direct polarization gradient, in addition to the implicit difference due to the induced dipoles being converged self-consistently. Specifically, the derivative of the matrix **C** leads to two terms

$$\boldsymbol{\nu}^t \frac{\partial \mathbf{C}}{\partial r_{i,\sigma}} \boldsymbol{\mu} = -\boldsymbol{\nu}^t \left(\frac{\partial \mathbf{T}^{(11)}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}^{(11)}}{\partial r_{i,\sigma}} \right) \boldsymbol{\mu} \tag{4.4.42}$$

The first term on the RHS occurs in vacuum and is described elsewhere[2, 3, 7, 9, 10], however, the final term is specific to GK. The gradient of one sub-matrix of the $\frac{\partial \mathbf{K}^{(11)}}{\partial r_{i,\sigma}}$ supermatrix is

$$\frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial r_{i,\sigma}} = \left(\frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial r_{i,\sigma}} \right)_{a_i, a_j} + \frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial a_i} \frac{\partial a_i}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial a_j} \frac{\partial a_j}{\partial r_{i,\sigma}} \tag{4.4.43}$$

The expression for the gradient of $\mathbf{K}_{ij}^{(11)}$ is simpler than those for the other GK interaction matrices because it is symmetric.

The veracity of the AMOEBA/GK energy gradients was checked using finite-differences of the energy, optimization of proteins to an RMS convergence criterion of 10^{-4} kcal/mole/Å and constant energy molecular dynamics. For example, at a mean temperature of 300 K the protein 1ETL showed a mean total energy of -361.20 kcal/mole with a standard deviation of just 0.25 kcal/mole over 1 nsec. The simulation started with a total energy of -361.25 and finished at -365.28 kcal/mole.

4.5 Validation and Application

GK is an approximation to the Poisson solution that extends GB to arbitrary order polarizable atomic multipoles. Here we test GK by comparing to numerical PMPB solutions in the limit of using a van der Waals definition of the solute-solvent interface parameterized using the Bondi radii set.[100] Specifically, the electrostatic solvation free

energy and total solvated dipole moment for a series of 55 proteins was compared using the PMPB and GK continuums. This test set based on PDB entries was recently proposed by Tjong and Zhou for studying the accuracy of analytic solvation models and is characterized by structures with less than 10% sequence identity, resolution better than 1.0 Å and less than 250 residues.[98] Amino acids with missing side-chains were changed to alanine if the C_β carbon was present and glycine if not. The TINKER pdbxyz program added missing hydrogen atoms. Histidine residues were made neutral with the δ-nitrogen protonated. All structures were optimized in vacuum to an RMS gradient of 5.0 kcal/mole/Å, with the goal being to remove bad contacts. The average heavy atom RMS distance from the crystal structure was 0.07 Å after optimization.

4.5.1 Electrostatic Solvation Free Energy of Proteins

Previous studies have shown that given accurate effective radii, GB predicts the electrostatic solvation energy of proteins to a mean unsigned error of approximately 1% relative to numerical Poisson calculations. In this section we investigate whether it is reasonable to expect similar performance from GK by comparing the electrostatic solvation free energy for a series of folded proteins to values computed using the PMPB model.

The PMPB calculations used a grid spacing of 0.31 Å and at least 8 Å between the edge of the solute-solvent boundary and the grid boundary. A finer grid spacing of 0.23 Å was also tried, which lowered the PMPB energy by approximately 2%, but did not

change the quality of the agreement between the two models. The interior of the protein was assigned a permittivity of 1.0 while the solvent was set to 78.3. The induced dipoles were deemed to have converged at a tolerance of 0.01 RMS Debye. Converging to a tighter tolerance of 10^{-6} RMS Debye only changed the electrostatic solvation free energy by 0.1% relative to the loser criteria, and was therefore deemed unnecessary. The constant in generalizing function c_f was optimized by hand to eliminate systematic error, which was found to occur at a value of 2.455.

The results are shown in Table 4.8. The mean signed relative difference is 0.0% a result of tuning the cross-term parameter. The mean unsigned relative difference is 0.9%, which is comparable to the most accurate GB methods. We anticipate using a different cross-term parameter when optimizing GK to reproduce PMPB calculations based on a molecular surface definition.

Table 4.8. The electrostatic solvation free energy (kcal/mole) for 55 proteins within the PMPB and GK continuum models. The number of atoms and total charge of each protein is listed along with the signed and unsigned relative difference of the GK model to PMPB.

			Energy		**% Difference**	
	N_{atoms}	**Q**	**PMPB**	**GK**	**Signed**	**Unsigned**
1A6M	2435	2	-2831	-2765	2.3	2.3
1AHO	936	0	-1161	-1158	0.3	0.3
1BYI	3383	-4	-3861	-3873	-0.3	0.3
1C75	985	-6	-1733	-1742	-0.5	0.5
1C7K	1927	-5	-2523	-2481	1.7	1.7
1CEX	2867	1	-3161	-3212	-1.6	1.6
1EB6	2566	-15	-5044	-5042	0.1	0.1
1EJG	642	0	-580	-614	-6.0	6.0
1ETL	140	-1	-246	-247	-0.5	0.5
1EXR	2240	-25	-8656	-8620	0.4	0.4

1F94	967	2	-1240	-1226	1.1	1.1
1F9Y	2535	-5	-2964	-2968	-0.2	0.2
1G4I	1842	-1	-2356	-2345	0.5	0.5
1G66	2794	-2	-2826	-2824	0.1	0.1
1GQV	2135	7	-2708	-2723	-0.6	0.6
1HJE	175	1	-264	-269	-2.0	2.0
1IQZ	1171	-17	-4663	-4729	-1.4	1.4
1IUA	1207	-1	-1400	-1419	-1.4	1.4
1J0P	1597	8	-2975	-2934	1.4	1.4
1K4I	3253	-6	-4085	-4099	-0.3	0.3
1KTH	885	0	-1469	-1448	1.4	1.4
1L9L	1226	11	-3182	-3150	1.0	1.0
1M1Q	1236	-4	-2084	-2077	0.3	0.3
1NLS	3564	-7	-4743	-4756	-0.3	0.3
1NWZ	1912	-6	-2768	-2760	0.3	0.3
1OD3	1893	-3	-2105	-2104	0.0	0.0
1OK0	1076	-5	-1578	-1571	0.5	0.5
1P9G	519	4	-814	-817	-0.4	0.4
1PQ7	3065	4	-2946	-2942	0.1	0.1
1R6J	1230	0	-1486	-1477	0.6	0.6
1SSX	2755	8	-3000	-2980	0.7	0.7
1TG0	1029	-12	-3017	-3014	0.1	0.1
1TQG	1660	-7	-2920	-2900	0.7	0.7
1TT8	2676	1	-2762	-2758	0.1	0.1
1U2H	1495	2	-2038	-2002	1.8	1.8
1UCS	997	0	-1027	-1042	-1.4	1.4
1UFY	1911	0	-2130	-2145	-0.7	0.7
1UNQ	1947	-1	-3217	-3155	1.9	1.9
1VB0	913	3	-1246	-1232	1.1	1.1
1VBW	1056	8	-1931	-1927	0.2	0.2
1W0N	1756	-5	-2380	-2356	1.0	1.0
1WY3	560	1	-750	-747	0.3	0.3
1X6Z	1720	-1	-2170	-2198	-1.3	1.3
1X8Q	2815	-1	-3739	-3714	0.7	0.7
1XMK	1268	1	-1723	-1724	0.0	0.0
1YK4	774	-8	-1893	-1920	-1.4	1.4
1ZZK	1243	1	-1730	-1699	1.8	1.8
2A6Z	3430	-3	-4203	-4186	0.4	0.4
2BF9	560	-2	-933	-940	-0.8	0.8
2CHH	1624	-3	-2128	-2131	-0.1	0.1
2CWS	3400	-3	-3651	-3616	1.0	1.0

2ERL	567	-6	-1178	-1179	0.0	0.0
2FDN	731	-8	-1746	-1796	-2.9	2.9
2FWH	1830	-6	-2495	-2502	-0.3	0.3
3LZT	1960	8	-2754	-2723	1.1	1.1
Mean	1692	-1.9	-2458	-2454	0.0	0.9

4.5.2 Dipole Moment of Solvated Proteins

The change in dipole moment as a function of environment for a polarizable solute is a relevant observable in terms of validating GK because it indicates whether or not the reaction field strength is consistent. The PMPB calculations are exactly equivalent to those described in the previous section. Furthermore, the same constant was used in the GK cross-term. In Table 4.9 it is observed that the total dipole moment of proteins within the GK continuum achieve a mean signed relative difference of -2.7% and a mean unsigned percent difference of 2.7%. This indicates a small, but systematic underestimation of the reaction field. In all cases, for both PMPB and GK models, the reaction field factor was greater than one, except for 1P9G. In this case, the vacuum dipole moment decreased from 18 to 15 and 13 Debye in the PMPB and GK models, respectively. Overall, the mean reaction field factor for the 55 proteins was 1.28 in the PMPB model and 1.24 in GK.

Table 4.9. The total dipole moment (Debye) for 55 proteins in vacuum and within the PMPB and GK continuum models are presented. The signed and unsigned percent error of the GK model relative to PMPB is given along with the reaction field factor under both models.

	Dipole Moment	**% Difference**	**Reaction Field Factor**

	Vacuum	PMPB	GK	Signed	Unsigned	PMPB	GK
1A6M	191.5	252.1	242.6	-3.7	3.7	1.32	1.27
1AHO	119.3	143.6	142.6	-0.7	0.7	1.20	1.20
1BYI	295.8	357.4	343.1	-4.0	4.0	1.21	1.16
1C75	125.0	167.2	165.7	-0.9	0.9	1.34	1.33
1C7K	229.3	310.3	302.7	-2.4	2.4	1.35	1.32
1CEX	451.0	599.7	574.3	-4.2	4.2	1.33	1.27
1EB6	217.9	281.0	274.6	-2.3	2.3	1.29	1.26
1EJG	37.4	49.0	48.9	-0.3	0.3	1.31	1.31
1ETL	29.3	42.9	41.2	-3.8	3.8	1.46	1.41
1EXR	352.5	395.6	384.2	-2.9	2.9	1.12	1.09
1F94	90.7	116.7	113.0	-3.2	3.2	1.29	1.25
1F9Y	138.4	166.0	161.9	-2.5	2.5	1.20	1.17
1G4I	87.9	102.1	97.9	-4.1	4.1	1.16	1.11
1G66	226.5	279.9	273.5	-2.3	2.3	1.24	1.21
1GQV	314.6	394.5	385.2	-2.4	2.4	1.25	1.22
1HJE	48.3	61.2	60.4	-1.4	1.4	1.27	1.25
1IQZ	86.1	110.7	107.2	-3.1	3.1	1.29	1.25
1IUA	107.5	146.1	141.5	-3.2	3.2	1.36	1.32
1J0P	105.2	148.7	142.3	-4.3	4.3	1.41	1.35
1K4I	130.1	163.0	159.6	-2.1	2.1	1.25	1.23
1KTH	117.1	152.1	148.9	-2.1	2.1	1.30	1.27
1L9L	422.8	525.9	517.0	-1.7	1.7	1.24	1.22
1M1Q	261.7	318.1	311.2	-2.2	2.2	1.22	1.19
1NLS	244.9	331.8	313.0	-5.7	5.7	1.35	1.28
1NWZ	83.2	130.2	126.9	-2.5	2.5	1.56	1.53
1OD3	115.2	165.9	160.9	-3.0	3.0	1.44	1.40
1OK0	149.4	193.7	189.1	-2.4	2.4	1.30	1.27
1P9G	17.7	14.6	13.0	-10.7	10.7	0.82	0.74
1PQ7	46.4	49.6	49.1	-1.1	1.1	1.07	1.06
1R6J	86.8	108.8	106.7	-1.9	1.9	1.25	1.23
1SSX	66.0	93.8	89.9	-4.2	4.2	1.42	1.36
1TG0	236.9	316.8	311.1	-1.8	1.8	1.34	1.31
1TQG	355.4	489.5	477.3	-2.5	2.5	1.38	1.34
1TT8	339.6	450.3	434.3	-3.6	3.6	1.33	1.28
1U2H	157.1	206.0	200.6	-2.6	2.6	1.31	1.28
1UCS	111.1	133.0	132.9	0.0	0.0	1.20	1.20
1UFY	94.0	105.9	102.3	-3.4	3.4	1.13	1.09
1UNQ	601.1	735.2	718.8	-2.2	2.2	1.22	1.20
1VB0	132.2	158.2	155.0	-2.0	2.0	1.20	1.17
1VBW	94.4	117.0	114.0	-2.6	2.6	1.24	1.21

1W0N	114.9	155.4	150.0	-3.5	3.5	1.35	1.31
1WY3	63.7	96.4	93.6	-3.0	3.0	1.51	1.47
1X6Z	294.2	366.7	355.9	-2.9	2.9	1.25	1.21
1X8Q	183.8	244.2	237.6	-2.7	2.7	1.33	1.29
1XMK	272.8	356.1	347.0	-2.6	2.6	1.31	1.27
1YK4	66.1	83.7	83.6	-0.2	0.2	1.27	1.26
1ZZK	195.2	246.5	241.6	-2.0	2.0	1.26	1.24
2A6Z	84.1	105.0	101.4	-3.4	3.4	1.25	1.21
2BF9	255.7	290.6	288.4	-0.7	0.7	1.14	1.13
2CHH	267.4	335.7	329.2	-1.9	1.9	1.26	1.23
2CWS	168.6	220.5	211.0	-4.3	4.3	1.31	1.25
2ERL	81.2	108.1	105.6	-2.3	2.3	1.33	1.30
2FDN	78.3	93.2	93.4	0.3	0.3	1.19	1.19
2FWH	104.9	146.3	142.9	-2.3	2.3	1.39	1.36
3LZT	178.5	214.6	209.8	-2.3	2.3	1.20	1.18
Mean	173.2	220.9	215.0	-2.7	2.7	1.28	1.24

5 Implicit Solvents for the AMOEBA Force Field

In addition to the PMPB and GK continuum electrostatics models described in Chapter 3 and Chapter 4, respectively, an apolar estimator is needed to complete the thermodynamic cycle that is the basis for an implicit solvent (see Figure 1.1 on p. 5). In this chapter we describe novel cavitation and dispersion terms and outline their parameterization based on explicit water simulations on the solutes listed below in Table 5.1. Given the apolar contribution to the solvation, parameterization of the electrostatic term is completed in order to match experimental solvation free energies.

Table 5.1. The solvent assessable surface area (SASA) and solvent excluded volume (SEV) for the 39 small molecules used to parameterize PMPB and GK based implicit solvents. The solvent assessable surface area (SASA) and solvent excluded volume (SEV) were defined used AMOEBA R_{min} values and solvent probe radius of 1.4 Å.

Molecule	**SASA (Å^2)**	**SEV (Å^3)**
acetic acid	225.88	297.52
formic acid	183.89	224.99
ethanol	226.35	299.53
isopropanol	258.69	362.42
methanol	186.76	229.55
propanol	259.81	363.19
acetaldehyde	211.83	272.59
formaldehyde	167.45	197.74
butane	281.69	404.59
methane	170.27	201.91

acetamide	230.81	307.25
dimethylacetamide	297.95	436.56
dimethylformamide	267.11	376.19
formamide	190.86	236.68
n-methylacetamide	266.25	373.63
n-methylformamide	233.13	310.55
propamide	265.75	373.14
ammonia	147.74	165.97
dimethylamine	232.97	311.98
ethylamine	233.70	313.34
methylamine	194.49	242.86
propylamine	267.55	377.30
trimethylamine	263.06	372.23
benzene	275.86	394.35
cresol	322.46	482.36
ethylbenzene	342.77	523.07
phenol	288.66	418.47
toluene	311.18	460.87
ethylimidazole	314.06	463.89
imidazole	239.56	325.14
ethylindole	386.33	611.36
indole	325.51	489.46
n-methylpyrrolidine	305.27	456.54
pyrrolidine	272.80	392.53
dimethylsulfide	242.22	329.40
ethanethiol	243.52	332.48
methylethylsulfide	276.55	394.35
methanethiol	206.28	264.15
water	133.89	143.70

5.1 Cavitation Free Energy

The Lum-Chandler-Weeks theory of hydrophobicity predicts contrasting behavior for the cavitation free energy of small and large solutes.[103-106] At all length scales, the driving force for phase separation is proportional to solute volume, while the cost to form an interface is proportional to surface area. These competing factors manifest in a cross-

over in the dependence of the cavitation free energy between volume scaling for small solutes and surface area scaling for large solutes, which occurs for a spherical cavity at a radius of approximately 1 nm.

$$\Delta G(r) \propto \begin{cases} \text{Volume} & r < \sim 1 \text{ nm} \\ \text{Cross-Over} & \sim 1 \text{ nm} \leq r \leq \sim 2 \text{ nm} \\ \text{Surface Area} & \sim 2 \text{ nm} < r \end{cases} \tag{5.1.1}$$

For solutes with more general shapes, such as biomolecules, the cavitation cost is neither proportional to volume nor surface area, but rather some local mixture of the two regimes. For example, the cost to form a cavity for an extended chain would scale with more volume character than would a compact spherical conformation with similar surface areas. One can imagine protein conformations that have both extended loops and large compact regions, suggesting that ad-hoc surface area or volume cavitation terms that do not consider local conformation are too simplistic. It is beyond the scope of the current work to develop a general functional form for the cavitation free energy of a solute of arbitrary size and shape, although recent work that attempts to adjust effective surface tension based on local context is promising.[107] Fortunately, as small molecule cavitation free energies are simply proportional to volume, the magnitude of the cavitation term in AMOEBA implicit solvents is only anticipated to change for macromolecules, but not the small molecule parameterization discussed here.

5.1.1 Cavitation Measurements

Solutes were simulated in explicit water by defining purely repulsive, but smooth, solute-solvent interactions according to

$$\mathrm{U}_{\mathrm{rep}}\left(r_{ij}\right)=\begin{cases}\mathrm{U}_{14-7}\left(r_{ij}\right)+\varepsilon_{ij} & r_{ij}<r_{ij}^{0}\\ 0 & r_{ij}\geq r_{ij}^{0}\end{cases} \tag{5.1.2}$$

where ε_{ij} and r_{ij}^{0} are the potential well depth and minimum energy distance for the buffered 14-7 potential $\mathrm{U}_{14\text{-}7}\left(r_{ij}\right)$ used in the AMOEBA force field, respectively, given by

$$\mathrm{U}_{14-7}\left(r_{ij}\right)=\varepsilon_{ij}\left(\frac{1.07r_{ij}^{0}}{r_{ij}+0.07r_{ij}^{0}}\right)^{7}\left(\frac{1.12\left(r_{ij}^{0}\right)^{7}}{r_{ij}^{7}+0.12\left(r_{ij}^{0}\right)^{7}}-2\right) \tag{5.1.3}$$

where r_{ij} is the separation between atomic sites i and j. Combining rules for heterogeneous pairs are given by

$$\varepsilon_{ij}=\frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\varepsilon_{ii}^{1/2}+\varepsilon_{jj}^{1/2})^{2}} \tag{5.1.4}$$

and

$$r_{ij}^{0}=\frac{(r_{ii}^{0})^{3}+(r_{jj}^{0})^{3}}{(r_{ii}^{0})^{2}+(r_{jj}^{0})^{2}} \tag{5.1.5}$$

Solute electrostatics, both permanent multipoles and atomic polarizabilities, were set to zero.

All simulations were run under the NPT ensemble for 600 psec, with the first 100 psec discarded from subsequent analysis. Initially 216 water molecules were used to solvate the purely repulsive solute, but the equilibrated densities were in the range of 0.95 g/cc, well below 1.0 g/cc. For the acetic acid simulation increasing the box size to 512 water molecules lead to a mean density of 0.991 with standard deviation of 0.011 over the 500 psec of data collection, which was considered acceptable. For all simulations the Berendsen weak coupling thermostat and barostat were employed with time constants of 0.1 and 2.0 psec, respectively.[93] Long range electrostatics were treated using particle mesh Ewald (PME) summation with a cutoff for real space interactions of 7.0 Å and an Ewald coefficient of 0.54 $Å^{-1}$.[94] The PME methodology used tinfoil boundary conditions, a 54 x 54 x 54 charge grid and 6th order B-spline interpolation. van der Waals interactions were smoothly truncated to zero at 12.0 Å using a switching window of width 1.2 Å. Simulations were run using TINKER version 4.2, although custom modifications were required to correctly account for changes in box size due to the NPT conditions during analysis of trajectories.[95] Specifically, the box size for each snap shot was saved during the sampling runs and read back in during reprocessing.

The Gibbs free energy to increase or decrease the SASA or SEV of a solute, and therefore the surface tension (ST) or solvent pressure (SP), respectively, can be determined by a novel free energy perturbation approach. Specifically, the free energy change in moving from a state with potential energy defined by $U(\lambda_0)$ to a state with potential energy $U(\lambda_1)$ can be computed using the Zwanzig relationship[108]

$$\Delta G_{0\to 1} = -k_B T \ln\left\langle e^{-\left(U(\lambda_0)-U(\lambda_1)\right)/k_B T}\right\rangle_{\lambda_0},$$

where k_B is Boltzmann's constant, T is the absolute temperature, λ is a coupling parameter that defines a continuous transformation between solute SASA or SEV values, and $\langle\ \rangle_{\lambda_0}$ indicates an ensemble averaging in state λ_0.

To compute SP and ST, a transformation was defined by moving all atoms toward or away from the center of mass by a displacement of 0.01 Å. This distance was chosen such that the change in free energy was much less than $k_B T$. In this way the potential energy depends on the coupling parameter as follows

$$U\left(\mathbf{X} \pm \lambda \cdot 0.01 \cdot \hat{\mathbf{X}}, \mathbf{Y}\right), \tag{5.1.6}$$

where **X** are the solute coordinates, **Y** are the solvent coordinates, and $\hat{\mathbf{X}}$ contains unit vectors from the solute center of mass to each atom. The surface tension and solvent pressure were then computed using the λ=0 trajectory and a two-sided average based on growing $\Delta G_{0\to G}$ and shrinking $\Delta G_{0\to S}$ the solute size. Note that the solute was not rigid, but was allowed to sample its internal degrees of freedom, such that the change in volume was not constant for each snapshot. In hindsight, it would be more rigorous to leave the solute rigid, but this inconsistency is not expected to have a noticeable effect on the mean solvent pressure used to parameterize the implicit solvent cavitation term. Furthermore, flexible solutes are more appropriate when computing the mean solute-solvent enthalpy, which will be needed in the following section on dispersion free energy.

$$\mathrm{ST} = \frac{1}{2}\left[\frac{\Delta \mathrm{G}_{0\to G}}{\left\langle \mathrm{SASA}_G - \mathrm{SASA}_0 \right\rangle_0} + \frac{\Delta \mathrm{G}_{0\to S}}{\left\langle \mathrm{SASA}_S - \mathrm{SASA}_0 \right\rangle_0}\right] \tag{5.1.7}$$

$$\mathrm{SP} = \frac{1}{2}\left[\frac{\Delta \mathrm{G}_{0\to G}}{\left\langle \mathrm{SEV}_G - \mathrm{SEV}_0 \right\rangle_0} + \frac{\Delta \mathrm{G}_{0\to S}}{\left\langle \mathrm{SEV}_S - \mathrm{SEV}_0 \right\rangle_0}\right] \tag{5.1.8}$$

An alternative approach to computing the cavitation free energy for each small molecule would be to grow in a solute sized cavity; however this entails many more trajectories and is impractical for probing the SP and ST for large macromolecules. Therefore, the approach presented here has been developed with an eye toward using it to collect cavitation target data for validating implicit solvent models at the length scale of proteins and nucleic acids.

The data collected for small molecules is shown below in Table 5.2. We note that the standard deviation of the ST over the test set of small molecules is proportionately larger than that of the SP. This supports the notion that SP is relatively constant over the length scale of small molecules. The mean SP will be used as a parameter for the implicit solvent cavitation term. Furthermore, assuming a constant SP allows a rough estimation of the cavitation free energy for each small molecule, which is useful to estimate the error in this term of the model. Cavitation free energies computed in this manner from explicit water simulations should not be expected to be extremely precise, but are nevertheless useful for judging whether the magnitude of the cavitation term is reasonable at an affordable computational cost.

Table 5.2. Calculated surface tension and solvent pressure are used to determine self-consistent cavitation free energies. The computed standard errors on the ST were all below 0.001 for the ST measurements and below 0.0005 for the SP.

Molecule	**ST (kcal/mol/Å^2)**	**ST x SASA (kcal/mol)**	**SP (kcal/mol/Å^3)**	**SP x SEV (kcal/mol)**
acetic acid	0.0605	13.68	0.0361	10.76
formic acid	0.0599	11.01	0.0376	8.45
ethanol	0.0617	13.96	0.0360	10.77
isopropanol	0.0587	15.18	0.0324	11.72
methanol	0.0553	10.32	0.0344	7.89
propanol	0.0582	15.13	0.0321	11.67
acetaldehyde	0.0549	11.63	0.0333	9.07
formaldehyde	0.0522	8.75	0.0339	6.71
butane	0.0600	16.89	0.0320	12.94
methane	0.0417	7.10	0.0268	5.41
acetamide	0.0630	14.55	0.0371	11.41
dimethylacetamide	0.0633	18.86	0.0341	14.91
dimethylformamide	0.0625	16.71	0.0349	13.14
formamide	0.0542	10.35	0.0337	7.97
n-methylacetamide	0.0638	16.98	0.0355	13.25
n-methylformamide	0.0596	13.90	0.0349	10.83
propamide	0.0595	15.80	0.0331	12.36
ammonia	0.0360	5.32	0.0246	4.08
dimethylamine	0.0565	13.16	0.0320	10.00
ethylamine	0.0564	13.17	0.0324	10.16
methylamine	0.0546	10.62	0.0335	8.13
propylamine	0.0587	15.72	0.0320	12.09
trimethylamine	0.0604	15.88	0.0321	11.96
benzene	0.0670	18.49	0.0367	14.48
cresol	0.0644	20.78	0.0343	16.52
ethylbenzene	0.0630	21.60	0.0328	17.14
phenol	0.0649	18.72	0.0352	14.74
toluene	0.0661	20.58	0.0355	16.36
ethylimidazole	0.0633	19.87	0.0340	15.75
imidazole	0.0682	16.35	0.0385	12.51
ethylindole	0.0659	25.47	0.0336	20.53
indole	0.0687	22.35	0.0365	17.85
n-methylpyrrolidine	0.0617	18.83	0.0314	14.32
pyrrolidine	0.0638	17.40	0.0337	13.22
dimethylsulfide	0.0614	14.86	0.0352	11.58
ethanethiol	0.0577	14.06	0.0329	10.94
methylethylsulfide	0.0636	17.59	0.0346	13.66

methanethiol	0.0567	11.69	0.0345	9.10
water	0.0375	5.02	0.0260	3.74
mean	0.0590		0.0334	
standard deviation	0.0077		0.0028	

5.1.2 Cavitation Model and Parameterization

Given the assumption that LCW theory holds for molecular shapes that are not strictly spherical, including unfolded and folded proteins, it is possible to map a solute conformation $\mathbf{X}$ to an effective radius $r(\mathbf{X})$ using the solvent accessible surface area (SASA).

$$r(\mathbf{X}) = \sqrt{\mathrm{SASA}(\mathbf{X})/4\pi} \tag{5.1.9}$$

The free energy of cavitation can then be modeled as a piecewise continuous function of the effective radius.

$$\Delta G_{\chi}(r) = \begin{cases} \lambda \cdot \frac{4}{3}\pi r^3 & \mathrm{r} \le \chi \\ \gamma \cdot 4\pi r^2 & \chi < \mathrm{r} \end{cases} \tag{5.1.10}$$

In the volume scaling regime cavitation free energy is defined by the product of SP λ (units of kcal/mole/Å^3) with SEV, while in the surface area scaling regime it is defined by the product of ST γ (kcal/mole/Å^2) with SASA. For our model, the SP was parameterized using the simulations described in the previous section, while the limiting ST is conservatively chosen to be 0.080. This is between the known experimental value of 0.103 for the limiting macroscopic case of an infinite air-water interface and the mean

value measured from the explicit water simulations described above. Further refinement based on explicit water protein simulations is anticipated. Given the SP and ST, the cross-over point χ is uniquely defined as

$$\chi = \frac{3 \cdot \gamma}{\lambda} \tag{5.1.11}$$

to give 7.29 Å.

This simple definition is of limited use because the transition between the volume scaling regime and the surface area scaling regime must have continuous first (and ideally second) derivatives to be amenable for molecular dynamics and optimization algorithms. Therefore, we now consider the use of a multiplicative switch $s_v(r)$ to smoothly turn off the volume scaling cavitation energy and a second switch $s_{sa}(r)$ to smoothly turn on the surface area scaling term. Each switch acts over a window of length w centered on the cross-over point.

$$\Delta G_{symmetric}(r) = \begin{cases} \lambda \cdot \frac{4}{3}\pi r^3 & \mathrm{r} \le \chi - w/2 \\ \lambda \cdot \frac{4}{3}\pi r^3 \cdot s_v(r) & \chi - w/2 \le \mathrm{r} < \chi + w/2 \\ +\gamma \cdot 4\pi r^2 \cdot s_{sa}(r) & \\ \gamma \cdot 4\pi r^2 & \chi + w/2 \le \mathrm{r} \end{cases} \tag{5.1.12}$$

The volume scaling switch $s_v(r)$ is a 5th order polynomial whose 6 coefficients are uniquely determined by constraining its value to be 0 at $\chi - w/2$ and 1 at $\chi + w/2$, as well as requiring its first and second derivatives at that these locations to vanish.

$$
\begin{aligned}
s_v(r) &= c_5 r^5 + c_4 r^4 + c_3 r^3 + c_2 r^2 + c_1 r + c_0, \\
c_5 &= 6/d \\
c_4 &= -15(b+e) \\
c_3 &= 10\left(b^2 + 4be + e^2\right) \\
c_2 &= -30eb(b+e) \\
c_1 &= 30b^2e^2 \\
c_0 &= e^3\left(e^2 - 5be + 10b^2\right) \\
d &= b^5 - 10b^2e^2 + 5be^4 - e^5 - 5b^4e + 10e^2b^3
\end{aligned}
\tag{5.1.13}
$$

where b is the radius where the switching begins and e is the radius where the switching ends. The surface area scaling switch in this symmetric case is

$$
s_{sa}(r) = 1 - s_v(r) \tag{5.1.14}
$$

The behavior of the symmetric switched cavitation free energy shows a modest peak at the cross-over point, which was removed using an asymmetric switch

$$
\Delta G_{asymmetric}(r) = \begin{cases} \lambda \cdot \frac{4}{3}\pi r^3 & \mathrm{r} \le \chi - w \\ \lambda \cdot \frac{4}{3}\pi r^3 \cdot s_v(r) + \gamma \cdot 4\pi r^2 \cdot s_{sa}(r) & \chi - w \le \mathrm{r} < \chi + w \\ \gamma \cdot 4\pi r^2 \cdot s_{sa}(r) & \chi + w \le \mathrm{r} < \chi + w + o \\ \gamma \cdot 4\pi r^2 & \chi + w + o \le \mathrm{r} \end{cases} \tag{5.1.15}
$$

where the window w is 3.5 Å and the offset o is 0.4 Å such that surface area scaling is switched off more quickly than in the symmetric case.

The quality of the resulting model for small molecules is shown below in Figure 5.1. The mean unsigned difference between the cavitation free energy computed via the molecule specific (actual) SP given in Table 5.2 and that based on the mean (constant) SP is 0.56 kcal/mol. The only apparent systematic error is for very small molecules with

volumes of approximately 200 Å^3 or less, including water, ammonia and methane. For comparison, we also present a cavitation model based on ST in Figure 5.2, which shows a mean unsigned difference of 1.03 kcal/mol. The ST model overestimates the cavitation free energy for the smallest molecules in the parameterization set, but underestimates it for largest solutes. This supports the physical picture that SP is relatively constant for solutes with an effective radius below about 1 nm, while ST is not.

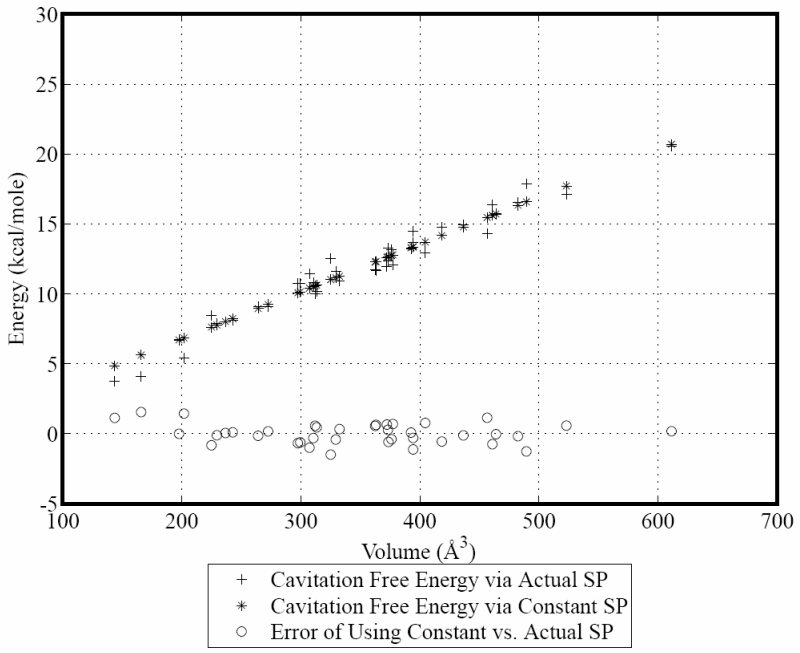

Figure 5.1. Cavitation free energy for AMOEBA small molecules via SP.

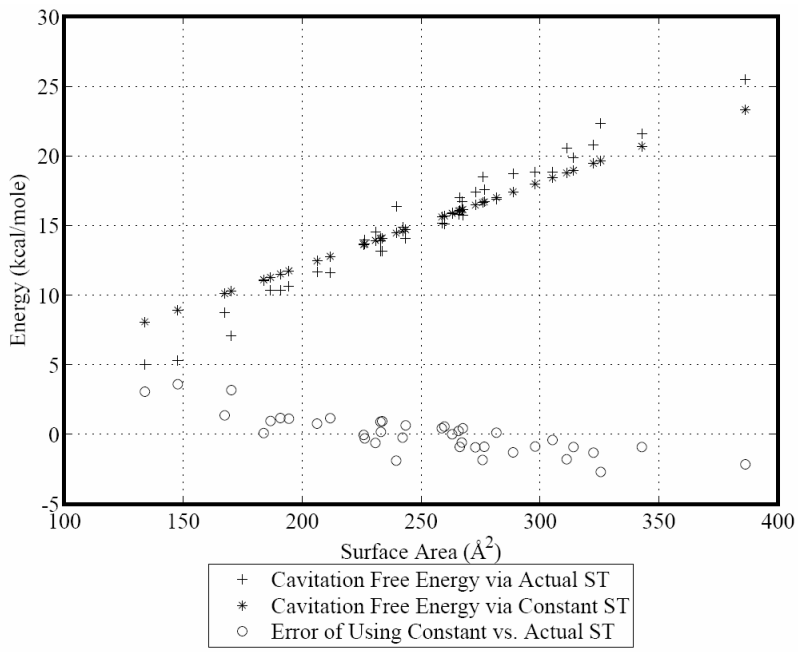

Figure 5.2. Cavitation free energy for AMOEBA small molecules via ST.

5.2 Dispersion Free Energy

Work by Gallicchio, Kubo and Levy has demonstrated that the free energy of adding dispersion interactions to the WCA repulsive potential, thereby restoring the full Lennard-Jones interaction, is very nearly equal to the change in solute-solvent enthalpy for a series of small alkanes studied using free energy perturbation (FEP).[109]

$$\Delta G_{disp} \simeq \langle U_{14\text{-}7} \rangle - \langle U_{rep} \rangle \tag{5.2.1}$$

This lead to their suggestion of a dispersion free energy estimator based on Born radii, such that the dispersion free energy of atom *i* is

$$\Delta G_{GKL} = \sum_{i=1}^{n} \frac{-16\pi\rho_w \varepsilon_{iw} \sigma_{iw}^6}{3R_i^3}, \tag{5.2.2}$$

where ρ_w is the number density of water, ε_{iw} and σ_{iw} are the well depth and sigma value of the interaction of atom *i* with the TIP3P water model, respectively, n is the number of solute atoms and R_i is the Born radius.[47, 48] In effect, the term acts like a tail correction, assuming solvent to be a continuum outside the solute and integrating the $1/r^6$ attractive portion a 6-12 Lennard- Jones potential. In the limit of a spherical solute, use of the Born radii in Eq. (5.2.2) is exact, however for other geometries it is an approximation.

5.2.1 Dispersion Measurements

To obtain parameterization data for the dispersion term, a second set of explicit water simulations were completed in analogous fashion to those described in the previous section on cavitation, except that the solute-solvent interactions were calculated with the full buffered 14-7 potential rather than the WCA repulsive potential $\mathrm{U}_{\mathrm{rep}}\left(r_{ij}\right)$ given in Eq. (5.1.2). As before, the solute multipoles and polarizabilities were set to zero. The average solute-solvent enthalpy was calculated for both sets of simulations and the results are shown below in Table 5.3. The standard error for the computed solute-solvent enthalpies was less than 0.05 kcal/mol in all cases, and therefore their sum is always below 0.1 kcal/mole. Also given are the results of our novel analytic dispersion free energy model $\Delta \mathrm{G}_{\mathrm{disp}}$ described in the next section.

Table 5.3. The average solute-solvent enthalpy was calculated from two sets of explicit water simulations as described in the text. Taking their difference gives an estimate for the dispersion free energy. The value of the implicit solvent dispersion term is shown in the 4th column, along with its error relative to the explicit water estimate. All values are in kcal/mol.

Molecule	$\langle U_{rep} \rangle$	$\langle U_{14\text{-}7} \rangle$	$\langle U_{14\text{-}7} \rangle - \langle U_{rep} \rangle$	ΔG_{disp}	**Signed Error**	**Unsigned Error**
acetic acid	1.2	-6.9	-8.1	-8.4	-0.3	0.3
formic acid	1.0	-4.9	-6.0	-6.1	-0.2	0.2
ethanol	1.3	-6.3	-7.6	-7.9	-0.3	0.3
isopropanol	1.3	-7.9	-9.2	-9.5	-0.3	0.3
methanol	1.0	-4.4	-5.4	-5.6	-0.2	0.2
propanol	1.3	-8.2	-9.5	-9.8	-0.3	0.3
acetaldehyde	1.1	-6.1	-7.2	-7.4	-0.3	0.3
formaldehyde	0.9	-4.0	-4.9	-5.0	0.0	0.0
butane	1.4	-9.4	-10.8	-10.7	0.1	0.1
methane	0.9	-3.5	-4.4	-4.4	-0.1	0.1
acetamide	1.3	-7.5	-8.8	-9.0	-0.2	0.2
dimethylacetamide	1.5	-11.1	-12.6	-12.5	0.1	0.1
dimethylformamide	1.4	-9.4	-10.8	-10.9	-0.2	0.2
formamide	1.0	-5.6	-6.6	-6.8	-0.2	0.2
n-methylacetamide	1.4	-9.5	-10.9	-11.2	-0.2	0.2
n-methylformamide	1.2	-7.7	-8.9	-9.1	-0.2	0.2
propamide	1.3	-9.2	-10.5	-10.7	-0.2	0.2
ammonia	0.8	-2.6	-3.4	-3.4	0.0	0.0
dimethylamine	1.2	-7.0	-8.1	-8.3	-0.2	0.2
ethylamine	1.2	-7.0	-8.2	-8.2	0.0	0.0
methylamine	1.1	-5.0	-6.1	-6.1	0.0	0.0
propylamine	1.3	-8.8	-10.1	-10.0	0.1	0.1
trimethylamine	1.3	-8.7	-10.0	-9.8	0.2	0.2
benzene	1.4	-10.5	-11.8	-11.7	0.1	0.1
cresol	1.5	-12.6	-14.2	-14.1	0.1	0.1
ethylbenzene	1.6	-13.5	-15.1	-14.7	0.5	0.5
phenol	1.4	-11.0	-12.4	-12.5	-0.1	0.1
toluene	1.6	-12.0	-13.5	-13.3	0.2	0.2
ethylimidazole	1.5	-12.4	-14.0	-13.6	0.4	0.4
imidazole	1.3	-8.7	-10.0	-10.3	-0.3	0.3
ethylindole	1.8	-17.8	-19.6	-17.9	1.6	1.6
indole	1.6	-14.7	-16.3	-15.6	0.7	0.7
n-methylpyrrolidine	1.5	-11.4	-13.0	-11.9	1.1	1.1
pyrrolidine	1.4	-9.8	-11.2	-10.6	0.6	0.6

dimethylsulfide	1.3	-8.4	-9.7	-10.1	-0.4	0.4
ethanethiol	1.3	-8.2	-9.4	-9.7	-0.3	0.3
meetsulfide	1.4	-10.1	-11.6	-12.0	-0.4	0.4
methanethiol	1.1	-6.3	-7.4	-7.7	-0.3	0.3
water	0.7	-1.8	-2.6	-2.5	0.1	0.1
mean			-9.7	-9.7	0.0	0.3

5.2.2 Dispersion Model and Parameterization

Our goal for the dispersion free energy model was to remove use of the Born radii from the GKL model ΔG_{GKL} given in Eq. (5.2.2) and instead integrate the true WCA attractive potential outside of the solute cavity for each atom. This analytic approach is based on the HCT pairwise descreening method used for GK.[32, 33] As described in the previous section on cavitation, the AMOEBA Lennard-Jones interactions are based on a buffered-14-7 potential. Therefore, the underlying pairwise integration machinery will need to integrate the constant portion of the WCA potential for $r < R_{io}$ and both $1/r^7$ and $1/r^{14}$ elsewhere. Here R_{io} is the minimum energy separation for solute atom i with an AMOEBA water oxygen. The general analytic form for the dispersion free energy $\Delta \mathrm{G}_{\mathrm{disp}}(\mathbf{X})$ of a solute with coordinates **X** is then given by

$$\Delta \mathrm{G}_{\mathrm{disp}}(\mathbf{X}) = \rho_w \sum_{i=1}^{n} \int_{R_i}^{\infty} \int_0^{\pi} \int_0^{2\pi} \mathrm{U}_{\mathrm{WCA}}(r) \mathrm{W}(r, \theta, \phi, \mathbf{X}, \mathbf{R}) \sin\theta r^2 d\phi d\theta dr \qquad (5.2.3)$$

where W takes the value unity if the point (r,θ,ϕ) is located in the solvent, but zero otherwise, ρ_w is the number density of water and **R** are the set of intrinsic radii that specify the solute cavity for purposes of the dispersion calculation. These are set to

$$R_i = R_{\min,i} + d \tag{5.2.4}$$

where the base radius is the AMOEBA $R_{\min}$ value and d is the single parameter in the model that will be fit against the explicit water simulation results. Inverting the integration domain and applying the HCT pairwise approximation gives

$$\Delta G_{\text{disp}}(\mathbf{X}) = \rho_w \sum_{i=1}^{n} \left[U_{\text{tail}}(R_i) - 4\pi \sum_{j\neq i} \int_L^U U_{\text{WCA}}(r)\, H(r, r_{ij}, sR_i)\, r^2 dr \right] \tag{5.2.5}$$

where H is the fraction of the area of the current spherical integration shell of radius r that is covered by atom j located a distance r_{ij} from atom i and whose radius is scaled to sR_j

$$H(r, r_{ij}, sR_j) = \frac{1}{2} - \frac{1}{4}\frac{r_{ij}^2 + r^2 - (sR_j)^2}{r_{ij} r}. \tag{5.2.6}$$

We note that the scale factor s was parameterized during development of GK as described in section 4.1 (p. 89) and accounts for overlap between the volumes of nearby atoms. The WCA potential uses a simplified form the buffered 14-7 for interactions of solute atoms with water

$$\begin{aligned}\mathrm{U}_{\mathrm{WCA}}(r) &= \mathrm{U}_{\mathrm{WCA,o}}(r) + 2\,\mathrm{U}_{\mathrm{WCA,h}}(r) \\ &= \begin{cases} -\varepsilon_{io} & r < R_{io} \\ \varepsilon_{io} R_{io}^{7}\left(\dfrac{R_{io}^{7}}{r^{14}} - \dfrac{2}{r^{7}}\right) & R_{io} < r \end{cases} \\ &\quad +2\begin{cases} -\varepsilon_{ih} & r < R_{ih} \\ \varepsilon_{ih} R_{ih}^{7}\left(\dfrac{R_{ih}^{7}}{r^{14}} - \dfrac{2}{r^{7}}\right) & R_{ih} < r \end{cases}\end{aligned} \tag{5.2.7}$$

where the well depths and minimum energy distances are based on the mixing rules in (5.1.4) and (5.1.5) for atom i with the AMOEBA water model.[3] The difference between this 14-7 potential and the buffered 14-7 potential is negligible for separations greater than the minimum energy distance, which is the only portion in use. The analytic tail correction based on Eq. (5.2.7) for the interaction with the water oxygen gives

$$\begin{aligned}\mathrm{U}_{\mathrm{tail},o}(R_i) &= \int_{R_i}^{\infty} \mathrm{U}_{\mathrm{WCA},o}(r)\, 4\pi r^2 dr \\ &= \begin{cases} -\dfrac{4}{3}\pi\varepsilon_{io}\left(R_{io}^{3} - R_i\right) - \varepsilon_{io} 18 R_{io}^{3}\pi & r < R_{io}\,, \\ \varepsilon_{io} R_{io}^{7}\pi\left(\dfrac{4R_{io}^{7}}{11R_i^{11}} - \dfrac{2}{R_i^{4}}\right) & R_{io} < r \end{cases}\end{aligned} \tag{5.2.8}$$

and the tail correction for the interaction between a solute atom and hydrogen is analogous.

The final piece to this model is the solution to the integral in Eq. (5.2.5) above. If integration of the WCA dispersion begins inside the minimum energy distance $b < R_{io}$ then a contribution of

$$-\varepsilon_{io}\int_{L}^{U} \mathrm{H}\left(r_{ij}, sR_j\right) r^2 dr = -\varepsilon_{io}\left[-4\pi r^2\left(3r^2 - 8r_{ij}r + 6r_{ij}^2 - 6\left(sR_j\right)^2\right)\Big/ 48 r_{ij}\right]_{L}^{U} \tag{5.2.9}$$

is included. The lower limit L is b or $r_{ij} - \rho_j$, whichever is greater. The upper limit U of this integral is R_{io} or $r_{ij} + \rho_j$, whichever is smaller. If $r_{ij} + \rho_j$ is greater than R_{io}, the integration result outside R_{io} is

$$\begin{aligned}
&\varepsilon_{io} R_{io}^{14} \int_L^U \frac{1}{r^{12}} \mathrm{H}\left(r_{ij}, sR_j\right) dr - 2\varepsilon_{io} R_{io}^{7} \int_L^U \frac{1}{r^{5}} \mathrm{H}\left(r_{ij}, sR_j\right) dr \\
&= 4\pi\varepsilon_{io} R_{io}^{14} \left[\left(-120 r_{ij} r + 66 r^2 + 55 r_{ij}^2 - 55\left(sR_j\right)^2 \right) \Big/ 2640 r_{ij} r^{12} \right]_L^U \\
&-8\pi\varepsilon_{io} R_{io}^{7} \left[\left(-15 r_{ij} r + 10 r^2 + 6 r_{ij}^2 - 6\left(sR_j\right)^2 \right) \Big/ 120 r_{ij} r^{5} \right]_L^U
\end{aligned} \tag{5.2.10}$$

where the upper limit is always $r_{ij} + \rho_j$. As before, the lower limit L is b or $r_{ij} - \rho_j$, whichever is greater, unless this result is inside the minimum energy distance R_{io}. In this case, a contribution up to R_{io} has already been included from Eq. (5.2.9) and L takes the value R_{io}.

Shown above in Table 5.3 are the results of parameterization of this dispersion estimator against the explicit water simulation results. It was found that the optimal value of the parameter d was 0.36 Å. The average unsigned error in this term, given the assumption of Eq. (5.2.1), is only 0.3 kcal/mole. This is a remarkable result and gives confidence that the dispersion free energy can be accurately modeled in a continuum fashion. We also note from Figure 5.3 below that although dispersion free energy is correlated with surface area, they are not strictly proportional. It is obvious that an attempt to fit a line through this data hurts the quality of the model, and therefore combining cavitation and dispersion into a single apolar term is not recommended. It should also be pointed out that use of the HCT overlap scale factor of 0.690 that is

consistent with the Bondi radii used thus far with GK may not be optimal for the larger radii used in the dispersion calculation. Adjustment to this parameter based on dispersion target data for proteins is anticipated.

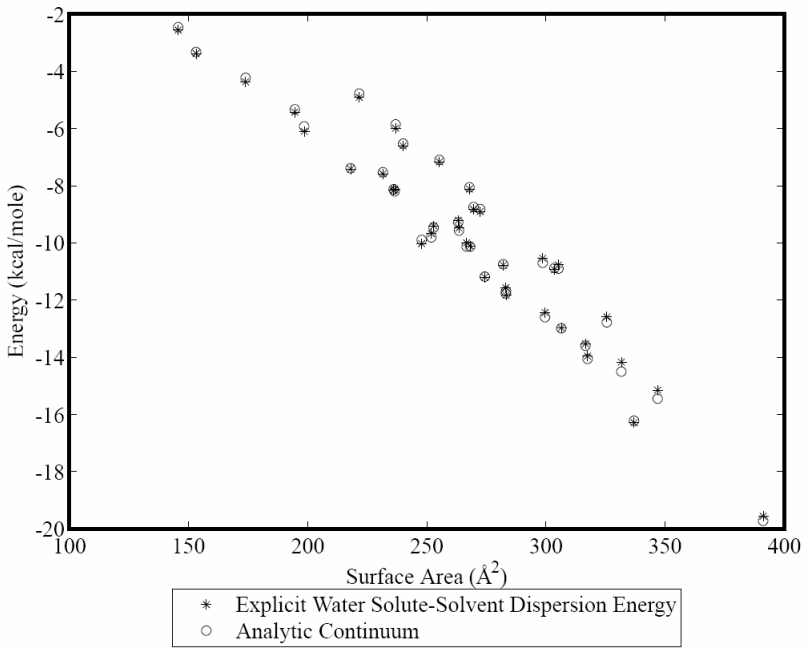

Figure 5.3. A comparison of the analytic continuum dispersion free energy with results from explicit water simulations show good agreement over a range of small molecule sizes.

5.3 Solvation Free Energy of Small Molecules

A subset of the 39 small molecules used in the previous two sections, those with known experimental solvation free energies, will now be used for an initial parameterization of the of the PMPB and GK electrostatic terms.[110] Although a general strategy can be outlined and preliminary indications on the overall quality of the models

will be presented, it is difficult to avoid over-fitting until more AMOEBA small molecules are available, especially those with net charge. If the electrostatic term could be fit with a single parameter, as is the case for the cavitation and dispersion terms, there would be no difficulty. However, at the length scale of small molecules, continuum electrostatics is very sensitive to the definition of the solute-solvent boundary. Therefore, further refinement in the future is unavoidable. On the other hand, for modest sized proteins the total dipole moment appears to be rather insensitive to detailed parameterization (see Table 3.8 on p.71).

A successful strategy for defining the solute-solvent boundary has been presented by Barone, Cassi and Tomasi, which offsets the boundary of functional groups based on solvation free energy trends.[111] Using this approach for both PMPB and GK solute-solvent boundaries gave mean unsigned errors of 0.6 and 0.7 kcal/mol, respectively, as shown in Table 5.4 below.

Table 5.4. Solvation free energy of AMOEBA solutes in both PMPB and GK based implicit solvents compared to experiment. The PMPB and GK values include the same apolar term. All values are in kcal/mol.

	Solvation Energy			**Signed Error**		**Unsigned Error**	
Molecule	**Expt.**	**PMPB**	**GK**	**PMPB**	**GK**	**PMPB**	**GK**
acetic acid	-6.7	-7.5	-7.6	-0.8	-0.8	0.8	0.8
ethanol	-5.0	-5.4	-5.0	-0.4	0.0	0.4	0.0
isopropanol	-4.8	-4.8	-4.5	-0.1	0.3	0.1	0.3
methanol	-5.1	-5.6	-5.4	-0.5	-0.3	0.5	0.3
propanol	-4.8	-5.1	-4.9	-0.3	0.0	0.3	0.0
acetaldehyde	-3.5	-3.7	-2.5	-0.2	1.0	0.2	1.0
formaldehyde	-2.8	-3.7	-2.5	-1.0	0.2	1.0	0.2
butane	2.1	1.7	1.8	-0.4	-0.3	0.4	0.3
methane	2.0	1.6	0.9	-0.4	-1.1	0.4	1.1
acetamide	-9.7	-10.8	-10.3	-1.1	-0.6	1.1	0.6
dimethylacetamide	-8.5	-5.9	-8.6	2.6	-0.1	2.6	0.1

dimethylformamide	-7.8	-8.4	-8.6	-0.6	-0.8	0.6	0.8
(n)-methylacetamide	-10.1	-9.2	-9.4	0.9	0.7	0.9	0.7
(n)-methylformamide	-10.0	-9.9	-9.2	0.1	0.8	0.1	0.8
propamide	-9.7	-10.0	-9.7	-0.3	0.0	0.3	0.0
ammonia	-4.3	-4.2	-7.0	0.1	-2.7	0.1	2.7
dimethylamine	-4.3	-4.4	-4.8	-0.1	-0.5	0.1	0.5
ethylamine	-4.5	-4.4	-3.9	0.1	0.6	0.1	0.6
methylamine	-4.6	-5.2	-4.6	-0.7	0.0	0.7	0.0
propylamine	-4.4	-4.2	-3.7	0.2	0.7	0.2	0.7
trimethylamine	-3.2	-2.2	-2.4	1.0	0.8	1.0	0.8
benzene	-0.9	-2.2	-1.3	-1.3	-0.4	1.3	0.4
cresol	-6.1	-5.8	-5.4	0.4	0.8	0.4	0.8
ethylbenzene	-0.8	-0.5	1.7	0.3	2.5	0.3	2.5
phenol	-6.6	-6.8	-7.2	-0.2	-0.6	0.2	0.6
toluene	-0.9	-1.3	0.3	-0.4	1.2	0.4	1.2
imidazole	-10.3	-12.9	-12.0	-2.6	-1.7	2.6	1.7
pyrrolidine	-5.5	-4.2	-2.4	1.3	3.1	1.3	3.1
dimethylsulfide	-1.5	-1.8	-1.6	-0.3	0.0	0.3	0.0
ethanethiol	-1.3	-1.4	-1.6	-0.1	-0.3	0.1	0.3
methylethylsulfide	-1.5	-1.6	-1.5	-0.1	0.0	0.1	0.0
methanethiol	-1.2	-1.7	-1.8	-0.5	-0.6	0.5	0.6
water	-6.3	-6.6	-6.6	-0.3	-0.3	0.3	0.3
mean				-0.2	0.0	0.6	0.7

6 Spherical Solvent Boundary Potential for Multipoles

A spherical solvent boundary potential (SSBP) uses explicit water molecules inside a spherical domain and a continuum outside to capture the effect of solvation on a system of interest.[112] The advantage of using a spherical boundary is that the Poisson-Boltzmann equation can be solved for the reaction potential at any point within the sphere in terms of an infinite series of Legendre polynomials. This statement holds for any arbitrary collection multipole moments within the spherical boundary.[28] However, the infinite series may converge slowly for complicated charge distributions, or not at all in practice due to the finite precision of numerical calculations.

This motivated us to extend the approximation used by Grycuk for charge-charge interactions within a low dielectric sphere and surrounded by a high dielectric solvent to higher order interactions.[96] This approach was described in section 4.1.2 (p. 86) during derivation of GK. In fact, we note that if the GK cross-term was “perfect”, which it is not, it should produce expressions similar to those below for the interaction between multipole components within a spherical solute. This observation is useful for providing insight into how GK might be improved in the future. However, we note that Grycuk concluded that the widely used GB function performed better than his closed form

solution for the interaction energy between charges $W_{0,0}$ shown below.[96] In effect, this conclusion implies that direct use of the formulas below in a GK-like model produces results that are over-fit to spherical geometry.

6.1 Pairwise Electrostatic Solvation Free Energy

Beginning from the vector formulas for the reaction field energy between point multipoles given in Appendix B of the paper by Kong and Ponder[28], which are not repeated here, we applied Grycuk's approximation.[96] Shown below in Table 6.1 are the results through quadrupole-quadrupole, which is sufficient for implementing a SSBP for the AMOEBA force field.[113] The resulting expressions completely eliminate dependencies on infinite series of Legendre polynomials. The validity of the results is demonstrated by simplifying the general pairwise terms to the special case of self-energies, which is the subject of the next section.

Table 6.1. Closed form expressions for the pairwise electrostatic solvation free energies between two off-center multipole components within a sphere of radius *a* up to quadrupole order are given. The vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are relative to the center of the sphere. When $\mathbf{r}_1 = \mathbf{r}_2$ the formulas are reduced to self-energies, which are given in Table 6.2. Kong and Ponder have previously reported infinite series solutions in terms of Legendre polynomials in Appendix B of their work.[28] The convention for repeated summation over Greek subscripts is assumed and $\hat{\mathbf{r}}$ is a unit vector in the direction $\mathbf{r}$.

W_{l_1,l_2}	**Pairwise electrostatic solvation free energy**
$W_{0,0}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{q_1 q_2}{af}$
$W_{0,1}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q_1\mu_{2,\alpha}\frac{r_1\left(a^2\hat{r}_{1,\alpha}-r_1r_2\hat{r}_{2,\alpha}\right)}{a^5f^3}$
$W_{1,1}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\mu_{1,\alpha}\mu_{2,\beta}\left[\frac{3r_1r_2\left(a^2\hat{r}_{2,\alpha}-r_1r_2\hat{r}_{1,\alpha}\right)\left(a^2\hat{r}_{1,\beta}-r_1r_2\hat{r}_{2,\beta}\right)}{a^9f^5}+\frac{a^2\delta_{\alpha,\beta}-2r_1r_2\hat{r}_{1,\alpha}\hat{r}_{2,\beta}}{a^5f^3}\right]$
$W_{0,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q_1\Theta_{2,\alpha\beta}\frac{1}{3}\left[\frac{3r_1^2\left(a^2\hat{r}_{1,\alpha}-r_1r_2\hat{r}_{2,\alpha}\right)\left(a^2\hat{r}_{1,\beta}-r_1r_2\hat{r}_{2,\beta}\right)}{a^9f^5}-\frac{r_1^2\delta_{\alpha\beta}}{a^5f^3}\right]$
$W_{1,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\mu_{1,\alpha}\Theta_{2,\beta\gamma}\frac{1}{3}\left(15\frac{A}{a^{13}f^7}+\frac{3B}{a^9f^5}-2\frac{r_1\hat{r}_{1,\alpha}\delta_{\beta\gamma}}{a^5f^3}\right)$
$W_{2,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\Theta_{1,\alpha\beta}\Theta_{2,\gamma\delta}\frac{1}{9}\left(-105\frac{C}{a^{17}f^9}-15\frac{D}{a^{13}f^7}-2\frac{E}{a^9f^5}-\frac{2}{3}\frac{\delta_{\alpha\beta}\delta_{\gamma\delta}}{a^9f^5}\right)$

$$f=\sqrt{1-\frac{2\cos(\theta)r_1r_2}{a^2}+\frac{r_1^2r_2^2}{a^4}} \tag{5.2.11}$$

$$A=r_1^2r_2\left(a^2\hat{r}_{2,\alpha}-r_1r_2\hat{r}_{1,\alpha}\right)\left(a^2\hat{r}_{1,\beta}-r_1r_2\hat{r}_{2,\beta}\right)\left(a^2\hat{r}_{1,\gamma}-r_1r_2\hat{r}_{2,\alpha}\right) \tag{5.2.12}$$

$$\begin{aligned}B=r_1\Big[&r_1^2r_2^2\hat{r}_{1,\alpha}\left(\delta_{\beta\gamma}+4\hat{r}_{2,\beta}\hat{r}_{2,\gamma}\right)\\&-r_1r_2a^2\left(2\hat{r}_{1,\alpha}\hat{r}_{1,\beta}\hat{r}_{2,\gamma}+2\hat{r}_{1,\alpha}\hat{r}_{1,\gamma}\hat{r}_{2,\beta}+\hat{r}_{2,\alpha}\delta_{\beta\gamma}+\hat{r}_{2,\beta}\delta_{\alpha\gamma}+\hat{r}_{2,\gamma}\delta_{\alpha\beta}\right)\\&+a^4\left(\hat{r}_{1,\beta}\delta_{\alpha\gamma}+\hat{r}_{1,\gamma}\delta_{\alpha\beta}\right)\Big]\end{aligned} \tag{5.2.13}$$

$$C = r_1^2 r_2^2 \left(a^2 \hat{r}_{2,\alpha} - \hat{r}_{1,\alpha} r_1 r_2\right)\left(a^2 \hat{r}_{2,\beta} - \hat{r}_{1,\beta} r_1 r_2\right) \\ \left(a^2 \hat{r}_{2,\gamma} - \hat{r}_{1,\gamma} r_1 r_2\right)\left(a^2 \hat{r}_{2,\delta} - \hat{r}_{1,\delta} r_1 r_2\right) \tag{5.2.14}$$

$$\begin{aligned} D = r_1 r_2 \Big\{ & -a^6 \left[\hat{r}_{1,\gamma} \left(\hat{r}_{2,\alpha} \delta_{\beta\delta} + \hat{r}_{2,\beta} \delta_{\alpha\delta} \right) + \hat{r}_{1,\delta} \left(\hat{r}_{2,\alpha} \delta_{\beta\gamma} + \hat{r}_{2,\beta} \delta_{\alpha\gamma} \right) \right] \\ & + a^4 r_1 r_2 \Big[\hat{r}_{1,\alpha} \hat{r}_{1,\gamma} \delta_{\beta\delta} + \hat{r}_{1,\alpha} \hat{r}_{1,\delta} \delta_{\beta\gamma} + \hat{r}_{1,\beta} \hat{r}_{1,\delta} \delta_{\alpha\gamma} + \hat{r}_{1,\beta} \hat{r}_{1,\gamma} \delta_{\alpha\delta} + \hat{r}_{1,\gamma} \hat{r}_{1,\delta} \delta_{\alpha\beta} \\ & + \hat{r}_{2,\alpha} \hat{r}_{2,\beta} \delta_{\gamma\delta} + \hat{r}_{2,\alpha} \hat{r}_{2,\gamma} \delta_{\beta\delta} + \hat{r}_{2,\alpha} \hat{r}_{2,\delta} \delta_{\beta\gamma} + \hat{r}_{2,\beta} \hat{r}_{2,\gamma} \delta_{\alpha\delta} + \hat{r}_{2,\beta} \hat{r}_{2,\delta} \delta_{\alpha\gamma} \\ & + 2 \left(\hat{r}_{1,\alpha} \hat{r}_{1,\gamma} \hat{r}_{2,\beta} \hat{r}_{2,\delta} + \hat{r}_{1,\alpha} \hat{r}_{1,\delta} \hat{r}_{2,\beta} \hat{r}_{2,\gamma} + \hat{r}_{1,\beta} \hat{r}_{1,\gamma} \hat{r}_{2,\alpha} \hat{r}_{2,\delta} + \hat{r}_{1,\beta} \hat{r}_{1,\delta} \hat{r}_{2,\alpha} \hat{r}_{2,\gamma} \right) \Big] \\ & - a^2 r_1^2 r_2^2 \Big[\hat{r}_{1,\alpha} \hat{r}_{2,\beta} \delta_{\gamma\delta} + \hat{r}_{1,\alpha} \hat{r}_{2,\gamma} \delta_{\beta\delta} + \hat{r}_{1,\alpha} \hat{r}_{2,\delta} \delta_{\beta\gamma} + \hat{r}_{1,\beta} \hat{r}_{2,\alpha} \delta_{\gamma\delta} \\ & + \hat{r}_{1,\beta} \hat{r}_{2,\gamma} \delta_{\alpha\delta} + \hat{r}_{1,\beta} \hat{r}_{2,\delta} \delta_{\alpha\gamma} + \hat{r}_{1,\gamma} \hat{r}_{2,\delta} \delta_{\alpha\beta} + \hat{r}_{1,\delta} \hat{r}_{2,\gamma} \delta_{\alpha\beta} \\ & + 4 \left(\hat{r}_{1,\alpha} \hat{r}_{1,\beta} \hat{r}_{1,\gamma} \hat{r}_{2,\delta} + \hat{r}_{1,\alpha} \hat{r}_{1,\beta} \hat{r}_{1,\delta} \hat{r}_{2,\gamma} + \hat{r}_{1,\alpha} \hat{r}_{2,\beta} \hat{r}_{2,\gamma} \hat{r}_{2,\delta} + \hat{r}_{1,\beta} \hat{r}_{2,\alpha} \hat{r}_{2,\gamma} \hat{r}_{2,\delta} \right) \Big] \\ & + r_1^3 r_2^3 \left(\hat{r}_{1,\alpha} \hat{r}_{1,\beta} \delta_{\gamma\delta} + \hat{r}_{2,\gamma} \hat{r}_{2,\delta} \delta_{\alpha\beta} + 8 \hat{r}_{1,\alpha} \hat{r}_{1,\beta} \hat{r}_{2,\gamma} \hat{r}_{2,\delta} \right) \Big\} \end{aligned} \tag{5.2.15}$$

$$\begin{aligned} E = \Bigg[& -\frac{a^4}{2} \left(\delta_{\alpha\gamma} \delta_{\beta\delta} + \delta_{\beta\gamma} \delta_{\alpha\delta} \right) \\ & + a^2 r_1 r_2 \left(\hat{r}_{1,\alpha} \hat{r}_{2,\beta} \delta_{\gamma\delta} + \hat{r}_{1,\alpha} \hat{r}_{2,\gamma} \delta_{\beta\delta} + \hat{r}_{1,\alpha} \hat{r}_{2,\delta} \delta_{\beta\gamma} + \hat{r}_{1,\beta} \hat{r}_{2,\alpha} \delta_{\gamma\delta} \right. \\ & \left. + \hat{r}_{1,\beta} \hat{r}_{2,\gamma} \delta_{\alpha\delta} + \hat{r}_{1,\beta} \hat{r}_{2,\delta} \delta_{\alpha\gamma} + \hat{r}_{1,\gamma} \hat{r}_{2,\delta} \delta_{\alpha\beta} + \hat{r}_{1,\delta} \hat{r}_{2,\gamma} \delta_{\alpha\beta} \right) \\ & - 2 r_1^2 r_2^2 \left(2 \hat{r}_{1,\alpha} \hat{r}_{2,\gamma} \hat{r}_{1,\beta} \hat{r}_{2,\delta} + \hat{r}_{1,\alpha} \hat{r}_{1,\beta} \delta_{\gamma\delta} + \hat{r}_{2,\gamma} \hat{r}_{2,\delta} \delta_{\alpha\beta} + \frac{1}{4} \delta_{\alpha\beta} \delta_{\alpha\gamma} \right) \Bigg] \end{aligned} \tag{5.2.16}$$

6.2 Electrostatic Solvation Self-Energy

As a check of the general pairwise terms reported in the previous section, we present two special cases in Table 6.2 below. First, we assume that both multipole components are located at the same point within the sphere by setting $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$, to give self-energies. Next, we further simplify the expressions by restricting the solutions to

self-energies of a multipole at the center of the sphere by setting r = 0. Finally, we mention again that the results of this chapter may be of use in motivating not only a SSBP for AMOEBA, but also future enhancements to GK.

Table 6.2. Here we present closed form expressions for the self-energy for two off-center multipole components at the same site within a spherical solute of radius a. As the multipole approaches the center of the sphere $r \to 0$, the formulas simplify to well-known solutions. Kong and Ponder have previously reported infinite series solutions in Appendix B of their work.[28] The convention for repeated summation over Greek subscripts is assumed and $\hat{\mathbf{r}}$ is a unit vector in the direction **r**.

W'_{l_1,l_2}	**Electrostatic solvation self-energy**	$r = 0$
$W'_{0,0}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{q^2}{2}\frac{a}{\left(a^2-r^2\right)}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{q^2}{2a}$
$W'_{0,1}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q\mu_\alpha\frac{ar\hat{r}_\alpha}{\left(a^2-r^2\right)^2}$	0
$W'_{1,1}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{\mu_\alpha\mu_\beta}{2}\frac{a\left(r^2\hat{r}_\alpha\hat{r}_\beta+a^2\delta_{\alpha\beta}\right)}{\left(a^2-r^2\right)^3}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{\mu_\alpha\mu_\beta\delta_{\alpha\beta}}{2a^3}$
$W'_{0,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q\Theta_{\alpha\beta}\frac{ar^2\hat{r}_\alpha\hat{r}_\beta}{\left(a^2-r^2\right)^3}$	0
$W'_{1,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\mu_\alpha\Theta_{\beta\gamma}\frac{ar}{\left(a^2-r^2\right)^4}$ $\times\left[r^2\hat{r}_\alpha\hat{r}_\beta\hat{r}_\gamma+a^2\left(\hat{r}_\beta\delta_{\alpha\gamma}+\hat{r}_\gamma\delta_{\alpha\beta}\right)\right]$	0
$W'_{2,2}$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{\Theta_{\alpha\beta}\Theta_{\gamma\delta}}{6}\frac{a}{\left(a^2-r^2\right)^5}$ $\times\left[3a^2r^2\left(\hat{r}_\alpha\hat{r}_\gamma\delta_{\beta\delta}+\hat{r}_\alpha\hat{r}_\delta\delta_{\beta\gamma}+\hat{r}_\beta\hat{r}_\gamma\delta_{\alpha\delta}+\hat{r}_\beta\hat{r}_\delta\delta_{\alpha\gamma}\right)\right.$ $\left.+a^4\left(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma}\right)+3r^4\hat{r}_\alpha\hat{r}_\beta\hat{r}_\gamma\hat{r}_\delta\right]$	$\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)$ $\times\frac{\Theta_{\alpha\beta}\Theta_{\gamma\delta}\left(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma}\right)}{6a^5}$

7 Conclusions

All of the methodology described in this dissertation is implemented in the TINKER molecular modeling package.[95] This will facilitate use of PMPB and GK based implicit solvents using a variety of algorithms including molecular dynamics, Monte Carlo and a range of optimization methods. Additionally, parallelization of the LPBE calculations using existing approaches in APBS, which have been applied to fixed partial charge models, would be an important improvement in terms of speed and increasing the size of systems that can be routinely studied. Future work will include validation of the complete implicit solvent models against observables for protein systems.

Further improvements in both the PMPB and GK continuum electrostatics models may depend on reconciling deficiencies that emerge in treating local, specific molecular interactions. For example, both the Clausius-Mossotti[85, 86] and Onsager[5] theories for predicting the permittivity of a liquid break down for those that “associate”, such as water. Here association is defined as short range ordering that leads to correlations in the orientations and positions of neighboring groups, such as hydrogen bonding pairs. Theory by Kirkwood[114] and Fröhlich[115] introduced a correction factor to explicitly account for this deviation from continuum behavior. More recently, Rick and Berne showed that no parameterization of the dielectric boundary for a water molecule in water could

simultaneously fit the electrostatic free energy and reaction potential to within 20%, mainly due to nonlinear electrostriction.[116] This effect, inherent to both numerical and analytic continuum electrostatic models, may be a current limiting factor to their accuracy.

7.1 Polarizable Multipole Poisson-Boltzmann

We have presented methodology required to determine the energy and gradient for the AMOEBA force field in conjunction with numerical solutions to the LPBE, which captures the electrostatic response of solvent by treating it as a dielectric continuum. The PMPB model was then applied to a series of proteins that were also studied using explicit water simulations. The resulting increases in dipole moment found using each approach were in excellent agreement. This indicates that the continuum assumption is a reasonable approximation at the length scale of the systems studied here.

The methodology presented here is also expected to be useful for the development of continuum electrostatics models for coarse grained potentials. For example, Golubkov and Ren have recently described a generalized coarse grain model based on point multipoles and Gay-Berne potentials, which saves several orders of magnitude over all atom models.[117] In addition, we have used the PMPB model as a gold standard in order to test the GK analytic approximation discussed below.

7.2 Generalized Kirkwood

Since its introduction in 1990, GB has proven to be capable of capturing the electrostatic response of the solvent environment to solutes. It has been successfully applied to molecular dynamics simulations, scoring protein conformations and the prediction of binding affinities.[38] However, all GB models are limited in their precision due to truncation at atomic monopoles. Applications of recent interest, including high-resolution homology modeling, design of protein-protein interactions and design of proteins with enzymatic activity may require improved force field electrostatics.[50, 51, 118] We suggest that the AMOEBA force field coupled with the GK continuum model is a promising improvement.[46]

There are two main differences between GB and GK. First, the GK self-energy of a permanent multipole site depends on Kirkwood's solution for the electrostatic solvation energy of a spherical particle with arbitrary charge distribution, which is reduced to Born's formula in the case of a monopole. Second, the GK cross-term is formulated by averaging a simple auxiliary potential for each multipole site, which reduces to the GB cross-term for monopole interactions.

We have implemented GK for the AMOEBA force field, including energy gradients, within the TINKER package.[95] The model was tested against numerical PMPB calculations of the electrostatic solvation free energy for a series of 55 diverse proteins

and showed a mean unsigned percent error of 1.0. The fidelity of the reaction field of GK relative to PMPB can be inferred from the total solvated dipole moment of each protein, which showed GK to have a mean unsigned percent error of 1.4.

The next step in the implementation of GK for AMOEBA solutes was parameterization of a complete implicit solvent model by addition of an apolar term. The overall model was parameterized against neutral small molecules solvation free energies.

GK may be useful for developing new continuum models based on electron densities derived from electronic structure calculations. For example, Cramer and Truhlar have successfully employed GB in their SMX series of solvation models.[65, 66, 71, 119] GK would also offer an analytic alternative to the numerical distributed multipole solvation model of Rinaldi *et al.*[61, 62]

Appendix A Finite-Difference Representation of the LPBE

The finite-difference representation of the LPBE for a uniform grid spacing is

$$\begin{aligned}
&\varepsilon_x(i,j,k)\left[\Phi(i+1,j,k)-\Phi(i,j,k)\right]+\varepsilon_x(i-1,j,k)\left[\Phi(i-1,j,k)-\Phi(i,j,k)\right]\\
&+\varepsilon_y(i,j,k)\left[\Phi(i,j+1,k)-\Phi(i,j,k)\right]+\varepsilon_y(i,j-1,k)\left[\Phi(i,j-1,k)-\Phi(i,j,k)\right]\\
&+\varepsilon_z(i,j,k)\left[\Phi(i,j,k+1)-\Phi(i,j,k)\right]+\varepsilon_z(i,j,k-1)\left[\Phi(i,j,k-1)-\Phi(i,j,k)\right]\\
&+\bar{\kappa}^2(i,j,k)\Phi(i,j,k)h^2=-4\pi\frac{q(i,j,k)}{h}
\end{aligned} \quad ,(A.1)$$

where h is the grid spacing, $\Phi(i,j,k)$ is the electrostatic potential, $\bar{\kappa}^2(i,j,k)$ is the modified Debye-Hückel screening factor and $q(i,j,k)$ is the fractional charge. The permittivity is specified by three separate arrays, ε_x, ε_y and ε_z, where each is shifted along its respective grid branch such that $\varepsilon_x(i,j,k)$ represents the location $(x_i+h/2, y_i, z_i)$ for the grid point (x_i, y_i, z_i). Eq. (A.1) is the basis for formulating the LPBE as a linear system of equations, which are represented compactly by Eq. (2.2.2) introduced in the section on fixed charge LPBE (p. 19).

Appendix B Representation of the Delta-Functional Using B-splines

The delta functional δ is defined by

$$\int_{-\infty}^{\infty} \delta(x-a)\,dx = 1 \tag{B.1}$$

and $\delta(x-a)=0$ for $x \neq a$. An approximate discrete 1-dimensional realization of this definition (approximate because the width is not infinitesimally small) is

$$\sum_{i=1}^{5} \mathrm{W}(x_i, a) = 1, \tag{B.2}$$

where the function W has been defined in Eq. (3.1.6) via 5th order B-splines and $\{x_1, ..., x_5\}$ are the 5 closest grid points to a. In the limit of infinitesimal grid spacing, the properties of the Delta functional are met exactly by expressing Eq. (B.2) above as a continuous integral

$$\int_{-\varepsilon}^{\varepsilon} \mathrm{W}(x, a)\,dx = 1, \tag{B.3}$$

where $\varepsilon > 0$.

The value at a of any function known to be defined over the grid can then be determined as

$$\int_{-\varepsilon}^{\varepsilon} \mathrm{W}(x,a) f(x) dx = f(a), \tag{B.4}$$

and the negative of its gradient as

$$\begin{aligned} \int_{-\varepsilon}^{\varepsilon} \left(\nabla \mathrm{W}(x,a)\right) f(x) dx &= \mathrm{W}(x,a) f(x)\Big|_{-\varepsilon}^{\varepsilon} - \int_{-\varepsilon}^{\varepsilon} \mathrm{W}(x,a) \nabla f(x) dx \\ &= -\nabla f(a) \end{aligned} \tag{B.5}$$

Further differentiations can be found in an analogous fashion, limited only by the continuity of the B-spline.

Appendix C Permanent and Polarization PMPB Forces

After solving the linear system, the permanent electrostatic solvation forces are determined via Eq. (3.4.3), which in the limit of infinitesimal grid spacing becomes

$$\mathrm{F}_{i,\gamma} = -\frac{\partial \Delta \mathrm{G}_{\mathrm{M}}}{\partial s_{i,\gamma}} = -\int_V \left[\Phi^{\mathrm{M}} \frac{\partial \rho_{\mathrm{M}}}{\partial s_{i,\gamma}} + \frac{1}{8\pi} \Phi_{\mathrm{s}}^{\mathrm{M}} \nabla \left(\frac{\partial \varepsilon_s}{\partial s_{i,\gamma}} \nabla \Phi_{\mathrm{s}}^{\mathrm{M}} \right) - \frac{1}{8\pi} \left(\Phi_{\mathrm{s}}^{\mathrm{M}} \right)^2 \frac{\partial \bar{\kappa}^2}{\partial s_{i,\gamma}} \right] d^3\mathbf{r} \text{ ,(C.1)}$$

where γ represents differentiation with respect to either the x-, y- or z-coordinate of atom *i*. The three terms on the RHS of Eq. (C.1) are usually referred to as the reaction field (RF) force, dielectric boundary (DB) force and ionic boundary (IB) force, respectively. We briefly review the implementation of these forces, in order to develop the foundation necessary to discuss additional details of realizing the polarization forces.

C.1 Permanent Reaction Field Force and Torque

The γ-component of the "Permanent Reaction Field Force" $\mathbf{F}_i^{\text{Perm RF}}$ for atom *i* is

$$\mathrm{F}_{i,\gamma}^{\text{Perm RF}} = -\frac{\partial}{\partial s_{i,\gamma}} \left[q_i \mathbf{B}_i - d_{i,\alpha} \nabla_\alpha \mathbf{B}_i + \frac{\Theta_{i,\alpha\beta}}{3} \nabla_\alpha \nabla_\beta \mathbf{B}_i \right] \Phi^{\mathrm{M}} \qquad \text{(C.2)}$$

where $\mathbf{B}_i$ is a single column of the B-spline matrix in Eq. (3.1.11), $\mathrm{d}_{i,\alpha}$ is the α-component of the permanent dipole, $\Theta_{i,\alpha\beta}$ is the $\alpha\beta$ component of the quadrupole and the

convention for summation over the α and β subscripts is implied. There is also an associated "Permanent Reaction Field Torque" $\boldsymbol{\tau}_i^{\text{Perm RF}}$, whose x-component is

$$\tau_{i,x}^{\text{Perm q-Phi}} = d_{i,y}\mathrm{E}_{\mathrm{RF},i,z}^{\mathrm{M}} - d_{i,z}\mathrm{E}_{\mathrm{RF},i,y}^{\mathrm{M}} - \frac{2}{3}\left[\Theta_{i,y\alpha}\nabla_\alpha \mathrm{E}_{\mathrm{RF},i,z}^{\mathrm{M}} - \Theta_{i,z\alpha}\nabla_\alpha \mathrm{E}_{\mathrm{RF},i,y}^{\mathrm{M}}\right], \tag{C.3}$$

where $\mathrm{E}_{\mathrm{RF},i,\alpha}^{\mathrm{M}}$ is the α-component of the permanent multipole reaction field. The y- and z-components are analogous, and we note that all torques are equivalent to forces on neighboring atoms that define the local frame of the multipole.

C.2 Direct Polarization Reaction Field Force and Torque

Similarly, the third term of the polarization gradient given in Eq. (3.7.18) results in a "Direct Polarization Reaction Field Force"

$$\mathrm{F}_{i,\gamma}^{\text{Direct RF}} = \frac{1}{2}\frac{\partial}{\partial s_{i,\gamma}}\left[\left(-\mu_{i,\alpha}\nabla_\alpha \mathbf{B}_i\right)\boldsymbol{\Phi}^{\mathrm{M}} + \left(q_i\mathbf{B}_i - d_{i,\alpha}\nabla_\alpha \mathbf{B}_i + \frac{\Theta_{i,\alpha\beta}}{3}\nabla_\alpha\nabla_\beta \mathbf{B}_i\right)\boldsymbol{\Phi}^{\mu}\right], \tag{C.4}$$

while the fifth term we label the "Non-Local Direct Polarization Reaction Field Force"

$$\mathrm{F}_{i,\gamma}^{\text{NL-Direct RF}} = \frac{1}{2}\frac{\partial}{\partial s_{i,\gamma}}\left[\left(-\nu_{i,\alpha}\nabla_\alpha \mathbf{B}_i\right)\boldsymbol{\Phi}^{\mathrm{M}} + \left(q_i\mathbf{B}_i - d_{i,\alpha}\nabla_\alpha \mathbf{B}_i + \frac{\Theta_{i,\alpha\beta}}{3}\nabla_\alpha\nabla_\beta \mathbf{B}_i\right)\boldsymbol{\Phi}^{\nu}\right], \tag{C.5}$$

respectively. The label "non-local" is used to denote that the term $\boldsymbol{\nu}$ results from omitting or scaling the contribution to the intramolecular field of permanent multipoles that are in a 1-5 connected or closer, as opposed to the induced dipoles $\boldsymbol{\mu}$ that result from the AMOEBA group based polarization scheme. Additionally, the x-component of the

torques, $\boldsymbol{\tau}_i^{\text{Direct RF}}$ and $\boldsymbol{\tau}_i^{\text{NL-Direct RF}}$, on the permanent moments due to the continuum reaction field of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are, respectively,

$$\tau_{i,x}^{\text{Direct RF}} = \frac{1}{2}\left[d_{i,y}\mathrm{E}_{\text{RF},i,z}^{\mu} - d_{\sigma,z}\mathrm{E}_{\text{RF},i,y}^{\mu} - \frac{2}{3}\left(\Theta_{i,y\alpha}\nabla_{\alpha}\mathrm{E}_{\text{RF},i,z}^{\mu} - \Theta_{i,z\alpha}\nabla_{\alpha}\mathrm{E}_{\text{RF},i,y}^{\mu} \right) \right] \quad \text{(C.6)}$$

and

$$\tau_{i,x}^{\text{NL-Direct RF}} = \frac{1}{2}\left[d_{i,y}\mathrm{E}_{\text{RF},i,z}^{\nu} - d_{\sigma,z}\mathrm{E}_{\text{RF},i,y}^{\nu} - \frac{2}{3}\left(\Theta_{i,y\alpha}\nabla_{\alpha}\mathrm{E}_{\text{RF},i,z}^{\nu} - \Theta_{i,z\alpha}\nabla_{\alpha}\mathrm{E}_{\text{RF},i,y}^{\nu} \right) \right]. \quad \text{(C.7)}$$

C.3 Mutual Polarization Reaction Field Force

The last reaction field force results from the seventh term of Eq. (3.7.18) and is due to mutual polarization

$$\mathrm{F}_{i,\gamma}^{\text{Mutual RF}} = -\frac{1}{2}\frac{\partial}{\partial s_{i,\gamma}}\left(\mu_{i,\alpha}\nabla_{\alpha}\mathbf{B}_i\boldsymbol{\Phi}^{\nu} + \nu_{i,\alpha}\nabla_{\alpha}\mathbf{B}_i\boldsymbol{\Phi}^{\mu} \right) \quad \text{(C.8)}$$

C.4 Permanent Dielectric Boundary Force

The second term in Eq. (C.1), the “Permanent Dielectric Boundary Force”, is determined from Eq. (A.1) as

$$
\begin{aligned}
\mathrm{F}_{i,\gamma}^{\text{Perm DB}} = -\frac{h}{8}\sum_{i,j,k}\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right) \\
&\left\{\frac{\partial\varepsilon_x\left(i,j,k\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i+1,j,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right]\right. \\
&+\frac{\partial\varepsilon_x\left(i-1,j,k\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i-1,j,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_y\left(i,j,k\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j+1,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_y\left(i,j-1,k\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j-1,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_z\left(i,j,k\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k+1\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
&\left.+\frac{\partial\varepsilon_z\left(i,j,k-1\right)}{\partial r_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k-1\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right]\right\}
\end{aligned}
\tag{C.9}
$$

where the partial derivatives of the permittivity depend on Eqs. (3.2.1), (3.2.3) and the heptic characteristic function presented in Eqs. (3.2.1) through (3.2.5).

$$
\begin{aligned}
\frac{\partial\varepsilon_x\left(i,j,k\right)}{\partial r_{i,\gamma}} &= \left(1-\varepsilon_s\right)\frac{\mathrm{H}_x\left(i,j,k\right)}{\mathrm{H}_{xi}\left(i,j,k\right)}\frac{\partial\mathrm{H}_x\left(i,j,k\right)}{\partial r_{i,\gamma}} \\
&= \left[\varepsilon_x\left(i,j,k\right)-1\right]\frac{\mathrm{H}'_{xi}\left(i,j,k\right)}{\mathrm{H}_{xi}\left(i,j,k\right)}\frac{-r_{i,\gamma}}{r_i}
\end{aligned}
,
\tag{C.10}
$$

where H_x and H_{xi} are the characteristic function of the solute and atom i for the x-branch of the cubic grid at $\left(x_i+h/2,\, y_j,\, z_k\right)$, respectively, and the vector $\mathbf{r}_i$ is the distance from the atomic center to the grid point.

C.5 Direct and Mutual Polarization Dielectric Boundary Forces

The fourth, sixth and eighth terms in Eq. (3.7.18) result in dielectric boundary force components. For example, the "Direct Polarization Dielectric Boundary Force" is

$$
\begin{aligned}
\mathrm{F}_{i,\gamma}^{\text{Direct DB}} = -\frac{h}{8}\sum_{i,j,k}\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right) & \\
\Bigg\{ & \frac{\partial\varepsilon_x\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i+1,j,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
& +\frac{\partial\varepsilon_x\left(i-1,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i-1,j,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
& +\frac{\partial\varepsilon_y\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j+1,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
& +\frac{\partial\varepsilon_y\left(i,j-1,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j-1,k\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
& +\frac{\partial\varepsilon_z\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k+1\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right] \\
& +\frac{\partial\varepsilon_z\left(i,j,k-1\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k-1\right)-\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)\right]\Bigg\}
\end{aligned}
\tag{C.11}
$$

or

$$
\begin{aligned}
\mathrm{F}_{i,\gamma}^{\text{Direct DB}} = -\frac{h}{8}\sum_{i,j,k}\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right) \\
&\left\{\frac{\partial\varepsilon_x\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i+1,j,k\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right]\right. \\
&+\frac{\partial\varepsilon_x\left(i-1,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i-1,j,k\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_y\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i,j+1,k\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_y\left(i,j-1,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i,j-1,k\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right] \\
&+\frac{\partial\varepsilon_z\left(i,j,k\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i,j,k+1\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right] \\
&\left.+\frac{\partial\varepsilon_z\left(i,j,k-1\right)}{\partial \mathrm{r}_{i,\gamma}}\left[\Phi_{\mathrm{s}}^{\mu}\left(i,j,k-1\right)-\Phi_{\mathrm{s}}^{\mu}\left(i,j,k\right)\right]\right\}
\end{aligned}
\tag{C.12}
$$

where the superscript on the solvated potentials have been exchanged between Eq. (C.11) and Eq. (C.12). In other words, both Eqs. (C.11) and (C.12) are equivalent to numerical precision and either may be implemented. Analogous expressions for the sixth and eighth terms in Eq. (3.7.18) are referred to as the “Non-Local Direct Polarization Dielectric Boundary Force” and the “Mutual Polarization Dielectric Boundary Force”, respectively.

C.6 Permanent Ionic Boundary Force

The last term in Eq. (C.1), the “Permanent Ionic Boundary Force”, is determined from Eq. (A1) as

$$
\mathrm{F}_{i,\gamma}^{\text{Perm IB}} = \frac{h^3}{8\pi}\sum_{i,j,k}\Phi_{\mathrm{s}}^{\mathrm{M}}\left(i,j,k\right)^2\frac{\partial\overline{\kappa}^2\left(i,j,k\right)}{\partial r_{i,\gamma}} \tag{C.13}
$$

using

$$\begin{aligned}\frac{\partial \bar{\kappa}^2(i,j,k)}{\partial r_{i,\gamma}} &= \bar{\kappa}_b^2 \frac{\mathrm{H}(i,j,k)}{\mathrm{H}_i(i,j,k)} \frac{\partial \mathrm{H}_i(i,j,k)}{\partial r_{i,\gamma}} \\ &= \bar{\kappa}^2(i,j,k) \frac{\mathrm{H}_i'(i,j,k)}{\mathrm{H}_i(i,j,k)} \frac{-r_{i,\gamma}}{r_i}\end{aligned} \tag{C.14}$$

where H and H_i are the characteristic function of the solute and atom *i*, respectively, and the vector $\mathbf{r}_i$ is the distance from the atomic center to the grid point.

C.7 Direct and Mutual Polarization Ionic Boundary Force

The fourth, sixth and eighth terms in Eq. (3.7.18) result in ionic boundary force components. For example, the "Direct Polarization Ionic Boundary Force" is

$$\mathrm{F}_{i,\gamma}^{\text{Direct IB}} = \frac{h^3}{8\pi} \sum_{i,j,k} \boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{M}}(i,j,k) \boldsymbol{\Phi}_{\mathrm{s}}^{\mu}(i,j,k) \frac{\partial \bar{\kappa}^2(i,j,k)}{\partial r_{i,\gamma}} \tag{C.15}$$

Analogous expressions for the sixth and eighth terms in Eq. (3.7.18) are termed the "Non-Local Direct Polarization Ionic Boundary Force" and the "Mutual Polarization Ionic Boundary Force", respectively. Note the difference between Eqs. (C.13) and (C.15); specifically the potential is squared in Eq. (C.13), but is asymmetric in Eq. (C.15)

Appendix D Gradients of the Generalized Kirkwood Tensors

Table 7.1. Gradients of $A^{(0)}_{\{0,0,0\}}$.

m	$\{m\}$	$A^{(0)}_{\{0,0,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$t_{(0,0)}$
1	1,0,0	$xt_{(0,1)}$
	0,1,0	$yt_{(0,1)}$
	0,0,1	$zt_{(0,1)}$
2	2,0,0	$t_{(0,1)} + x^2 t_{(0,2)}$
	1,1,0	$xyt_{(0,2)}$
	1,0,1	$xzt_{(0,2)}$
	0,2,0	$t_{(0,1)} + y^2 t_{(0,2)}$
	0,1,1	$yzt_{(0,2)}$
	0,0,2	$t_{(0,1)} + z^2 t_{(0,2)}$
3	3,0,0	$3xt_{(0,2)} + x^3 t_{(0,3)}$
	2,1,0	$yt_{(0,2)} + x^2 yt_{(0,3)}$
	2,0,1	$zt_{(0,2)} + x^2 zt_{(0,3)}$
	1,2,0	$xt_{(0,2)} + xy^2 t_{(0,3)}$
	1,1,1	$xyzt_{(0,3)}$
	1,0,2	$xt_{(0,2)} + xz^2 t_{(0,3)}$
	0,3,0	$3yt_{(0,2)} + y^3 t_{(0,3)}$
	0,2,1	$zt_{(0,2)} + y^2 zt_{(0,3)}$
	0,1,2	$yt_{(0,2)} + yz^2 t_{(0,3)}$
	0,0,3	$3zt_{(0,2)} + z^3 t_{(0,3)}$

Table 7.2. Gradients of $A^{(1)}_{1,0,0}$.

m	$\{m\}$	$A^{(1)}_{\{1,0,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$xt_{(1,0)}$
1	1,0,0	$t_{(1,0)} + x^2 t_{(1,1)}$
	0,1,0	$xyt_{(1,1)}$
	0,0,1	$xzt_{(1,1)}$
2	2,0,0	$3xt_{(1,1)} + x^3 t_{(1,2)}$
	1,1,0	$yt_{(1,1)} + x^2 yt_{(1,2)}$
	1,0,1	$zt_{(1,1)} + x^2 zt_{(1,2)}$
	0,2,0	$xt_{(1,1)} + xy^2 t_{(1,2)}$
	0,1,1	$xyzt_{(1,2)}$
	0,0,2	$xt_{(1,1)} + xz^2 t_{(1,2)}$
3	3,0,0	$3t_{(1,1)} + 6x^2 t_{(1,2)} + x^4 t_{(1,3)}$
	2,1,0	$3xyt_{(1,2)} + x^3 yt_{(1,3)}$
	2,0,1	$3xzt_{(1,2)} + x^3 zt_{(1,3)}$
	1,2,0	$t_{(1,1)} + x^2 t_{(1,2)} + y^2 t_{(1,2)} + x^2 y^2 t_{(1,3)}$
	1,1,1	$yzt_{(1,2)} + x^2 yzt_{(1,3)}$
	1,0,2	$t_{(1,1)} + x^2 t_{(1,2)} + z^2 t_{(1,2)} + x^2 z^2 t_{(1,3)}$
	0,3,0	$3xyt_{(1,2)} + xy^3 t_{(1,3)}$
	0,2,1	$xzt_{(1,2)} + xy^2 zt_{(1,3)}$
	0,1,2	$xyt_{(1,2)} + xyz^2 t_{(1,3)}$
	0,0,3	$3xzt_{(1,2)} + xz^3 t_{(1,3)}$

Table 7.3. Gradients of $A^{(1)}_{\{0,1,0\}}$.

m	$\{m\}$	$A^{(1)}_{\{0,1,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$yt_{(1,0)}$
1	1,0,0	$xyt_{(1,1)}$
	0,1,0	$t_{(1,0)} + y^2 t_{(1,1)}$
	0,0,1	$yzt_{(1,1)}$
2	2,0,0	$yt_{(1,1)} + x^2 yt_{(1,2)}$
	1,1,0	$xt_{(1,1)} + xy^2 t_{(1,2)}$
	1,0,1	$xyzt_{(1,2)}$
	0,2,0	$3yt_{(1,1)} + y^3 t_{(1,2)}$
	0,1,1	$zt_{(1,1)} + y^2 zt_{(1,2)}$
	0,0,2	$yt_{(1,1)} + yz^2 t_{(1,2)}$
3	3,0,0	$3xyt_{(1,2)} + x^3 yt_{(1,3)}$
	2,1,0	$t_{(1,1)} + x^2 t_{(1,2)} + y^2 t_{(1,2)} + x^2 y^2 t_{(1,3)}$
	2,0,1	$yzt_{(1,2)} + x^2 yzt_{(1,3)}$
	1,2,0	$3xyt_{(1,2)} + xy^3 t_{(1,3)}$
	1,1,1	$xzt_{(1,2)} + xy^2 zt_{(1,3)}$
	1,0,2	$xyt_{(1,2)} + xyz^2 t_{(1,3)}$
	0,3,0	$3t_{(1,1)} + 6y^2 t_{(1,2)} + y^4 t_{(1,3)}$
	0,2,1	$3yzt_{(1,2)} + y^3 zt_{(1,3)}$
	0,1,2	$t_{(1,1)} + y^2 t_{(1,2)} + z^2 t_{(1,2)} + y^2 z^2 t_{(1,3)}$
	0,0,3	$3yzt_{(1,2)} + yz^3 t_{(1,3)}$

Table 7.4. Gradients of $A^{(1)}_{\{0,0,1\}}$.

m	$\{m\}$	$A^{(1)}_{\{0,0,1\},\{m_1,m_2,m_3\}}$
0	0,0,0	$zt_{(1,0)}$
1	1,0,0	$xzt_{(1,1)}$
	0,1,0	$yzt_{(1,1)}$
	0,0,1	$t_{(1,0)} + z^2 t_{(1,1)}$
2	2,0,0	$zt_{(1,1)} + x^2 zt_{(1,2)}$
	1,1,0	$xyzt_{(1,2)}$
	1,0,1	$xt_{(1,1)} + xz^2 t_{(1,2)}$
	0,2,0	$zt_{(1,1)} + y^2 zt_{(1,2)}$
	0,1,1	$yt_{(1,1)} + yz^2 t_{(1,2)}$
	0,0,2	$3zt_{(1,1)} + z^3 t_{(1,2)}$
3	3,0,0	$3xzt_{(1,2)} + x^3 zt_{(1,3)}$
	2,1,0	$yzt_{(1,2)} + x^2 yzt_{(1,3)}$
	2,0,1	$t_{(1,1)} + x^2 t_{(1,2)} + z^2 t_{(1,2)} + x^2 z^2 t_{(1,3)}$
	1,2,0	$xzt_{(1,2)} + xy^2 zt_{(1,3)}$
	1,1,1	$xyt_{(1,2)} + xyz^2 t_{(1,3)}$
	1,0,2	$3xzt_{(1,2)} + xz^3 t_{(1,3)}$
	0,3,0	$3yzt_{(1,2)} + y^3 zt_{(1,3)}$
	0,2,1	$t_{(1,1)} + y^2 t_{(1,2)} + z^2 t_{(1,2)} + y^2 z^2 t_{(1,3)}$
	0,1,2	$3yzt_{(1,2)} + yz^3 t_{(1,3)}$
	0,0,3	$3t_{(1,1)} + 6z^2 t_{(1,2)} + z^4 t_{(1,3)}$

Table 7.5. Gradients of $A^{(2)}_{\{2,0,0\}}$.

m	$\{m\}$	$A^{(2)}_{\{2,0,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$x^2 t_{(2,0)}$
1	1,0,0	$2xt_{(2,0)} + x^3 t_{(2,1)}$
	0,1,0	$x^2 yt_{(2,1)}$
	0,0,1	$x^2 zt_{(2,1)}$
2	2,0,0	$2t_{(2,0)} + 5x^2 t_{(2,1)} + x^4 t_{(2,2)}$
	1,1,0	$2xyt_{(2,1)} + x^3 yt_{(2,2)}$
	1,0,1	$2xzt_{(2,1)} + x^3 zt_{(2,2)}$
	0,2,0	$x^2 t_{(2,1)} + x^2 y^2 t_{(2,2)}$
	0,1,1	$x^2 yzt_{(2,2)}$
	0,0,2	$x^2 t_{(2,1)} + x^2 z^2 t_{(2,2)}$
3	3,0,0	$12xt_{(2,1)} + 9x^3 t_{(2,2)} + x^5 t_{(2,3)}$
	2,1,0	$2yt_{(2,1)} + 5x^2 yt_{(2,2)} + x^4 yt_{(2,3)}$
	2,0,1	$2zt_{(2,1)} + 5x^2 zt_{(2,2)} + x^4 zt_{(2,3)}$
	1,2,0	$2xt_{(2,1)} + x^3 t_{(2,2)} + 2xy^2 t_{(2,2)} + x^3 y^2 t_{(2,3)}$
	1,1,1	$2xyzt_{(2,2)} + x^3 yzt_{(2,3)}$
	1,0,2	$2xt_{(2,1)} + x^3 t_{(2,2)} + 2xz^2 t_{(2,2)} + x^3 z^2 t_{(2,3)}$
	0,3,0	$3x^2 yt_{(2,2)} + x^2 y^3 t_{(2,3)}$
	0,2,1	$x^2 zt_{(2,2)} + x^2 y^2 zt_{(2,3)}$
	0,1,2	$yx^2 t_{(2,2)} + x^2 yz^2 t_{(2,3)}$
	0,0,3	$3x^2 zt_{(2,2)} + x^2 z^3 t_{(2,3)}$

Table 7.6. Gradients of $A^{(2)}_{\{1,1,0\}}$.

m	$\{m\}$	$A^{(2)}_{\{1,1,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$xyt_{(2,0)}$
1	1,0,0	$yt_{(2,0)} + x^2 yt_{(2,1)}$
	0,1,0	$xt_{(2,0)} + xy^2 t_{(2,1)}$
	0,0,1	$xyzt_{(2,1)}$
2	2,0,0	$3xyt_{(2,1)} + x^3 yt_{(2,2)}$
	1,1,0	$t_{(2,0)} + x^2 t_{(2,1)} + y^2 t_{(2,1)} + x^2 y^2 t_{(2,2)}$
	1,0,1	$yzt_{(2,1)} + x^2 yzt_{(2,2)}$
	0,2,0	$3xyt_{(2,1)} + xy^3 t_{(2,2)}$
	0,1,1	$xzt_{(2,1)} + xy^2 zt_{(2,2)}$
	0,0,2	$xyt_{(2,1)} + xyz^2 t_{(2,2)}$
3	3,0,0	$3yt_{(2,1)} + 6x^2 yt_{(2,2)} + x^4 yt_{(2,3)}$
	2,1,0	$3xt_{(2,1)} + 3xy^2 t_{(2,2)} + x^3 t_{(2,2)} + x^3 y^2 t_{(2,3)}$
	2,0,1	$3xyzt_{(2,2)} + x^3 yzt_{(2,3)}$
	1,2,0	$3yt_{(2,1)} + 3x^2 yt_{(2,2)} + y^3 t_{(2,2)} + x^2 y^3 t_{(2,3)}$
	1,1,1	$zt_{(2,1)} + x^2 zt_{(2,2)} + y^2 zt_{(2,2)} + x^2 y^2 zt_{(2,3)}$
	1,0,2	$yt_{(2,1)} + x^2 yt_{(2,2)} + yz^2 t_{(2,2)} + x^2 yz^2 t_{(2,3)}$
	0,3,0	$3xt_{(2,1)} + 6xy^2 t_{(2,2)} + xy^4 t_{(2,3)}$
	0,2,1	$3xyzt_{(2,2)} + xy^3 zt_{(2,3)}$
	0,1,2	$xt_{(2,1)} + xy^2 t_{(2,2)} + xz^2 t_{(2,2)} + xy^2 z^2 t_{(2,3)}$
	0,0,3	$3xyzt_{(2,2)} + xyz^3 t_{(2,3)}$

Table 7.7. Gradients of $A^{(2)}_{\{1,0,1\}}$.

m	$\{m\}$	$A^{(2)}_{\{1,0,1\},\{m_1,m_2,m_3\}}$
0	0,0,0	$xzt_{(2,0)}$
1	1,0,0	$zt_{(2,0)} + x^2zt_{(2,1)}$
	0,1,0	$xyzt_{(2,1)}$
	0,0,1	$xt_{(2,0)} + xz^2t_{(2,1)}$
2	2,0,0	$3xzt_{(2,1)} + x^3zt_{(2,2)}$
	1,1,0	$yzt_{(2,1)} + x^2yzt_{(2,2)}$
	1,0,1	$t_{(2,0)} + x^2t_{(2,1)} + z^2t_{(2,1)} + x^2z^2t_{(2,2)}$
	0,2,0	$xzt_{(2,1)} + xy^2zt_{(2,2)}$
	0,1,1	$xyt_{(2,1)} + xyz^2t_{(2,2)}$
	0,0,2	$3xzt_{(2,1)} + xz^3t_{(2,2)}$
3	3,0,0	$3zt_{(2,1)} + 6x^2zt_{(2,2)} + x^4zt_{(2,3)}$
	2,1,0	$3xyzt_{(2,2)} + x^3yzt_{(2,3)}$
	2,0,1	$3xt_{(2,1)} + 3xz^2t_{(2,2)} + x^3t_{(2,2)} + x^3z^2t_{(2,3)}$
	1,2,0	$zt_{(2,1)} + x^2zt_{(2,2)} + y^2zt_{(2,2)} + x^2y^2zt_{(2,3)}$
	1,1,1	$yt_{(2,1)} + x^2yt_{(2,2)} + yz^2t_{(2,2)} + x^2yz^2t_{(2,3)}$
	1,0,2	$3zt_{(2,1)} + 3x^2zt_{(2,2)} + z^3t_{(2,2)} + x^2z^3t_{(2,3)}$
	0,3,0	$3xyzt_{(2,2)} + xy^3zt_{(2,3)}$
	0,2,1	$xt_{(2,1)} + xy^2t_{(2,2)} + xz^2t_{(2,2)} + xy^2z^2t_{(2,3)}$
	0,1,2	$3xyzt_{(2,2)} + xyz^3t_{(2,3)}$
	0,0,3	$3xt_{(2,1)} + 6xz^2t_{(2,2)} + xz^4t_{(2,3)}$

Table 7.8. Gradients of $A^{(2)}_{\{0,2,0\}}$.

m	$\{m\}$	$A^{(2)}_{\{0,2,0\},\{m_1,m_2,m_3\}}$
0	0,0,0	$y^2 t_{(2,0)}$
1	1,0,0	$xy^2 t_{(2,1)}$
	0,1,0	$2yt_{(2,0)} + y^3 t_{(2,1)}$
	0,0,1	$y^2 zt_{(2,1)}$
2	2,0,0	$y^2 t_{(2,1)} + x^2 y^2 t_{(2,2)}$
	1,1,0	$2xyt_{(2,1)} + xy^3 t_{(2,2)}$
	1,0,1	$xy^2 zt_{(2,2)}$
	0,2,0	$2t_{(2,0)} + 5y^2 t_{(2,1)} + y^4 t_{(2,2)}$
	0,1,1	$2yzt_{(2,1)} + y^3 zt_{(2,2)}$
	0,0,2	$y^2 t_{(2,1)} + y^2 z^2 t_{(2,2)}$
3	3,0,0	$3xy^2 t_{(2,2)} + x^3 y^2 t_{(2,3)}$
	2,1,0	$2yt_{(2,1)} + y^3 t_{(2,2)} + 2x^2 yt_{(2,2)} + x^2 y^3 t_{(2,3)}$
	2,0,1	$y^2 zt_{(2,2)} + x^2 y^2 zt_{(2,3)}$
	1,2,0	$2xt_{(2,1)} + 5xy^2 t_{(2,2)} + xy^4 t_{(2,3)}$
	1,1,1	$2xyzt_{(2,2)} + xy^3 zt_{(2,3)}$
	1,0,2	$xy^2 t_{(2,2)} + xy^2 z^2 t_{(2,3)}$
	0,3,0	$12yt_{(2,1)} + 9y^3 t_{(2,2)} + y^5 t_{(2,3)}$
	0,2,1	$2zt_{(2,1)} + 5y^2 zt_{(2,2)} + y^4 zt_{(2,3)}$
	0,1,2	$2yt_{(2,1)} + y^3 t_{(2,2)} + 2yz^2 t_{(2,2)} + y^3 z^2 t_{(2,3)}$
	0,0,3	$3y^2 zt_{(2,2)} + y^2 z^3 t_{(2,3)}$

Table 7.9. Gradients of $A^{(2)}_{\{0,1,1\}}$.

m	$\{m\}$	$A^{(2)}_{\{0,1,1\},\{m_1,m_2,m_3\}}$
0	0,0,0	$yzt_{(2,0)}$
1	1,0,0	$xyzt_{(2,1)}$
	0,1,0	$zt_{(2,0)} + y^2zt_{(2,1)}$
	0,0,1	$yt_{(2,0)} + yz^2t_{(2,1)}$
2	2,0,0	$yzt_{(2,1)} + x^2yzt_{(2,2)}$
	1,1,0	$xzt_{(2,1)} + xy^2zt_{(2,2)}$
	1,0,1	$xyt_{(2,1)} + xyz^2t_{(2,2)}$
	0,2,0	$3yzt_{(2,1)} + y^3zt_{(2,2)}$
	0,1,1	$t_{(2,0)} + z^2t_{(2,1)} + y^2t_{(2,1)} + y^2z^2t_{(2,2)}$
	0,0,2	$3yzt_{(2,1)} + yz^3t_{(2,2)}$
3	3,0,0	$3xyzt_{(2,2)} + x^3yzt_{(2,3)}$
	2,1,0	$zt_{(2,1)} + y^2zt_{(2,2)} + x^2zt_{(2,2)} + x^2y^2zt_{(2,3)}$
	2,0,1	$yt_{(2,1)} + x^2yt_{(2,2)} + yz^2t_{(2,2)} + x^2yz^2t_{(2,3)}$
	1,2,0	$3xyzt_{(2,2)} + xy^3zt_{(2,3)}$
	1,1,1	$xt_{(2,1)} + xy^2t_{(2,2)} + xz^2t_{(2,2)} + xy^2z^2t_{(2,3)}$
	1,0,2	$3xyzt_{(2,2)} + xyz^3t_{(2,3)}$
	0,3,0	$3zt_{(2,1)} + 6y^2zt_{(2,2)} + y^4zt_{(2,3)}$
	0,2,1	$3yt_{(2,1)} + 3yz^2t_{(2,2)} + y^3t_{(2,2)} + y^3z^2t_{(2,3)}$
	0,1,2	$3zt_{(2,1)} + 3y^2zt_{(2,2)} + z^3t_{(2,2)} + y^2z^3t_{(2,3)}$
	0,0,3	$3yt_{(2,1)} + 6yz^2t_{(2,2)} + yz^4t_{(2,3)}$

Table 7.10. Gradients of $A^{(2)}_{\{0,0,2\}}$.

m	$\{m\}$	$A^{(2)}_{\{0,0,2\},\{m_1,m_2,m_3\}}$
0	0,0,0	$z^2 t_{(2,0)}$
1	1,0,0	$xz^2 t_{(2,1)}$
	0,1,0	$yz^2 t_{(2,1)}$
	0,0,1	$2zt_{(2,0)} + z^3 t_{(2,1)}$
2	2,0,0	$z^2 t_{(2,1)} + x^2 z^2 t_{(2,2)}$
	1,1,0	$xyz^2 t_{(2,2)}$
	1,0,1	$2xzt_{(2,1)} + xz^3 t_{(2,2)}$
	0,2,0	$z^2 t_{(2,1)} + y^2 z^2 t_{(2,2)}$
	0,1,1	$2yzt_{(2,1)} + yz^3 t_{(2,2)}$
	0,0,2	$2t_{(2,0)} + 5z^2 t_{(2,1)} + z^4 t_{(2,2)}$
3	3,0,0	$3xz^2 t_{(2,2)} + x^3 z^2 t_{(2,3)}$
	2,1,0	$yz^2 t_{(2,2)} + x^2 yz^2 t_{(2,3)}$
	2,0,1	$2zt_{(2,1)} + z^3 t_{(2,2)} + 2x^2 zt_{(2,2)} + x^2 z^3 t_{(2,3)}$
	1,2,0	$xz^2 t_{(2,2)} + xy^2 z^2 t_{(2,3)}$
	1,1,1	$2xyzt_{(2,2)} + xyz^3 t_{(2,3)}$
	1,0,2	$2xt_{(2,1)} + 5xz^2 t_{(2,2)} + xz^4 t_{(2,3)}$
	0,3,0	$3yz^2 t_{(2,2)} + y^3 z^2 t_{(2,3)}$
	0,2,1	$2zt_{(2,1)} + z^3 t_{(2,2)} + 2y^2 zt_{(2,2)} + y^2 z^3 t_{(2,3)}$
	0,1,2	$2yt_{(2,1)} + 5yz^2 t_{(2,2)} + yz^4 t_{(2,3)}$
	0,0,3	$12zt_{(2,1)} + 9z^3 t_{(2,2)} + z^5 t_{(2,3)}$

References

1. Flory, P. J., *Statistical Mechanics Of Chain Molecules*. Butterworth-Heinemann Ltd: 1969.

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

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

4. Roux, B.; Simonson, T., Implicit solvent models. *Biophysical Chemistry* **1999,** 78, (1-2), 1-20.

5. Onsager, L., Electric moments of molecules in liquids. *Journal of the American Chemical Society* **1936,** 58, (8), 1486-1493.

6. Baker, N. A., Poisson-Boltzmann methods for biomolecular electrostatics. *Methods in Enzymology* **2004,** 383, 94-118.

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

8. Ponder, J. W.; Case, D. A., Force fields for protein simulations. In *Advances in Protein Chemistry*, Academic Press: 2003; Vol. Volume 66, pp 27-85.

9. Ren, P. Y.; Ponder, J. W., Polarizable Atomic Multipole Based AMOEBA Potential for Protein Modeling (submitted). *Journal of Physical Chemistry B* **2007**.

10. Ren, P. Y.; Ponder, J. W., Polarizable Atomic Multipole Based Intermolecular Potential for Small Organic Molecules (submitted). *Journal of Physical Chemistry B* **2007**.

11. Warwicker, J.; Watson, H. C., Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *Journal of Molecular Biology* **1982,** 157, (4), 671-9.

12. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **2001,** 98, (18), 10037-41.

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

14. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** 268, (5214), 1144-1149.

15. Baker, N. A., Improving implicit solvent simulations: a Poisson-centric view. *Current Opinion in Structural Biology* **2005,** 15, (2), 137-143.

16. Maple, J. R.; Cao, Y. X.; Damm, W. G.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A., A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation* **2005,** 1, (4), 694-715.

17. Cortis, C. M.; Langlois, J. M.; Beachy, M. D.; Friesner, R. A., Quantum mechanical geometry optimization in solution using a finite element continuum electrostatics method. *Journal of Chemical Physics* **1996,** 105, (13), 5472-5484.

18. Friedrichs, M.; Zhou, R. H.; Edinger, S. R.; Friesner, R. A., Poisson-Boltzmann analytical gradients for molecular modeling calculations. *Journal of Physical Chemistry B* **1999,** 103, (16), 3057-3061.

19. Im, W.; Beglov, D.; Roux, B., Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications* **1998,** 111, (1-3), 59-75.

20. Thole, B. T., Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981,** 59, (3), 341-350.

21. Davis, M. E.; Mccammon, J. A., Calculating electrostatic forces from grid-calculated potentials. *Journal of Computational Chemistry* **1990,** 11, (3), 401-409.

22. Gilson, M. K.; Davis, M. E.; Luty, B. A.; Mccammon, J. A., Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *Journal of Physical Chemistry* **1993,** 97, (14), 3591-3600.

23. Niedermeier, C.; Schulten, K., Molecular-dynamics simulations in heterogeneous dielectrica and Debye-Huckel media - application to the protein bovine pancreatic trypsin inhibitor. *Molecular Simulation* **1992,** 8, (6), 361-387.

24. Gilson, M. K., Molecular dynamics simulation with a continuum electrostatic model of the solvent. *Journal of Computational Chemistry* **1995,** 16, (9), 1081-1095.

25. Micu, A. M.; Bagheri, B.; Ilin, A. V.; Scott, L. R.; Pettitt, B. M., Numerical considerations in the computation of the electrostatic free energy of interaction within the Poisson-Boltzmann theory. *Journal of Computational Physics* **1997,** 136, (2), 263-271.

26. Born, M., Volumen und Hydratationsarme der Ionen. *Z Phys* **1920,** 1, (1), 45-48.

27. Kirkwood, J. G., Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *Journal of Chemical Physics* **1934,** 2, (7), 351-361.

28. Kong, Y.; Ponder, J. W., Calculation of the reaction field due to off-center point multipoles. *Journal of Chemical Physics* **1997,** 107, (2), 481-492.

29. Schaefer, M.; Karplus, M., A comprehensive analytical treatment of continuum electrostatics. *Journal of Physical Chemistry* **1996,** 100, (5), 1578-1599.

30. Schaefer, M.; Bartels, C.; Karplus, M., Solution conformations and thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *Journal of Molecular Biology* **1998,** 284, (3), 835-848.

31. Schaefer, M.; Froemmel, C., A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *Journal of Molecular Biology* **1990,** 216, (4), 1045-1066.

32. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters* **1995,** 246, (1-2), 122-129.

33. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *Journal of Physical Chemistry* **1996,** 100, (51), 19824-19839.

34. Jeancharles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C., Electrostatic contributions to solvation energies - comparison of free-energy perturbation and continuum calculations. *Journal of the American Chemical Society* **1991,** 113, (4), 1454-1455.

35. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society* **1990,** 112, (16), 6127-6129.

36. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997,** 101, (16), 3005-3014.

37. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., Performance comparison of Generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry* **2004,** 25, (2), 265-284.

38. Feig, M.; Brooks, C. L., 3rd, Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology* **2004,** 14, (2), 217-24.

39. Feig, M.; Im, W.; Brooks, C. L., Implicit solvation based on generalized Born theory in different dielectric environments. *Journal of Chemical Physics* **2004,** 120, (2), 903-911.

40. Tanizaki, S.; Feig, M., A Generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes. *Journal of Chemical Physics* **2005,** 122, (12).

41. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the Generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry* **2002,** 23, (14), 1297-1304.

42. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the Generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B* **2000,** 104, (15), 3712-3720.

43. Sigalov, G.; Fenley, A.; Onufriev, A., Analytical electrostatics for biomolecules: Beyond the Generalized Born approximation. *Journal of Chemical Physics* **2006,** 124, (12).

44. Sigalov, G.; Scheffel, P.; Onufriev, A., Incorporating variable dielectric environments into the Generalized Born model. *Journal of Chemical Physics* **2005,** 122, (9).

45. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified Generalized Born model. *Proteins-Structure Function and Bioinformatics* **2004,** 55, (2), 383-394.

46. Schnieders, M. J.; Ponder, J. W., Polarizable atomic multipole solutes in a Generalized Kirkwood continuum (to appear). *Journal of Chemical Theory and Computation* **2007**.

47. Gallicchio, E.; Zhang, L. Y.; Levy, R. M., The SGB/NP hydration free energy model based on the surface Generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. *Journal of Computational Chemistry* **2002,** 23, (5), 517-529.

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

49. Friesner, R. A.; Robert L. Baldwin, a. D. B., Modeling polarization in proteins and protein-ligand complexes: methods and preliminary results. In *Advances in Protein Chemistry*, Academic Press: 2005; Vol. Volume 72, pp 79-104.

50. Jaramillo, A.; Wodak, S. J., Computational protein design is a challenge for implicit solvation models. *Biophysical Journal* **2005,** 88, (1), 156-71.

51. Vizcarra, C. L.; Mayo, S. L., Electrostatics in computational protein design. *Current Opinion in Chemical Biology* **2005,** 9, (6), 622-6.

52. Cisneros, G. A.; Piquemal, J. P.; Darden, T. A., Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *Journal of Chemical Physics* **2006,** 125, (18).

53. Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A., Towards a force field based on density fitting. *Journal of Chemical Physics* **2006,** 124, (10).

54. Bashford, D.; Case, D. A., Generalized Born models of macromolecular solvation effects. *Annual Review of Physical Chemistry* **2000,** 51, 129-152.

55. Cammi, R.; Tomasi, J., Remarks on the use of the apparent surface charges (ASC) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *Journal of Computational Chemistry* **1995,** 16, (12), 1449-1458.

56. Cances, E.; Mennucci, B.; Tomasi, J., A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *Journal of Chemical Physics* **1997,** 107, (8), 3032-3041.

57. Mierts, S.; Scrocco, E.; Tomasi, J., Electrostatic interaction of a solute with a continuum. A direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics* **1981,** 55, (1), 117-129.

58. Tomasi, J., Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theoretical Chemistry Accounts* **2004,** 112, (4), 184-203.

59. Klamt, A., Conductor-like screening model for real solvents - a new approach to the quantitative calculation of solvation phenomena. *Journal of Physical Chemistry* **1995,** 99, (7), 2224-2235.

60. Klamt, A.; Schuurmann, G., COSMO - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society: Perkin Transactions 2* **1993**, (5), 799-805.

61. Rinaldi, D.; Bouchy, A.; Rivail, J. L., A self-consistent reaction field model of solvation using distributed multipoles. II: Second energy derivatives and application to vibrational spectra. *Theoretical Chemistry Accounts* **2006,** 116, (4-5), 664-669.

62. Rinaldi, D.; Bouchy, A.; Rivail, J. L.; Dillet, V., A self-consistent reaction field model of solvation using distributed multipoles. I. Energy and energy derivatives. *Journal of Chemical Physics* **2004,** 120, (5), 2343-2350.

63. Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *Journal of Physical Chemistry* **1996,** 100, (40), 16385-16398.

64. Cramer, C. J.; Truhlar, D. G., General parameterized SCF model for free energies of solvation in aqueous solution. *Journal of the American Chemical Society* **1991,** 113, (22), 8305-8311.

65. Cramer, C. J.; Truhlar, D. G., AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *Journal of Computer-Aided Molecular Design* **1992,** 6, (6), 629-666.

66. Cramer, C. J.; Truhlar, D. G., PM3-SM3 - a general parameterization for including aqueous solvation effects in the PM3 molecular orbital model. *Journal of Computational Chemistry* **1992,** 13, (9), 1089-1097.

67. Cramer, C. J.; Truhlar, D. G., An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992,** 256, (5054), 213-217.

68. Giesen, D. J.; Cramer, C. J.; Truhlar, D. G., A semiempirical quantum mechanical solvation model for solvation free energies in all alkane solvents. *Journal of Physical Chemistry* **1995,** 99, (18), 7137-7146.

69. Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G., A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents. *Theoretical Chemistry Accounts* **1997,** 98, (2-3), 85-109.

70. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G., SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *Journal of Chemical Theory and Computation* **2005,** 1, (6), 1133-1152.

71. Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G., Extension of the platform of applicability of the SM5.42R universal solvation model. *Theoretical Chemistry Accounts* **1999,** 103, (1), 9-63.

72. Thompson, J. D.; Cramer, C. J.; Truhlar, D. G., Density-functional theory and hybrid density-functional theory continuum solvation models for aqueous and organic solvents: universal SM5.43 and SM5.43R solvation models for any fraction of Hartree-Fock exchange. *Theoretical Chemistry Accounts* **2005,** 113, (2), 107-131.

73. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** 102, (18), 3586-3616.

74. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B., Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins* **1986,** 1, (1), 47-59.

75. Holst, M.; Saied, F., Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry* **1993,** 14, (1), 105-113.

76. Holst, M. J. S., F., Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods. *Journal of Computational Chemistry* **1995,** 16, (3), 337-364.

77. Zhou, Z. X.; Payne, P.; Vasquez, M.; Kuhn, N.; Levitt, M., Finite-difference solution of the Poisson-Boltzmann equation: Complete elimination of self-energy. *Journal of Computational Chemistry* **1996,** 17, (11), 1344-1351.

78. Stone, A. J., *The Theory of Intermolecular Forces*. Clarendon Press: Oxford, 1996; Vol. 32, p 264.

79. de Boor, C., *A Practical Guide to Splines*. Springer: New York, 2001; p 346.

80. Prabhu, N. V.; Zhu, P. J.; Sharp, K. A., Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *Journal of Computational Chemistry* **2004,** 25, (16), 2049-2064.

81. Grant, J. A.; Pickup, B. T.; Nicholls, A., A smooth permittivity function for Poisson-Boltzmann solvation methods. *Journal of Computational Chemistry* **2001,** 22, (6), 608-640.

82. Applequist, J., Traceless cartesian tensor forms for spherical harmonic functions - new theorems and applications to electrostatics of dielectric media. *Journal of Physics A: Mathematical and General* **1989,** 22, (20), 4303-4330.

83. Applequist, J., Maxwell-Cartesian spherical harmonics in multipole potentials and atomic orbitals. *Theoretical Chemistry Accounts* **2002,** 107, (2), 103-115.

84. Young, D. M., *Iterative Solutions of Large Linear Systems*. Acedemic Press: New York, 1971; p p. 570.

85. Böttcher, C. J. F., *Dielectrics in Static Fields*. 2 ed.; Elsevier Pub. Co.: Amsterdam, 1993; Vol. 1.

86. Böttcher, C. J. F., *Dielectrics in Static Fields*. 1 ed.; Elsevier Pub. Co.: Amsterdam, 1952; Vol. 1.

87. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic Acids Research* **2000,** 28, (1), 235-242.

88. Teeter, M. M., Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proceedings of the National Academy of Sciences* **1984,** 81, (19), 6014-6018.

89. Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O., Structural studies of the engrailed homeodomain. *Protein Science* **1994,** 3, (10), 1779-1787.

90. Dahiyat, B. I.; Mayo, S. L., De novo protein design: Fully automated sequence selection. *Science* **1997,** 278, (5335), 82-87.

91. Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L., 2 crystal structures of the B1 immunoglobulin binding domain of Streptococcal Protein-G and comparison with NMR. *Biochemistry* **1994,** 33, (15), 4721-4729.

92. McKnight, C. J.; Matsudaira, P. T.; Kim, P. S., NMR structure of the 35-residue villin headpiece subdomain. *Nature Structural Biology* **1997,** 4, (3), 180-184.

93. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984,** 81, (8), 3684-3690.

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

95. Ponder, J. W. *TINKER: Software Tools for Molecular Design*, 4.2; Saint Louis, MO, 2004.

96. Grycuk, T., Deficiency of the Coulomb-field approximation in the Generalized Born model: An improved formula for Born radii evaluation. *Journal of Chemical Physics* **2003,** 119, (9), 4817-4826.

97. Jackson, J. D., *Classical Electrodynamics*. 3rd edition ed.; John Wiley & Sons, Inc.: New York, 1998.

98. Tjong, H.; Zhou, H. X., GBr6: A parameterization-free, accurate, analytical Generalized Born method. *Journal of Physical Chemistry B* **2007**.

99. Tanford, C.; Kirkwood, J. G., Theory of protein titration curves. I. General equations for impenetrable spheres. *Journal of the American Chemical Society* **1957,** 79, (20), 5333-5339.

100. Bondi, A., van der Waals volumes and radii. *Journal of Physical Chemistry* **1964,** 68, (3), 441-451.

101. Challacombe, M.; Schwegler, E.; Almlof, J., Recurrence relations for calculation of the Cartesian multipole tensor. *Chemical Physics Letters* **1995,** 241, (1-2), 67-72.

102. McMurchie, L. E.; Davidson, E. R., One- and two-electron integrals over Cartesian Gaussian functions. *Journal of Computational Physics* **1978,** 26, (2), 218-231.

103. Huang, D. M.; Chandler, D., The hydrophobic effect and the influence of solute-solvent attractions. *Journal of Physical Chemistry B* **2002,** 106, (8), 2047-2053.

104. Huang, D. M.; Chandler, D., Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** 97, (15), 8324-8327.

105. Huang, D. M.; Geissler, P. L.; Chandler, D., Scaling of hydrophobic solvation free energies. *Journal of Physical Chemistry B* **2001,** 105, (28), 6704-6709.

106. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005,** 437, (7059), 640-647.

107. Chen, J. H.; Brooks, C. L., Critical importance of length-scale dependence in implicit modeling of hydrophobic interactions. *Journal of the American Chemical Society* **2007,** 129, (9), 2444-5.

108. Zwanzig, R. W., High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics* **1954,** 22, (8), 1420-1426.

109. Gallicchio, E.; Kubo, M. M.; Levy, R. M., Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *Journal of Physical Chemistry B* **2000,** 104, (26), 6271-6285.

110. Schnieders, M. J.; Ponder, J. W., Implicit solvents for the AMOEBA force field based on Poisson-Boltzmann and Generalized Kirkwood electrostatics (in preparation).

111. Barone, V.; Cossi, M.; Tomasi, J., A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *Journal of Chemical Physics* **1997,** 107, (8), 3210-3221.

112. Beglov, D.; Roux, B., Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *Journal of Chemical Physics* **1994,** 100, (12), 9050-9063.

113. Schnieders, M. J.; Ponder, J. W., Closed form solutions for the reaction field due to off-center point multipoles (in preparation).

114. Kirkwood, J. G., The dielectric polarization of polar liquids. *Journal of Chemical Physics* **1939,** 7, (10), 911-919.

115. Fröhlich, H., *Theory of Dielectrics*. 2 ed.; Oxford University Press: London, 1958.

116. Rick, S. W.; Berne, B. J., The aqueous solvation of water: A comparison of continuum methods with molecular dynamics. *Journal of the American Chemical Society* **1994,** 116, (9), 3949-3954.

117. Golubkov, P. A.; Ren, P., Generalized coarse-grained model based on point multipole and Gay-Berne potentials. *Journal of Chemical Physics* **2006,** 125, (6), 64103.

118. Marshall, S. A.; Vizcarra, C. L.; Mayo, S. L., One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. *Protein Science* **2005,** 14, (5), 1293-1304.

119. Thompson, J. D.; Cramer, C. J.; Truhlar, D. G., New universal solvation model and comparison of the accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM continuum solvation models for aqueous and organic solvation free energies and for vapor pressures. *Journal of Physical Chemistry A* **2004,** 108, (31), 6532-6542.

Curriculum Vitae

Michael J. Schnieders

Place of Birth Iowa City, IA

EDUCATION

Doctorate of Science, Biomedical Engineering, December 2007
Washington University, St. Louis, MO
Dissertation: The Theory and Effect of Solvent Environment on Biomolecules
Advisor: Jay W. Ponder
GPA 3.8

Bachelor of Science in Engineering, Biomedical Engineering, 1999
University of Iowa, Iowa City, IA
GPA 3.9, With High Distinction (Top 5% of University Class)
MCAT Physical Sciences 15, Biological Sciences 14, Verbal Reasoning 12
GRE Quantitative 770, Verbal 650, Analytic 680

HONORS/AFFILIATIONS

- Grace Norman Scholarship (2001)
- Rhodes Dunlap Scholarship (1998)
- National Barry Goldwater Excellence in Education Scholarship (1997)
- Alpha Eta Mu Beta Biomedical Engineering Honor Society, Top 20% of BME Class (1997)
- Paul D. Scholz Memorial Scholarship (1996)
- Tau Beta Pi Engineering Honor Society, Top 12.5% of Engineering Class (1996)
- Sigma Xi Scientific Research Society (1996)
- Stebler Scholarship (1995)
- University of Iowa Honor Society (1994-1999)

RESEARCH INTERESTS

- Theory and development of molecular models for biomolecular systems
- Application of molecular models to understand and develop treatments for
    1. Osteoarthritis
    2. Cystic Fibrosis

TEACHING INTERESTS

- Undergraduate biology and computer science courses
- Graduate computational biochemistry and statistical mechanics courses

RELATED EXPERIENCE

Research

1. Post-Doctoral Fellow, Laboratory of Professor Vijay Pande, Department of Chemistry, Stanford University, Palo Alto, CA, Fall 2007
    - Propose to apply the AMOEBA force field with Generalized Kirkwood implicit solvent to study the molecular constituents of osteoarthritis and cystic fibrosis using the Folding at Home distributed computing project.

2. Pre-Doctoral Fellow/D.Sc. Research, Laboratory of Professor Jay W. Ponder, Department of Biomedical Engineering, Washington University, Saint Louis, MO, 2001 - 2007
    - Derived and implemented a numerical continuum electrostatics model for a polarizable multipole force field based on the linearized Poisson-Boltzmann equation[1].
    - Derived and implemented an analytic approximation to solving the linearized Poisson-Boltzmann equation numerically that extends the generalized Born model to polarizable multipoles, termed Generalized Kirkwood[2].
    - Derived, implemented and parameterized complete implicit solvent models for the AMOEBA force field based on Poisson-Boltzmann or generalized Kirkwood electrostatics.

3. Howard Hughes Research Assistantship, Laboratory of Professor Thomas Brown, Orthopaedic Biomechanics Laboratory, University of Iowa, Iowa City, IA, 1997 – 1999
    - Studied the accuracy of a surgical drill guide for placing grafts or pins through the femoral neck and into the femoral head[3, 4]
    - Quantified the mechanical properties of osteonecrotic femoral heads and a composite fiberglass surrogate[5]

4. Research Assistant, Supervised by Dr. James Martin, Ponseti Biochemistry and Cell Biology Laboratory, University of Iowa, Iowa City, IA, 1996-1997
   - Applied image analysis techniques to measure staining of IGF-1 in articular cartilage from confocal microscopy images[6]

5. Laboratory Technician, Supervised by Kenneth Moore, Central Microscopy Research Facility, University of Iowa, Iowa City, IA, 1995-1996
   - Maintained equipment and reagents for processing of specimens
   - Supported researchers in their use of SEM, TEM, AFM and Confocal microscopy techniques

Teaching

1. Volunteer Teaching Assistant, Department of Biochemistry and Molecular Biophysics, Washington University, Saint Louis, MO, 2004-2007
   - Instructed students in the application of the TINKER and Force Field Explorer programs during a Computational Biochemistry course[7]

2. Volunteer Tutor, Department of Biochemistry and Molecular Biophysics, Washington University, Saint Louis, MO, 2003
   - Solicited to be a private tutor for graduate students taking an advanced course in Statistical Thermodynamics

3. Teaching Assistant, Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, 1998
   - Led two lab sections per week for the Computers in Engineering course
   - Graded assignments and tests

4. Howard Hughes Teaching Assistant, Department of Biology, University of Iowa, Iowa City, IA 1997
   - Lead two study sections per week for an undergraduate introductory biology course

PUBLICATIONS

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

2. Schnieders, M. J.; Ponder, J. W., Polarizable atomic multipole solutes in a generalized Kirkwood continuum (to appear). *Journal of Chemical Theory and Computation* **2007**.

3. Schnieders, M. J.; Dave, S. B.; Morrow, D. E.; Heiner, A. D.; Pedersen, D. R.; Brown, T. D., Assessing the accuracy of a prototype drill guide for fibular graft placement in femoral head necrosis. *Iowa Orthop J* **1997,** 17, 58-63.

4. Anderson, D. A.; Schnieders, M. J.; Heiner, A. D.; Pedersen, D. R.; Brown, T. D.; Brand, R. A., A Surgical Guide to Accurately Place Pins or Nails Within the Femoral Head. *Journal of Musculoskeletal Research* **1999,** 3, (3), 233.

5. Heiner, A. D.; Brown, T. D.; Schnieders, M. J., Structural behavior of composite fiberglass surrogate vs. natural human femoral heads: Implications for avascular necrosis modeling. *Transactions of the American Society of Biomechanics* **1997,** 21, 302.

6. Martin, J. A.; Ellerbroek, S. M.; Schnieders, M. J.; Buckwalter, J. A., Inhibition of IGF-1 Response in Osteoarthritic Cartilage: A Cause for Cartilage Degeneration. *Transactions of the Orthopaedic Research Society Meeting* **1997**.

7. Schnieders, M. J. *Force Field Explorer, Version 4.2*, 2004.

December 2007